Selection of Information Streams in Social Sensing: an Interdependence- and Cost-aware Ranking Method

Gabriele Gianini
Emirates ICT Innovation Centre
Khalifa University of Science and Technology, Abu Dhabi, UAE
Università degli Studi di Milano
Milano, Italy
gabriele.gianini@unimi.it

Corrado Mio
Emirates ICT Innovation Centre
Khalifa University of Science and Technology, Abu Dhabi, UAE
Università degli Studi di Milano
Milano, Italy
corrado.mio@ku.ac.ae

Francesco Viola
SESAR Lab
Dipartimento di Informatica
Università degli Studi di Milano
Milano, Italy
francesco.viola1@unimi.it

Jianyi Lin
Department of Mathematics
Khalifa University of Science and Technology (KUST)
Abu Dhabi, UAE
jianyi.lin@ku.ac.ae

Nawaf Almoosa
Emirates ICT Innovation Centre
Khalifa University of Science and Technology (KUST)
Abu Dhabi, UAE
nawaf.almoosa@ku.ac.ae

ABSTRACT

In this work we address the problem of critical source selection in social sensing. We propose an approach to the ranking of information streams, which is aware of the interdependence among streams (redundancy and synergies), of the cost of individual streams, and of the cost related to the integration of multiple streams. The method is based on the use of the Coalitional Game Theory concept of Power Index, and relies on the polynomial-time estimate of the stream sets characteristics. With respect to other works using a power index, the method takes into account that the problem has a non-trivial cost structure.

CCS CONCEPTS

• **Information systems → Extraction, transformation and loading**; *Multimedia information systems*.

KEYWORDS

Social Sensing; Critical Source Selection; Source Management; Coalitional Game Theory; Power Index; Shapley Value; Banzhaf Value

ACM Reference Format:

Gabriele Gianini, Corrado Mio, Francesco Viola, Jianyi Lin, and Nawaf Almoosa. 2020. Selection of Information Streams in Social Sensing: an Interdependence- and Cost-aware Ranking Method. In *12th International Conference on Management of Digital EcoSystems (MEDES '20), November 2–4, 2020, Virtual Event, United Arab Emirates.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3415958.3433099

MEDES '20, November 2–4, 2020, Virtual Event, United Arab Emirates

ACM ISBN 978-1-4503-8115-4/20/11…$15.00
https://doi.org/10.1145/3415958.3433099

1 INTRODUCTION

The proliferation of various kinds of sensors, in the availability of many people (e.g. smart-phones), along with the wide adoption of online social media (e.g. Twitter, Instagram, Facebook), has fostered the development of social sensing, a paradigm for collecting, organizing and analyzing observations, originated from individuals.

1.1 Truth discovery

The final goal of sensing is the so-called *truth discovery* – e.g. in the form of detection of a specific kind of event occurring in the real world (such as a traffic jam on a highway [12]), or the determination of the state of affairs of a given area (e.g. localization of dangerous potholes [23]) or in a wide region (e.g. monitoring the spread of a disease [26]). In the process, the information – originated from the diverse, possibly heterogeneous sources, that represent as many views of the phenomenon of interest – has to undergo some form of fusion [35], whose complexity may vary considerably from one case to the other, but is typically characterized by some recurrent challenges.

Due to the presence of humans in the loop, the possible unreliability, inaccuracy and imprecision of the gathered information (e.g. malicious claims) represent one of the main challenges [34]. A wide literature addresses this point, e.g. some authors approach the problem by gauging the contents of a multiplicity of observations to hedge against the different reliability of the observers (for a review see [1]).

1.2 Critical Source Selection

Another very frequent challenge is posed by the sheer number of sources itself. In relation to this challenge, the task of *critical source selection* consists in "*identifying a subset of critical sources that can help effectively reduce the computational complexity of the original truth-discovery problem at the same time improving the accuracy of the analysis results*" [17]. Indeed, not only gathering and scanning a large number of sources is by definition burdensome, but the analysis algorithms have to work on combinations of different streams. For instance event detection algorithms typically have to compare and merge several streams into a synthesis information flow, richly annotated and more suitable for the detection task: as a consequence, the computational complexity is often at least quadratic in the number of streams. Being able to single out a set of high quality sources (known as critical source set, that here we call, for short, the *golden bucket*), therefore, becomes a must.

The relationships among sources add an additional complexity to this selection challenge: on one side a human source can forward claims received from others (a re-tweet to followers), thus creating undesirable redundancy, on the other, sources providing complementary information can bring to light a truth that each individual source would not be able to convey if taken alone. Indeed, the latter is the main motivation behind the fusion of diverse information flows: one performs the fusion under the assumption that at least some of the different sources are synergetic in reconstructing the actual state of affairs.

Finally, the character of sources (speak rate, reliability, etc.) in general evolves with time, so that the golden bucket has to be kept up to date: sources have to be reassessed from time to time. Overall this involves a considerable management effort, considering that sources are note exploited in isolation, but become useful when all their synergetic relations are taken into account.

In this work we propose an approach to support the source selection and management (within the latter we restrict our attention to the periodical updates of the golden bucket) based on a concept related to the Coalitional Game Theory notion of Power Index [11]. Our approach relies on the polynomial-time estimate of the stream sets characteristics. With respect to other works using a power index, the method takes into account that the problem has a non-trivial cost structure.

The reminder of the paper is structured as follows. In Section 2 we overview the structure and rationale of the approach; in Section 3 we point to the related work and stress the difference between the proposed approach and the existing works using the power indices as a ranking criterion, both in terms of semantics and in terms of structural differences; in Section 4 we develop the approach.

2 OVERVIEW AND RATIONALE OF THE APPROACH

Power Indices are used in Coalitional Game Theory (CGT) to quantify the contribution by an individual player to a coalition in terms of the achievement of a given payoff (or return) by the coalition. They take into account the fact that, in typical collaborative tasks, the contributions do not compose additively.

2.1 Preliminaries and definitions

A Power Index based approach in social sensing can take into account the redundancy and synergy (i.e. negative and positive interactivity) of the information streams in a principled way: it assigns to each source a synthesis score representing a balance of its power in isolation plus its average level of interactivity with two or more other sources. This allows for more effective source selection and, at the same time, simplifies the periodic reassessment of the information source, allowing for source-by-source oriented decision about the inclusion in the golden bucket.

2.1.1 Some definition and notation. We provide the formal definition of the CGT power indices in the appropriate section. For the present preliminary discussion is sufficient to recall the terms of the analogies used along this work, and some basic characteristics of those indices.

In the present context a *player* corresponds to an information stream, whereas a *coalition S* corresponds to a set of streams fused together by a detection algorithm.

The *performance* of the coalition can be defined in many ways depending on the context, and based on the usefulness of the prediction results (e.g. it could be some measure of accuracy): we assume, for the sake of simplicity, that the performance metric is a real valued.

We denote an individual player by p, a coalition by S and its performance by $A(S)$. If there is a set N of $n = |N|$ players there are 2^n possible coalitions, 2^{n-1} of which contain player p.

The *marginal contribution* $\Delta_p(S)$ by a player p to a coalition S is defined as the difference between the performance of the coalition with and without that player

$$\Delta_p(S) \equiv A((S \setminus p) \cup p) - A((S \setminus p)) \tag{1}$$

2.1.2 Power Index definition. In order to take into account all the ways in which the player p can interact with others, one can consider the marginal contributions w.r.t. all the coalitions and take an average. A power index α_p for a player p is always defined as suitably weighted average over all the coalitions of the marginal contributions by the player:

$$\alpha_p = \sum_{S \in 2^{(N \setminus i)}} w_S \Delta_p(S) \tag{2}$$

where the w_S are predefined weights (the symbol $2^{(N \setminus i)}$ denotes the power set of the set $\{N \setminus i\}$). For instance, if all the marginal contributions enter into the average with the same weight $w_S = 1/2^{n-1}$, irrespectively of the coalition size, we obtain the Banzhaf Value α_p^{Ban} [5, 20], whereas if the cardinality $s = |S|$ of the coalition is taken into consideration by $w_S = (1/n)/(\binom{n-1}{s})$, one gets the Shapley Value α_p^{Sha} [24, 28, 29].

The different CGT power indices have been defined to fulfill different sets of axioms. E.g. the Shapley and the Banzhaf Value are the unique solutions to two slightly different sets of axioms expressing fair division of the revenue of a coalition: such a value is meant to quantify the influence of each player in achieving that revenue and to assign to the player a fair share of the revenue.

This is the intuitive reasoning behind the choice of a power index as a measure of player importance. By assigning the information streams a Game Theoretic importance and keeping the most valuable, intuitively, one can approximate the best coalition, in some sense.

2.2 The approach at a glance

An approach to the selection of the information streams based on a power index can be synthesized as follows:

(1) compute the power index of each stream based on recent historical data;
(2) use the power index as a score of the stream
(3) keep only the k top scored streams, in the golden bucket;
(4) periodically re-assess the streams
(5) decide on a stream-by-stream basis whether to add it to the bucket, keep it in the bucket, or drop it.

2.2.1 *Rationale.*

Beyond the CGT intuitive justifications, the mechanics behind the Power Index approach is the following. The different coalitions and their performances define the set function $A : 2^N \to \mathbb{R}$. Such function can be represented [14] as a polynomial of degree n

$$A(S) = \sum_{T \subset 2^N} a_T \prod_{p \in T} x_p \tag{3}$$

of the Boolean variables x_p,

$$x_p \equiv \begin{cases} 1 & if \quad p \in S \\ 0 & if \quad p \notin S \end{cases} \tag{4}$$

The a_T's are called Moebius coefficients. The function can be approximated by a first degree function $\alpha(S)$ which minimizes some distance criterion

$$\alpha(S) = \alpha_0 + \sum_{p \in S} \alpha_p x_p \tag{5}$$

where α_0 is an offset constant.

The coefficient α_p turns out to coincide with a power index of the element p. Which power index precisely depends on the definition of distance and minimization criterion: for instance if the distance is the Euclidean distance between A and α computed over all the sets of 2^N considered equivalent, one obtains the Banzhaf Value (it was shown by Hammer and Holzman [13]).

We are normally interested in the optimal coalition, which would be the maximum of A, an that we denote $S^{max} = \arg\max(A)$. Being unfeasible to check all the 2^n coalitions to find the max of the n-degree function A, one uses the first degree function α as a proxy.

If α is given, i.e. if the coefficients α_p are available, the maximum of α can be found easily: gathering the k highest coefficients α_p and setting their indicator variables x_p to 1 (on), and keeping to 0 (off) the other $(n-k)$ variables one gets the maximum of the function α over the level of sets of size k: the "on" variables identify the set maximizing α. That set can be used as a candidate maximum for A. Of course this is a rough approximation but can be computed in linear time and is appropriate to the situations in which the requirement of finding the actual maximum is not stringent, whereas finding a high performing coalition is accepted as a practical solution.

2.2.2 *Feasibility.*

The feasibility of the overall approach relies on the assumption that one can compute the power indices α_p with little effort. Indeed, although the definition (2) suggests that the computation of the power indices has exponential complexity, in practice, a sufficiently useful estimate of the index can be obtained in polynomial time by sampling a reasonable number of coalitions [10]: in fact one does not have to discover the exact value of the indices, but for the proposed approach one just needs to find out which ones are the highest k indices.

3 RELATED WORK

3.1 Social Sensing

A vast literature addresses the problem of source selection in the communities of networked sensing and data mining [4, 8, 15, 27, 33]. For example: Rekatsinas et al. [27] consider dynamic sources whose contents change with time; Dong et al. [8] focus on integration cost aware source selection. Uddin et al. [33] focus on diversifying the source selection based on the social connections among sources; and previous papers.

Previous works [16, 25, 34] have considered the source dependency by partitioning into mutually independent groups. Huang and Wang [17] propose a scheme to find the critical set of sources by explicitly exploring both interdependency and speak rate of sources in the context of social sensing. They formulate the problem as a multi-objective constraint optimization problem and develop an efficient solution algorithm.

3.2 Power Indices

The power indices in general, and the Shapley value [29] in particular, are the subject of wide literature [24], where they are considered as a solution of the fair division problem [28], as a centrality measure [30] or even as a transform within the Dempster-Shafer Theory of Evidence [31] (or Theory of Belief Functions).

It has often been used to assess the importance of components in a composite systems or processes. Among the recent examples, we can mention its application in tag sense disambiguation [19], and neural network pruning [32]. It has also been used in algorithm portfolio selection [9], feature selection [6, 7] and it has been proposed within Explanatory Artificial Intelligence for optimal model approximation [21, 22].

3.3 Comparison with a similar setting

The approach of the present work bears some analogy with the problems of classification rule selection and rule pool management in fraudulent credit card transaction detection as it is addressed in the work [10] (that work uses the Shapley Value as a scoring index). However the very nature of the involved objects (classification rules on one side and information streams on the other) makes the actual details of the methods rather different.

3.3.1 *A semantic difference.*

In the so-called *fraud detection scenario*, the rule pool is confronted with a very well defined event at time, coming from a single source: the transaction (there is a given, relatively simple, and fixed set of attributes associated to the transaction which take values in predefined ranges and one precise time-stamp). Furthermore the transactions used to assess the rules have a precise classification either in the category *legitimate* or in the category *fraudulent*. Thus we are dealing with a *classification task*.

In an *event detection* scenario (as an example of truth discovery) we have to reconstruct the event form different sources. We have categories, but not only the two most obvious (the event of interest, e.g. a traffic jam, being present or absent).

In fact there can be even an *intensity* associated to the event presence (a traffic jam has several degrees of severity, going from the slowing down of the traffic to the complete stop for extreme congestion/accidents [2, 3]). Furthermore, an event has a space structure (the speed of the traffic flow is space-variant, and time-variant, there might be waves of *car density*).

In short, an event, or more in general, the truth to be discovered, are richly structured phenomena and their description is very different from the one of credit card transactions.[1] Even so, should one manage to formulate the performance metric in terms of a single real number the formalism could be the same used in reference [10]. However the two setting are distinguished also by a structural difference, which forces the adoption of different modeling tools.

3.3.2 A structural difference: presence of costs. The main difference between the two scenarios is that whereas the the fraud detection scenario the players (the classification rules) have essentially the same utilization cost, the information fusion scenario players have an associated cost (intended in a sense that we specify below), which can change considerably from one stream to the other.

Indeed, within fraud detection, each rule has a computation cost but the differences from one rule to the other in terms of computation costs for a given transaction are negligible. On the contrary, being the information streams very complex objects, they can differ greatly in terms of the resources that have to be employed for their treatment. The differences between two sources can be due to their different information rate (e.g. number of tweets in Twitter), type of information communicated (text, images, videos) which determine a single stream cost in terms of processing complexity etc. Not last, one can assume that the use of different streams can be associated to different economic costs (e.g. the cost corresponded to the stream provider).

In short, using a player in the information fusion setting has a cost that has to be taken into account. The manager of a service for event detection from information streams must be aware of the costs and will try on one side to minimize them, while on the other side will try to maximize the performance.

This problem of multi-objective optimization can be brought back to a single variable problem, which can be addressed, then, within the framework of the power index approach: a balance quantity has to be defined as an appropriate function of performance metric and and cost metric.

A way of achieving this goal could be to transform the costs in suitable weights associated to individual players: an approach which uses a tool known as Weighted Power index. An example of this tool is the Weighted Shapley Value [18]: this power index assigns to the player a value that ends up being proportional to the ratio of its (un-weighted) Shapley Value and the a priori known player cost.

[1] In other words, in the fraud detection scenario we have Boolean categories to guess, whereas in the information fusion scenario we have – in the simplest case – a real variable to estimate (e.g. the intensity of the phenomenon). This impacts the choice of the performance quantifiers for a coalition: in the former scenario we can define the performance in terms of a classification performance metric (precision, recall, F2-score etc.), whereas in the latter scenario there is a larger spectrum of options. Each event estimate originates an error, thus one could choose as a performance metric the average error, or the maximum error or another aggregator, depending on the application. The semantic of the performance metric and therefore of the marginal contributions and of the power index are very different.

We follow a different pathway: we consider also non-individual costs and include the cost structure into the set function to be optimized.

4 THE APPROACH

As discussed in Section 2, the problem of selecting k critical sources out of n sources available in the repertoire $N = \{1, 2, \dots, n\}$ can be framed as a set function maximization problem (an NP-hard problem), and approached as a power index based maximization.

Normally, as in [10] one would define the set function to be maximized as the performance $A(S)$ of the set of streams $S \in 2^N$ (in detecting the event of interest) – quantified for instance through historical data – and would try to approximate A by a first degree function α (equation (5)), disregarding any cost-related consideration. The coefficients of the first degree polynomial α, once ranked, would indicate the set of elements forming the coalition which is the maximum of α, to be used as candidate maximum of A.

We argued in Section 3 that – differently from the case dealt with in [10] – the problem at hand has a non-negligible cost structure: each stream i has a cost $C(i)$, even when considered in isolation (economic cost of the acquisition, processing costs, etc.).

Here we add that also the integration of two streams has a cost which depends on the relationship between the two streams (consider for instance the task of time alignment: its cost depends on the patterns of both streams, and is not simply the sum of some time-stamp-related cost specific to each stream). This implies the presence of an at least pairwise cost structure: each pair of streams will have an associated cost $c(ij)$ in addition to the cost given by the sum of the individual costs. So

$$C(ij) = c_i + c_j + c_{ij}$$

To formulate the model without making too many restrictive hypotheses, one can assume that, to the cost of using the set S, contribute relationships of any order up to s. Considering all the possible sets one can, thus, define a Boolean set function, the cost function C, formally mirroring the performance function A (equation (3))

$$C(S) = \sum_{T \subset 2^N} c_T \prod_{p \in T} x_p \tag{6}$$

We observe that, also for this one, one can define a first degree approximation $\gamma(S)$, with $S \in 2^N$, and a power index γ_i from its coefficients (e.g. the Shapley or the Banzhaf Value).

We define the following *balance* function

$$B(S) \equiv u_0 A(S) - C(S) \tag{7}$$

with $S \in 2^N$, where u_0 is a constant that converts the performance value in economic value, so that $u_0 A(S)$ can be compared to the costs $C(S)$. Obviously $u_0 A(S)$ represents a linear utility function that maps a performance onto an economic value, but more sophisticated utility functions u could be devised: we use a linear function $u(A) \equiv u_0 A$ for the sake of illustration of the method.

Now $B(S)$ is the new objective function to maximize, and we can approach the task using the power index approach. Following the procedure used for A, we approximate B by a first degree polynomial β and obtain the power indices β_i. By additivity, the power indices if this function will be

$$\beta_i = u_0 \alpha_i - \gamma_i \tag{8}$$

In practice it is reasonable to assume that the cost function C can be well modeled by a second degree function (accounting for the cost of streams in isolations and the cost of streams in pairs). Therefore the overall computation can be simplified.

First one computes the power indices γ_i. For instance, both for the Shapley and the Banzhaf value, they will be

$$\gamma_i = c_0 + c_i + \frac{1}{2} \sum_{j \in N \setminus i} c_{ij} \tag{9}$$

All the components of this equation ideally should be computable a priori or by sampling based on features of the individual streams.

If the assumption holds that one can model the cost as a quadratic function the power index β_i to be used for ranking will be the following

$$\beta_i = u_0 \alpha_i - (c_0 + c_i + \frac{1}{2} \sum_{j \in N \setminus i} c_{ij}) \tag{10}$$

This expression allows for an effective source selection and simplifes the periodic reassessment of the information sources, allowing for source-by-source oriented decision about the inclusion in the golden bucket.

5 CONCLUSIONS

In this work we proposed a power index based approach to the ranking of information streams in social sensing, which is aware of the interdependence among streams (redundancy and synergies), of the cost of individual streams and the cost related to the integration of several streams. The method is based on polynomial time estimate of the stream sets characteristics.

We plan to extend this work by developing the approach in more detail and to address a real-world application case example.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their advise, which has significantly helped in improving the paper.

REFERENCES

[1] Charu C Aggarwal and Tarek Abdelzaher. 2013. Social sensing. In *Managing and mining sensor data*. Springer, 237–297.

[2] Ahmed Al Dhanhani, Ernesto Damiani, Rabeb Mizouni, and Di Wang. 2018. Analysis of Shapelet Transform Usage in Traffic Event Detection. In *2018 IEEE International Conference on Cognitive Computing (ICCC)*. IEEE, 41–48.

[3] Ahmed Al Dhanhani, Ernesto Damiani, Rabeb Mizouni, and Di Wang. 2019. Framework for traffic event detection using Shapelet Transform. *Engineering Applications of Artificial Intelligence* 82 (2019), 226–235.

[4] Haleh Amintoosi, Salil S Kanhere, and Mohammad Allahbakhsh. 2015. Trust-based privacy-aware participant selection in social participatory sensing. *Journal of Information Security and Applications* 20 (2015), 11–25.

[5] John F Banzhaf III. 1964. Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.* 19 (1964), 317.

[6] Shay Cohen, Gideon Dror, and Eytan Ruppin. 2007. Feature selection via coalitional game theory. *Neural Computation* 19, 7 (2007), 1939–1961.

[7] Shay Cohen, Eytan Ruppin, and Gideon Dror. 2005. Feature Selection Based on the Shapley Value.. In *IJCAI*, Vol. 5. 665–670.

[8] Xin Luna Dong, Barna Saha, and Divesh Srivastava. 2012. Less is more: Selecting sources wisely for integration. *Proceedings of the VLDB Endowment* 6, 2 (2012), 37–48.

[9] Alexandre Fréchette, Lars Kotthoff, Tomasz Michalak, Talal Rahwan, Holger H Hoos, and Kevin Leyton-Brown. 2016. Using the shapley value to analyze algorithm portfolios. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[10] Gabriele Gianini, Leopold Ghemmogne Fossi, Corrado Mio, Olivier Caelen, Lionel Brunie, and Ernesto Damiani. 2020. Managing a pool of rules for credit card fraud detection by a Game Theory based approach. *Future Generation Computer Systems* 102 (2020), 549–561.

[11] Michel Grabisch. [n.d.]. *Set functions, games and capacities in decision making*. Vol. 46. Springer.

[12] André Guéziec. 2014. Crowd sourced traffic reporting. US Patent 8,718,910.

[13] Peter L Hammer and Ron Holzman. 1992. Approximations of pseudo-Boolean functions; applications to game theory. *Zeitschrift für Operations Research* 36, 1 (1992), 3–21.

[14] Peter L Hammer and Sergiu Rudeanu. 1968. *Boolean methods in operations research and related areas*. Vol. 7. Springer-Verlag New York Inc.

[15] Mohammad Hosseini, Nooreddin Nagibolhosseini, Amotz Barnoy, Peter Terlecky, Hengchang Liu, Shaohan Hu, Shiguang Wang, Tanvir Amin, Lu Su, Dong Wang, et al. 2015. Joint source selection and data extrapolation in social sensing for disaster response. *arXiv preprint arXiv:1512.00500* (2015).

[16] Chao Huang and Dong Wang. 2016. Topic-aware social sensing with arbitrary source dependency graphs. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 1–12.

[17] Chao Huang and Dong Wang. 2017. Critical source selection in social sensing applications. In *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 53–60.

[18] Ehud Kalai and Dov Samet. 1987. On weighted Shapley values. *International Journal of Game Theory* 16, 3 (1987), 205–222.

[19] Meshesha Legesse, Gabriele Gianini, and Dereje Teferi. 2016. Selecting Feature-Words in Tag Sense Disambiguation Based on Their Shapley Value. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2016 12th International Conference on*. IEEE, 236–240.

[20] Ehud Lehrer. 1988. An axiomatization of the Banzhaf value. *International Journal of Game Theory* 17, 2 (1988), 89–99.

[21] Stan Lipovetsky and Michael Conklin. 2001. Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry* 17, 4 (2001), 319–330.

[22] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*. 4765–4774.

[23] Paul Marks. 2013. Crowds point out potholes on a map to speed up street repairs.

[24] Stefano Moretti and Fioravante Patrone. 2008. Transversality of the Shapley value. *TOP* 16, 1 (19 Apr 2008), 1. https://doi.org/10.1007/s11750-008-0044-5

[25] Ravali Pochampally, Anish Das Sarma, Xin Luna Dong, Alexandra Meliou, and Divesh Srivastava. 2014. Fusing data with correlations. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 433–444.

[26] Md Tahmid Rashid and Dong Wang. 2020. CovidSens: a vision on reliable social sensing for COVID-19. *Artificial Intelligence Review* (2020), 1–25.

[27] Theodoros Rekatsinas, Xin Luna Dong, and Divesh Srivastava. 2014. Characterizing and selecting fresh data sources. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 919–930.

[28] Lloyd S Shapley. 1953. *Additive and non-additive set functions*. Princeton University.

[29] Lloyd S Shapley and Martin Shubik. 1954. A method for evaluating the distribution of power in a committee system. *American political science review* 48, 03 (1954), 787–792.

[30] Oskar Skibski, Tomasz P Michalak, and Talal Rahwan. 2018. Axiomatic characterization of game-theoretic centrality. *Journal of Artificial Intelligence Research* 62 (2018), 33–68.

[31] Philippe Smets. 1998. *The Transferable Belief Model for Quantified Belief Representation*. Springer Netherlands, Dordrecht, 267–301. https://doi.org/10.1007/978-94-017-1735-9_9

[32] Julian Stier, Gabriele Gianini, Michael Granitzer, and Konstantin Ziegler. 2018. Analysing Neural Network Topologies: a Game Theoretic Approach. *Procedia Computer Science* 126 (2018), 234 – 243. https://doi.org/10.1016/j.procs.2018.07.257 Knowledge-Based and Intelligent Information and Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia.

[33] Md Yusuf S Uddin, Md Tanvir Al Amin, Hieu Le, Tarek Abdelzaher, Boleslaw Szymanski, and Tommy Nguyen. 2012. On diversifying source selection in social sensing. In *2012 Ninth International Conference on Networked Sensing (INSS)*. IEEE, 1–8.

[34] Dong Wang, Md Tanvir Amin, Shen Li, Tarek Abdelzaher, Lance Kaplan, Siyu Gu, Chenji Pan, Hengchang Liu, Charu C Aggarwal, Raghu Ganti, et al. 2014. Using humans as sensors: an estimation-theoretic perspective. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, 35–46.

[35] Dong Wang, Boleslaw K Szymanski, Tarek Abdelzaher, Heng Ji, and Lance Kaplan. 2019. The age of social sensing. *Computer* 52, 1 (2019), 36–45.