Probabilistic Knowledge Infusion through Symbolic Features for Context-Aware Activity Recognition

Luca Arrotta[a,*], Gabriele Civitarese[a], Claudio Bettini[a]

[a] *Università degli studi di Milano, Dipartimento di Informatica, Via Celoria, 18, 20133, Milan, Italy*

Abstract

In the general machine learning domain, solutions based on the integration of deep learning models with knowledge-based approaches are emerging. Indeed, such hybrid systems have the advantage of improving the recognition rate and the model's interpretability. At the same time, they require a significantly reduced amount of labeled data to reliably train the model. However, these techniques have been poorly explored in the sensor-based Human Activity Recognition (HAR) domain. The common-sense knowledge about activity execution can potentially improve purely data-driven approaches. While a few knowledge infusion approaches have been proposed for HAR, they rely on rigid logic formalisms that do not take into account uncertainty. In this paper, we propose P-NIMBUS, a novel knowledge infusion approach for sensor-based HAR that relies on probabilistic reasoning. A probabilistic ontology is in charge of computing symbolic features that are combined with the features automatically extracted by a CNN model from raw sensor data and high-level context data. In particular, the symbolic features encode probabilistic common-sense knowledge about the activities consistent with the user's surrounding context. These features are infused within the model before the classification layer. We experimentally evaluated P-NIMBUS on a HAR dataset of mobile devices sensor data that includes 14 different activities performed by 25 users. Our results show

*Corresponding author

Email addresses: `luca.arrotta@unimi.it` (Luca Arrotta), `gabriele.civitarese@unimi.it` (Gabriele Civitarese), `claudio.bettini@unimi.it` (Claudio Bettini)

that P-NIMBUS outperforms state-of-the-art neuro-symbolic approaches, with the advantage of requiring a limited amount of training data to reach satisfying recognition rates (i.e., more than 80% of F1-score with only 20% of labeled data).

Keywords: human activity recognition, neuro-symbolic, context-awareness

---

1. Introduction

Sensor-based Human Activity Recognition (HAR) is a research area that has been investigated in the last decade by many research groups due to its many applications, ranging from service personalization to healthcare and well-being [1]. Most of the existing approaches are based on supervised deep learning solutions [2, 3].

However, several open issues still limit the deployment of HAR methods in real-world scenarios. A major issue is the need for large labeled training data sets to build reliable recognition models. Moreover, the decisions of deep learning models are poorly interpretable. The integration of common-sense knowledge in sensor-based HAR approaches has the potential of mitigating the above-mentioned issues [4]. Indeed, the human knowledge about the relationships between human activities and the users' context (e.g., running is an activity that is usually performed outdoors but less likely on rainy days) has the potential to significantly improve standard approaches only based on machine learning.

Recently, Neuro-Symbolic AI (NeSy) methods are emerging. NeSy approaches enhance the capabilities of deep learning models with traditional symbolic AI approaches [5]. In these approaches, a symbolic module (often designed by domain experts through human knowledge) is embedded in data-driven classification to reduce the amount of necessary labeled data. Since NeSy deep models also rely on symbolic representations of the application domain, their decisions become inherently more transparent for humans.

Among the NeSy approaches proposed in the literature, knowledge infusion is particularly promising. Specifically, this technique consists of infusing external knowledge (e.g., from knowledge graphs) into the data-driven component of a system [6]. Knowledge infusion has mainly been experimented for NLP applications with promising results [7, 8].

In this work, we propose NIMBUS: a novel framework based on knowledge infusion for sensor-based HAR. NIMBUS combines, before the classification layer, the features automatically extracted from raw sensor data and high-level context data with the ones inferred by a context-aware symbolic reasoner. Such *symbolic features* encode common-sense knowledge about the consistency of activities with the user's surrounding context. For the sake of this work, we apply NIMBUS to sensor-based HAR based on mobile devices (e.g., smartphones, smartwatches).

In previous work, we presented a method to generate symbolic features based on ontological reasoning [9]. Even though our previous work exhibited promising results, it relies on a rigid formalism that does not take into account uncertainty.

As a novel technical contribution, we propose a new symbolic reasoning strategy of NIMBUS that is called P-NIMBUS: a Probabilistic method for *kNowledge InfuSion through syMBolic featUreS*. Differently from our previous work, P-NIMBUS is based on an ontology encoding probabilistic relationships between contexts and activities.

Our experiments on a dataset of 25 users performing 14 different activities show that P-NIMBUS outperforms other approaches based on deterministic ontologies and a state-of-the-art HAR probabilistic neuro-symbolic solution. Besides, we experimentally show that NIMBUS is significantly effective in labeled data scarcity scenarios.

The contributions of this paper are three-fold:

- We propose NIMBUS, a framework based on knowledge infusion for sensor-based HAR.
- We present different symbolic reasoning strategies that can be used by

NIMBUS. In particular, the novel contribution of this paper is P-NIMBUS which is based on probabilistic reasoning.

- We experimentally show the superiority of P-NIMBUS compared to alternative approaches.

2. Related work

2.1. Neuro-symbolic AI and knowledge infusion

The goal of Neuro-Symbolic AI (NeSy) is to combine the benefits of data-driven and knowledge-based AI approaches [5]. In recent years, data-driven deep learning models have been applied to several domain applications due to their ability to automatically extract meaningful features from raw data. Another benefit of data-driven approaches is their robustness with respect to data uncertainty. However, huge amounts of labeled training data are required to reliable train such models. Moreover, humans struggle to understand the rationale behind the models' outputs due to their lack of transparency.

On the other hand, knowledge-based approaches (e.g., reasoning methods based on formal logic) do not require training data since they are typically developed through an explicit symbolic representation of the human domain knowledge [10]. This makes such approaches also inherently more transparent for humans. However, complex domains like HAR would require very complicated rule-based models to cover all the possible scenarios that should be handled in real-world applications. Hence, this rigidity significantly limits the scalability of symbolic approaches.

The combination of data-driven and knowledge-based approaches through NeSy leads to different potential benefits [5, 11, 12] such as making the decisions of the model more interpretable for humans and reducing the amounts of training data required during the learning process.

Knowledge Infusion is a promising NeSy technique that directly incorporates external domain knowledge into a deep learning model. With this approach, the model automatically acquires human domain knowledge during training. For instance, knowledge-based constraints can be used to guide the training process of a generative model [13]. In [14], a specifically designed *Knowledge Infusion Layer* is used to combine the latent features extracted by an LSTM with external knowledge provided by knowledge graphs. Overall, knowledge infusion has been mainly consider in the *Natural Language Processing* (NLP) [7, 8] and the computer vision [15] domains. Solutions based on reinforcement learning can also benefit from knowledge infusion. For instance, a set of knowledge-based functions is used in [12] to initially guide the decisions of an agent that has little experience with its surrounding environment.

2.2. Neuro-symbolic AI methods for HAR

Recently, the sensor-based recognition of human activities has been mainly tackled through purely data-driven deep learning solutions [2, 3]. Most of the existing NeSy approaches for HAR have been proposed for environmental sensors in smart-home environments without considering the inertial measurements provided by the user's mobile devices, which is the focus of this work. Moreover, such methods take advantage of domain knowledge only before or after the training process of their data-driven component [16, 17], without considering any infusion technique. For instance, in [18], frequent patterns extracted from an unlabeled dataset through data mining techniques are associated with the corresponding user's activity based on domain knowledge. On the other hand, knowledge-driven reasoning is used in [19] to derive an initial activity model that is then adapted to the user's habits in a data-driven fashion.

Considering mobile devices' data, information about the context surrounding the user (e.g., her semantic location, local weather conditions) can be used to expand the set of activities that can be detected. Moreover, context data would be helpful to distinguish those activities that present similar motion patterns but that typically take place in different context conditions (e.g., standing and standing on a bus) [4]. However, the acquisition of training data in all the possible context conditions in which an activity could occur makes the use of data-driven solutions for context-aware HAR even more challenging. Hence, NeSy approaches that rely on formal knowledge-based models representing the relationships between context situations and activities could be a promising way to mitigate this issue.

Among the different approaches proposed in the literature [20], ontologies are the most common solution used to formally represent context data because of their automatic reasoning capabilities [21, 22, 23]. In this paper, we consider the ontologies originally proposed in previous works [4, 24].

The combination of data-driven approaches with ontological context reasoning has already been investigated in the literature. A machine learning model that detects low-level activities is integrated in [25] with an ontological reasoning process in charge of deriving higher-level activities based on the detected activities and other context information (e.g., semantic location and mood). Other NeSy approaches rely on context reasoning to refine the predictions of a data-driven activity classifier [16, 4]. However, these approaches rely on a rigid ontological formalism that cannot handle the probabilistic nature of the relationships between the user's context and activities.

In the HAR domain, few works focused on logic formalisms that support uncertainty reasoning. For instance, Markov Logic Networks (MLNs) [26] combine logic with probability theory to model knowledge-based hard and soft constraints. In order to handle uncertainty reasoning, different weights can be associated with such constraints. Unfortunately, MLNs methods have been proposed only for activity recognition in smart-home environments, based on environmental sensors data [27, 28, 10].

The only application of probabilistic context reasoning for context-aware HAR based on mobile devices is ProCAVIAR [24]. ProCAVIAR refines the probability distribution generated by the activity classifier using probabilistic reasoning. However, ProCAVIAR does not infuse the domain knowledge into the activity classifier to guide its learning process, thus limiting the potential benefits of NeSy approaches.

For this reason, we recently proposed DUSTIN [9] that uses ontological reasoning to derive domain-based symbolic features that are directly infused into a deep neural network. One of the problems of DUSTIN is that it relies on a too-rigid formalism rather than on probabilistic reasoning.

To the best of our knowledge, P-NIMBUS is the first knowledge infusion strategy for context-aware HAR that infuses domain knowledge into a deep neural network by relying on probabilistic context reasoning.

3. Knowledge infusion through symbolic features

In this section, we present NIMBUS: a framework for *kNowledge Infusion through syMBolic featUreS* specifically designed for context-aware sensor-based HAR on mobile devices. In this paper, with *context*, we mainly indicate the information about the environment which surrounds the user (e.g., current semantic location, the fact that she is indoors or outdoors, proximity to transportation routes, the current weather).

The rationale behind NIMBUS is that domain knowledge about HAR (e.g., cycling is more likely performed outdoors) can be exploited to drive the learning process of a purely data-driven activity classifier. In this work, we consider a knowledge model expressing relationships between high-level context data obtained by mobile devices (e.g., semantic location, weather) and the possible activities.

Given the current context C of the user, NIMBUS uses a symbolic reasoner to determine the set of activities that are *consistent* with C and, if possible, their degree of consistency. The information derived by the symbolic reasoner is then translated into a vector of *symbolic features*. Then, symbolic features are infused into the deep learning HAR model.

Besides improving the recognition rate, the infusion of domain knowledge can potentially reduce the amount of training data and epochs that would be necessary to learn such domain constraints in a purely data-driven fashion. At the same time, the knowledge used to train the activity classifier can also be used to partially understand the rationale behind its predictions, thus increasing the interpretability of the deep learning model.

Figure 1 depicts the overall architecture of NIMBUS. The user's mobile

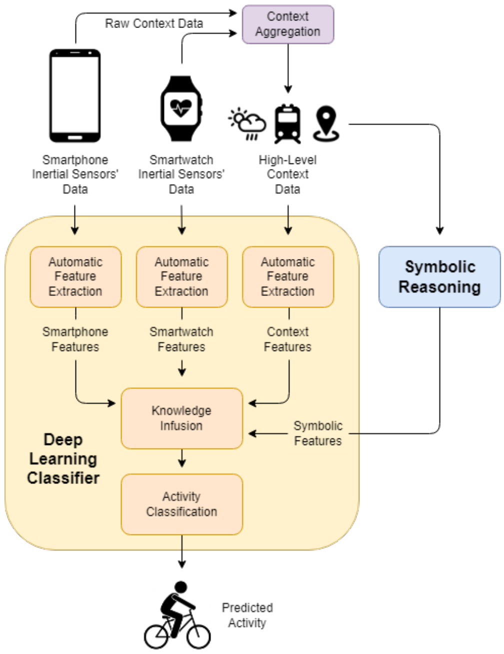


Figure 1: Overall architecture of NIMBUS: a knowledge infusion method for HAR based on symbolic features

devices (e.g., smartphone, smartwatch) generate two streams of raw sensor data: context and inertial sensor data. Raw context data (e.g., GPS) are provided to the Context Aggregation module. This module is in charge of generating high-level context data that better describe the user's surrounding context. For instance, given GPS data, this module interacts with a dedicated web service to obtain the semantic position of the user (e.g., park). High-level context data and raw inertial sensors data are provided as input to the Deep Learning Classifier. At the same time, high-level context data are also processed by the Symbolic Reasoning module. This module is in charge of inferring, for each activity, its degree of consistency considering the context that surrounds the user, according to common-sense knowledge about the HAR domain. For instance, the *walking* activity is consistent only when the user has a positive speed. The output of the Symbolic Reasoning module is encoded through a vector of symbolic features. Section 4 will present three alternative strategies to realize the Symbolic Reasoning module and to generate *symbolic features* vectors.

Finally, *symbolic features* are infused inside the Deep Learning Classifier. Note that raw inertial sensors data and high-level context data are first processed by *Automatic Feature Extraction* layers to automatically extract meaningful features. Then, automatically extracted features and *symbolic features* are provided to the *Knowledge Infusion* layers. These layers firstly combine, in the latent space, the features automatically extracted by the Automatic Feature Extraction layers with the ones inferred through the Symbolic Reasoning module. Then, the combined features are provided to a sequence of fully-connected layers in charge of learning correlations between input data and context-consistent activities. Finally, the Activity Classification layers output the activity currently performed by the user.

4. Symbolic reasoning approaches

In this section, we present three alternative symbolic reasoning methodologies that we designed to infer the symbolic features that are infused in the activity classifier, as explained in Section 3. Overall, the Symbolic Reasoning module (running locally on one of the user's mobile devices) analyzes the user's surrounding context to compute symbolic features encoding information about context-consistent activities. This module relies on ontological reasoning based on domain relationships between high-level context data and activities. In this work, we considered ontologies proposed in previous works [4, 24]. Specifically, each ontology models relationships between activities and high-level context data. We consider different context categories: the semantic place of the user, her presence in indoor or outdoor locations, her speed, her proximity to public transportation routes, her height variations, as well as weather conditions, environmental noise and light levels, and temporal context (e.g., the day of the week, the month, the season).

Periodically, the information related to the current context surrounding the user is automatically translated into ontological facts and included in the ontology. Most of the context data we considered present a one-to-one mapping with ontological facts. For instance, information about current weather conditions (e.g., it is sunny) provided by a public web service is automatically mapped to the corresponding ontological concept *Sunny(weather)*. On the other hand, scalar values about raw context data must be discretized by the CONTEXT AGGREGATION module before being mapped to the corresponding ontological facts. For instance, height variation measurements provided by the smartphone's barometer can be mapped to the following ontological concepts: *NegativeHeightVariation*, *NullHeightVariation*, and *PositiveHeightVariation*. Based on domain knowledge, the ontology explicitly states the context conditions where activities most likely occur. For instance, the activity *going downstairs* should take place while the height variation of the user is negative and her speed is positive. Given the current context conditions, the SYMBOLIC REASONING module uses the ontology to infer, for each activity, a confidence value about its consistency with respect to the context.

The confidence value inferred for each activity is used by NIMBUS to generate the vector of *symbolic features* that will be infused into the deep learning model. Each position of the vector encodes one of the activities that NIMBUS is able to detect. The specific values encoded in this vector depend on the adopted ontological reasoning approach.

Example 1. *Alice is using a system based on NIMBUS. In the beginning, Person(Alice) is added as a fact into the ontology.*[1] *The* CONTEXT AGGREGATION *module of NIMBUS derives that Alice is currently in a public park and that her speed is* 2 *km/h according to GPS data obtained from her smartphone. Such*

[1]Note that the ontology needs this axiom to associate contexts and activities to an instance of a person. It is important to note that the reasoning process does not depend on the specific user. Here we report *Person(Alice)* for the sake of clarity, but the actual symbolic reasoning module does not take into account the user's identity.

context data are automatically added to the ontology through two individuals: Park(place) and LowSpeed(speed). Then, NIMBUS adds facts about the relationships between Alice and the available context data: hasCurrentSymbolicLocation(Alice, place) and hasCurrentSpeed(Alice, speed). Finally, in order to infer the degree of context-consistency of the activity walking, the Symbolic Reasoning *module adds two other axioms to the ontology: Walking(currentActivity) and isPerforming(Alice, currentActivity). The context-consistency confidence of walking will be hence determined by the set of facts and the domain knowledge. This last step is repeated for each activity to generate symbolic feature vectors that will be infused into the data-driven classifier.*

In the following, we present and compare three different ontological reasoning strategies we designed: (1) S-NIMBUS: a solution that performs simple symbolical reasoning based on strict constraints, (2) O-NIMBUS: the one used in our previous work [9], and (3) P-NIMBUS: the novel technical contribution of this paper, a probabilistic version of O-NIMBUS. We specifically designed S-NIMBUS and P-NIMBUS for this work.

4.1. S-NIMBUS: ontology reasoning based on strict constraints

In the following, we present Strict-NIMBUS (S-NIMBUS for short): an approach based on a deterministic ontology that defines strict constraints imposing that activities may only occur in common context scenarios. Indeed, S-NIMBUS aims to increase the recognition rate by considering context constraints that are likely satisfied in the majority of the situations.

In particular, S-NIMBUS is based on the ontology presented in [4]. The knowledge model of S-NIMBUS only considers the most common context situations, while unusual context conditions are automatically excluded. For instance, S-NIMBUS models the *running* activity by imposing that it can take place only when the user is outdoors with a positive speed. Hence, *running* is context-consistent only when performed on these specific context conditions. In less common scenarios (e.g., the subject is running inside a gym), this method would fail, considering *running* as context-inconsistent.

S-NIMBUS generates the symbolic features vector by associating the value 1 with the context-consistent activities, 0 otherwise.

4.2. O-NIMBUS: ontology reasoning based on open constraints

In order to mitigate the limitations of S-NIMBUS, we propose Open-NIMBUS (O-NIMBUS for short): an approach based on a deterministic ontology with weaker constraints on the contexts that should be verified when an activity is performed. The use of weaker constraints allows the ontology to consider as consistent both common and uncommon context situations.

For instance, consider the *running* activity. In S-NIMBUS, this activity can only be performed outdoors with a positive speed. In O-NIMBUS, instead, this activity is consistent even when performed indoors (e.g., in the gym). Nonetheless, O-NIMBUS imposes that the user should not be still since this constraint should generally be true for *running*. Hence, to be consistent according to O-NIMBUS, *running* has only to be performed with positive speed.

Note that, despite considering only the most common contexts as possible for an activity, S-NIMBUSmay reach overall better performances based on the dataset being considered. Indeed, O-NIMBUS considers as consistent also those contexts that can rarely occur. This could have a negative impact on helping the classifier to recognize activities in the majority of the situations.

Similarly to S-NIMBUS, O-NIMBUS generates the symbolic features vector as follows: the value of a symbolic feature is 1 if the corresponding activity is context-consistent, 0 otherwise. Note that this strategy is exactly the one we previously presented (named DUSTIN) in [9].

4.3. P-NIMBUS: probabilistic ontology reasoning

The main novelty of this work is the Probabilistic-NIMBUS (P-NIMBUS for short) strategy that relies on probabilistic reasoning to take into account the intrinsic uncertainty that characterizes the relationships between the activities performed by the user and the surrounding context. In this work, we use the probabilistic ontology originally proposed in [24].

This approach considers both *hard* and *soft* constraints. Hard constraints capture those conditions that must always be satisfied to consider an activity as context-consistent. In particular, in P-NIMBUS the hard constraints are exactly the ones that are enforced in O-NIMBUS.

On the other hand, soft constraints encode probability distributions over context conditions that may change for each activity. Continuing the *running* example, even though this activity is typically performed outdoors, a person could also run indoors (e.g., in a gym) even if less likely. If outdoors, the more likely user's semantic location would be the park. Moreover, if the user is running outdoors, it is less likely that this will happen during bad weather.

Each soft constraint is associated with a degree of confidence. For example, the confidence value of the soft constraint *running can be performed indoors* should be lower than the confidence of the soft constraint *running can be performed outdoors*. Note that hard constraints are associated with 1 as a probability value.

P-NIMBUS generates a symbolic feature vector by associating the value 0 if one or more hard constraints of the corresponding activity are false. Otherwise, P-NIMBUS computes a probability value considering all hard and soft constraints that are true. Section 5.3 explains the specific strategy we adopted to generate such probabilities. The symbolic features vector is finally normalized.

5. Experimental evaluation

In this section, we describe the experimental evaluation that we carried out to evaluate the different symbolic reasoning strategies of NIMBUS that we presented in Section 4. First, we describe the dataset that we used for our experiments. Then, we illustrate the details of our experimental setup. We provide details about the specific pipeline we adopted (e.g., pre-processing, network architecture) and describe our evaluation methodology.

5.1. Dataset description

We evaluate NIMBUS on a HAR dataset [4] that includes labeled sensor and context data obtained from the mobile devices of 25 subjects. Specifically, each subject wore a smartwatch on the dominant hand's wrist and a smartphone in the pants' front pocket. Both devices collected raw inertial sensor data (i.e., accelerometer, gyroscope, and magnetometer). Moreover, this dataset includes several types of context data. In particular, the environment's brightness and noise levels are measured through the luminosity sensor and the microphone, respectively. The barometer provides measurements of the subjects' height variations, while the GPS measures their current speed. The dataset also includes temporal context data, like the moment of the day (e.g., morning, afternoon), the day of the week, the month, and the season. Finally, higher-level context data is obtained by combining public web services and data from the smartphone's built-in sensors. *Google's Places API* is used to derive the semantic places closest to the subjects (e.g., university, restaurants); *OpenWeatherMap* for the current local weather conditions (e.g., sunny); *Bing's Traffic API* for information about the nearby traffic situation, and *Transitland* for the public transportation routes and stops closest to the subjects.

Overall, the dataset includes 14 context-dependent activities: *brushing teeth*, *cycling*, *elevator down*, *elevator up*, *lying*, *moving by car*, *running*, *sitting*, *sitting on transport*, *stairs down*, *stairs up*, *standing*, *standing on transport*, and *walking*. The total number of activity instances is ≈ 350, covering almost 9 hours of labeled data.

However, data collection was performed in a scripted fashion, and the variety of context data that was actually collected is limited. For instance, the *running* activity was only collected outdoors, even if it could be performed in indoor locations like a gym. For this reason, based on collected context data and common-sense knowledge about the HAR domain, for each activity, we synthetically generated simulated context scenarios that partially replaced the context data of the original dataset. For instance, in the enhanced version of the dataset, *running* mainly takes place outdoors (90%), while, in a few cases, also in indoor environments (10%).

5.2. Evaluation methodology

We evaluated the three ontological reasoning strategies presented in Section 4 through the leave-one-subject-out cross-validation technique. At each fold, the test set consists of the data of a subject of the dataset, while all the other data are used for the training (90%) and the validation (10%) sets. We considered a maximum of 200 training *epochs* with *batches* containing 32 samples and adopting the Adam optimizer. We used the *early stopping* strategy to stop the learning process when the *validation loss* did not improve for 5 consecutive epochs. The results presented in the next section are computed by averaging the recognition rates obtained on the test set of each fold.

In our experiments, we will consider as *Baseline* the neural network of NIMBUS without symbolic reasoning and knowledge infusion. Hence, *Baseline* only considers context data as an input processed by the *Automatic Feature Extraction* module. Moreover, we also compare NIMBUS with a state-of-the-art probabilistic context refinement method that is called ProCAVIAR [24]. This method uses a probabilistic logic to refine the output probability distribution of the classifier. Hence, while this method also relies on probabilistic semantic reasoning, the knowledge is not infused directly in the deep learning model.

5.3. Experimental setup

In the following, we describe the specific setup of our experiments.

We segmented inertial and high-level context data into non-overlapping and fixed-size windows. Based on previous works on this dataset [4], we considered segmentation windows of 4 seconds to detect both simple (e.g., *standing*) and complex (e.g., *brushing teeth*) activities. The input shape of the *Automatic Feature Extraction* modules for inertial sensors data (for both the smartphone and the smartwatch) is $(9, 400)$[2]. On the other hand, high-level context data

[2] Each row corresponds to one axis of a tri-axial inertial sensor (i.e., accelerometer, gyro-

are represented in each segment with one-hot encoding. In order to reduce the intrinsic noise of inertial sensor data, we applied a median filter.

Our network architecture has been determined empirically, by choosing a simple structure that guaranteed good recognition rates. Even though more sophisticated deep learning models have been proposed for HAR, we select a simple solution to focus on the impact of knowledge infusion. In particular, each Automatic Feature Extraction that processes inertial data consist of two *convolutional* layers with 8 3×3 and 64 2×2 filters, respectively, separated by a 2×2 *max pooling* layer. The second *convolutional* layer is followed by another 2×2 *max pooling* layer, a *flatten* layer, and a *fully connected* layer with 128 neurons. On the other hand, the Automatic Feature Extraction focused on high-level context data is composed of a single *fully connected* layer with 8 neurons. The Knowledge Infusion module consists of a *concatenation* layer that combines inertial, context, and symbolic features. The combined features are then provided to a *dropout* layer with a dropout rate of 0.1 and a *fully connected* layer with 256 neurons. Finally, in the Activity Classification module, a *fully connected* layer based on the *softmax* activation function is used to obtain the final classification.

The symbolic reasoning strategy relied on the ontologies presented in previous works (i.e., S-NIMBUS on [4], O-NIMBUS on [9], and P-NIMBUS on [24]). All of these ontologies are defined through the OWL2 ontology language. The symbolic reasoner used for S-NIMBUS and O-NIMBUS is Pellet [29]. On the other hand, the symbolic reasoner used for P-NIMBUS is ELOG [30], which is based on a log-linear probabilistic model. In order to use ELOG, the axioms of the ontology should be properly annotated to specify a weight for each soft constraint.

scope, and magnetometer), and includes 400 samples, considering a sampling rate of 100 Hz and windows of 4 seconds.

5.4. Results

In the following, we present the results we obtained while evaluating the effectiveness of NIMBUS on the dataset described in Section 5.1.

5.4.1. Comparing symbolic reasoning strategies

Figure 2 compares the overall F1-score of the *Baseline* with the three ontological reasoning strategies described in Section 4. Overall, the infusion of

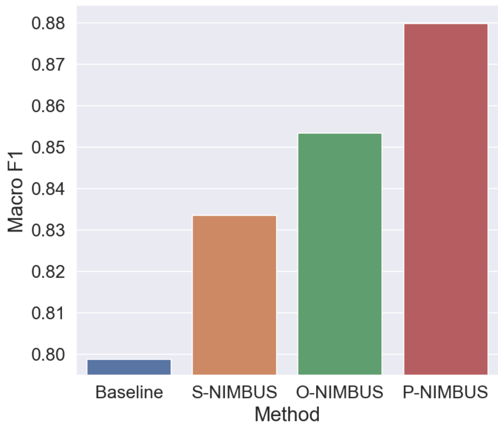


Figure 2: Overall results

symbolic features dramatically improves the recognition rates of the activity classifier. In particular, we observed that P-NIMBUS is the most effective strategy ($\approx +8\%$ compared to the *Baseline*). Indeed, probabilistic symbolic features reduce the well-known rigidity of the approaches based on deterministic reasoning.

Consistently, the S-NIMBUS approach is the least effective symbolic reasoning strategy (only $\approx +3\%$ compared to the *Baseline*). This is because this simple reasoning approach only considers the most likely scenarios, determining that activities performed in unlikely contexts (e.g., lying outdoors) are not consistent. While the O-NIMBUS approach slightly mitigates this problem ($\approx +5\%$ compared to the *Baseline*), it is still less effective than P-NIMBUS since it does not take into account uncertainty.

Figure 3 compares (in terms of F1-score) P-NIMBUS with ProCAVIAR [24]. These results suggest that P-NIMBUS outperforms ProCAVIAR, indicating

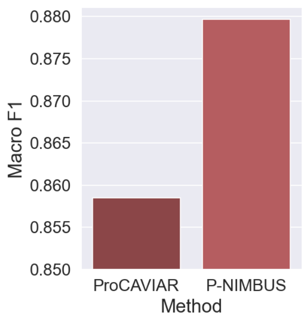


Figure 3: Comparison between P-NIMBUS and ProCAVIAR

that directly infusing probabilistic domain knowledge into the classifier is a more promising and effective approach.

Figure 4 depicts the confusion matrices obtained by the *Baseline* and the different symbolic reasoning strategies of NIMBUS. The results show that the methods based on symbolic features (especially P-NIMBUS) outperform the *Baseline* on different activities. Indeed, learning constraints directly from high-level context data is more complex than relying on symbolic features derived through symbolic reasoning. For instance, the *Baseline* sometimes confuses *elevator down* with *brushing teeth* even if *brushing teeth* does not occur with user's negative height variations. On the other hand, in these scenarios, the *symbolic features* of NIMBUS encodes *brushing teeth* as context-inconsistent, thus improving the recognition of *elevator down*.

Another interesting example is the *lying* activity, which is often miss-classified by the *Baseline*. In this case, S-NIMBUS improves the recognition of *lying* only slightly since this approach considers such an activity as context-consistent only when performed indoors. For this reason, the weaker constraints of O-NIMBUS dramatically improve the recognition rate of *lying* since it is assumed that this activity can also take place outdoors. P-NIMBUS further improves the recognition of *lying* thanks to soft constraints (i.e., this activity is possible outdoors but with a low probability).

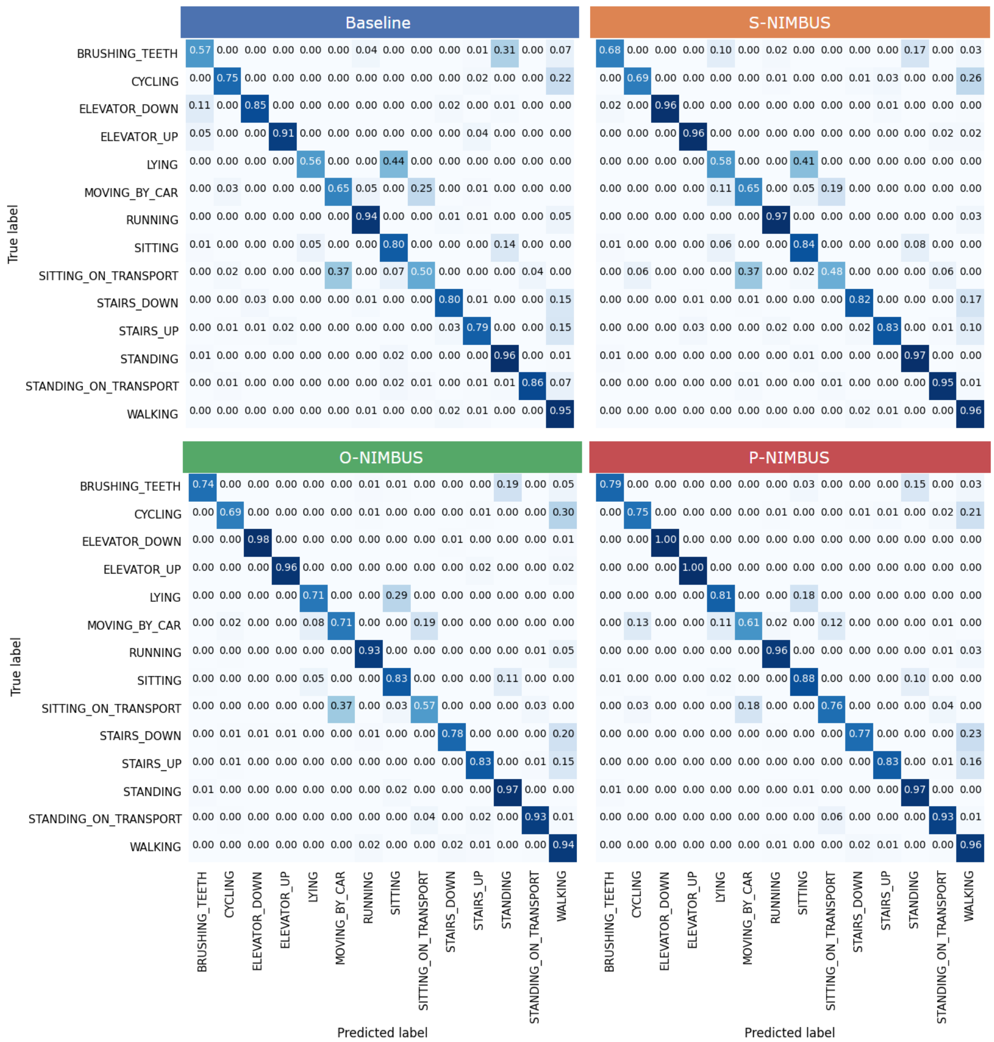


Figure 4: Confusion matrices comparison

5.4.2. Results with low labeled data availability

Figure 5 compares the different symbolic reasoning strategies of NIMBUS while varying the percentage of available labeled training data. We observed that P-NIMBUS significantly outperforms the other approaches with limited amounts of labeled data. For instance, when only 10% of training data is used, P-NIMBUS dramatically outperforms the *Baseline* by $\approx +30\%$, reaching an overall F1-score value of $\approx 76\%$. In this data scarcity scenario, also S-NIMBUS and O-NIMBUS outperform the *Baseline* in a similar way. However, only using hard constraints to derive the symbolic features reduces the benefits produced

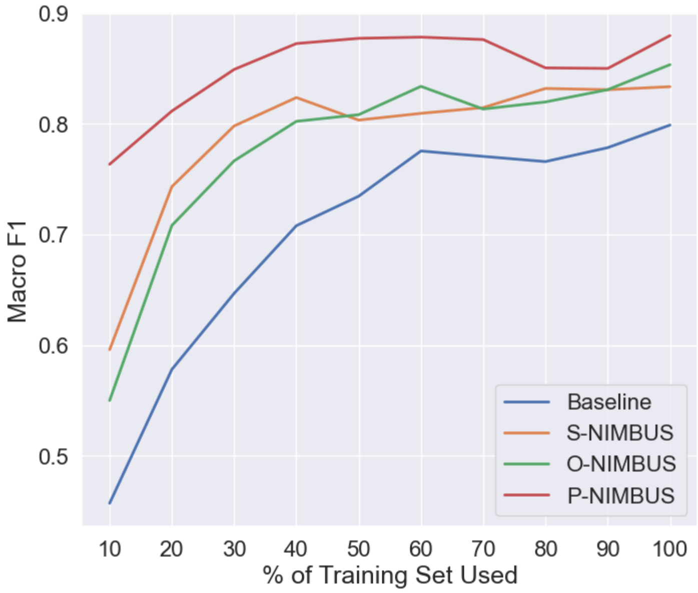


Figure 5: Overall performance while varying the available training data

by domain knowledge infusion on limited data availability.

Figure 6 provides activity-level recognition rates considering only 10% of training data. These results show how probabilistic reasoning increases the recognition rate of each considered activity. For instance, considering *elevator up/down* and *stairs up/down*, even if it should be easy to detect such activities based on the user's height variations, the *Baseline* struggles in recognizing them with a few training samples. This problem also affects S-NIMBUS and O-NIMBUS since height variations may also occur while performing other activities (e.g., *walking*). Since P-NIMBUS considers probabilistic constraints, it relies on the fact that height variations are less common when *walking* compared to taking the elevator or the stairs, thus significantly improving the recognition rate. Another example where probabilistic reasoning is particularly effective is with the *standing on transport* and *sitting on transport* activities. Indeed, even though it is possible to have a positive speed near transportation routes with several activities (e.g., *cycling*, *moving by car*, *running*), the probability of these events is lower.

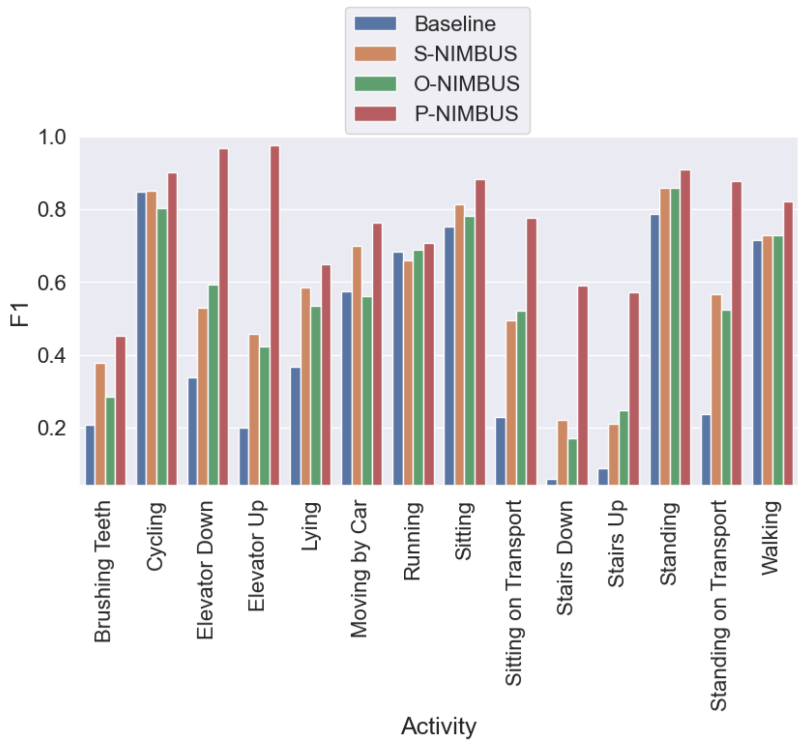


Figure 6: Activities results with 10% of training data

5.4.3. Impact on the number of epochs

Figure 7 compares the evolution of the F1-score during the learning process (i.e., at each training epoch) of the four different approaches. As previously mentioned, the training process is stopped based on the validation loss. Our

Figure 7: F1-score trend during training

results show that probabilistic reasoning is more complex than the ones of S-NIMBUS and O-NIMBUS. Thus, the network requires more epochs to converge (but still less than the *Baseline*). Nonetheless, P-NIMBUS reaches significantly high recognition rates even at the first epoch ($\approx 90\%$ of F1-score).

In general, compared to the *Baseline*, the results show that our approaches based on symbolic features significantly speed up the convergence of the learning process compared to the *Baseline*. This is due to the fact that constraints are quickly learned by the network, thus stabilizing the validation loss more rapidly.

Hence, these results suggest that knowledge infusion can potentially reduce the training effort of the classifier and, at the same time, reach higher recognition rates.

5.4.4. Impact of context data

Finally, we investigated the general impact of context data on the recognition rate. For this reason, we only considered the *Baseline* to exclude the influence of symbolic features on the recognition rate. Figure 8 compares the performance of our *Baseline* with a modified version that only considers inertial sensors (called *Inertial Only*). This result confirms that context data

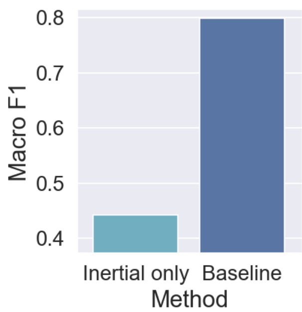
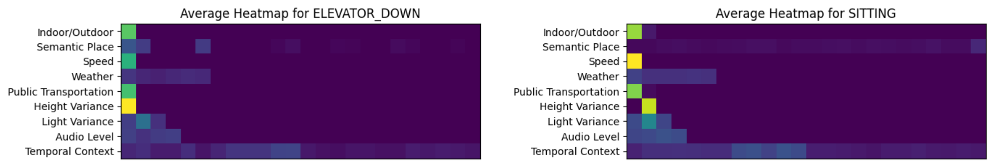


Figure 8: Impact of context data on the recognition rates of the activity classifier

dramatically improves the F1 score obtained by the *Baseline* ($\approx +35\%$ compared to *inertial only*). For instance, the user's speed and proximity to public transportation routes are essential to distinguish activities like *standing* and *standing on transport*.

Moreover, we also inspected the impact of each specific context information



(a) Elevator Down

(b) Sitting

Figure 9: Average relevance values obtained by RISE on a subset of the activity classes included in the considered dataset

on the recognition rate. Specifically, we considered an eXplainable AI (XAI) method, based on model induction, that generates explanations about the output of the activity classifier. Model induction XAI methods generate perturbed versions of an input sample to analyze how the different input features impact the classifier's output. This approach makes it possible to derive the most important input features for classification. LIME [31] is one of the most promising model induction methods. However, the process it uses to generate the input perturbations is unsuitable for one-hot encoded data (like context data in NIMBUS). Hence, we decided to rely on the model induction method RISE [32]. As explanations, RISE produces a relevance value for each context information, where the higher the relevance, the more this context information was important for the classifier's decision. We applied RISE on the correct predictions made by the *Baseline* model on the test set at each iteration of the cross-validation process. Then, for each dataset activity, we averaged the relevance values. Figure 9 shows the average relevance values obtained through RISE on two activity classes of the considered dataset. Note that, for the sake of visualization, we reshaped the one-hot encoded vectors of high-level context data into a matrix where each row represents a different type of context information. For instance, the first row of the matrix encodes information about the presence of the user in indoor or outdoor environments, while the third row codifies the speed of the user: the first four pixels of this row respectively represent *null*, *low*, *medium*, and *high* speeds.

Figure 9a depicts, for instance, the average heatmap related to the *elevator* *down* activity. As expected, the highest relevance value is associated with the negative height variations of the user. At the same time, also the user's presence in indoor environments, a null speed, and the absence of public transportation routes and stops were relatively important to distinguish this activity from the others. However, other context data like light variance, audio level, and temporal context are not useful for classifying this activity.

On the other hand, considering *sitting* (Figure 9b), the most discriminating context information is the null speed of the user. Indeed, this information is crucial, for instance, to distinguish *sitting* from *sitting on transport*.

6. Discussion

6.1. Context data collection

In this work, we assume that context data can be continuously collected and that they are constantly available. However, considering real-world scenarios, this assumption is not completely realistic.

Indeed, in order to be collected, several high-level context data (e.g., semantic location) require interaction with external web services. Continuous network communication may negatively impact the device's resources and latency (i.e., context information is not perfect in real-time).

However, it is important to point out that such high-level contexts do not change so rapidly, while activity recognition is continuously performed every few seconds (e.g., in our experiments, the segmentation window is 4 seconds). Hence, it is possible to design a strategy to obtain new information from web services with a low number of web service calls. For instance, considering semantic location, it is possible to perform a query only when GPS data exhibit significant changes. As another example, the weather web service could be queried with a low periodicity (e.g., every hour).

Thanks to these strategies, it is also possible to run our method when the mobile devices are not connected to the internet for short periods. However, if the mobile devices are offline for a long time period, the system would consider a limited amount of context information, possibly impacting the recognition rate.

In future work, we will investigate in detail such practical aspects, also considering new strategies to adapt the model based on the Quality of Service.

6.2. Alternative knowledge infusion strategies

In this work, we proposed a technique to infuse HAR knowledge through symbolic features. Our results show that our method outperforms ProCAVIAR [24], a state-of-the-art approach based on probabilistic context refinement. However, similarly to ProCAVIAR, our approach requires to execute symbolic reasoning also during classification. This setting may not be suitable for real-world deployments on mobile devices due to the computational complexity of ontologies.

We are currently investigating alternative knowledge infusion approaches where the deep learning classifier learns the common-sense knowledge constraints without requiring symbolic reasoning during inference. For example, this could be achieved by designing a semantic loss function in charge of driving the learning process through knowledge. An alternative solution is to train a separate neural network that approximates the symbolic reasoner proposed in this work by mapping high-level context data to symbolic features, thus alleviating the ontological reasoning task.

6.3. Generalizability of the approach

In this work, we used a dataset with rich contextual information and a significant amount of activity labels. Unfortunately, we could not find other datasets in the HAR based on mobile devices with these characteristics. In future work, we are interested in exploring our method also in the context of complex activity recognition in smart-home environments.

However, we are also interested in understanding if our approach could also be applied in different domains. In general, it could be applied to domains where:

- a portion of input data does not directly reveal high-level context information (e.g., inertial sensors in our domain).
- a portion of input data reveals high-level context information (e.g., GPS in our domain).
- it is possible to use common-sense knowledge to define relationships between context and the classification task.

For instance, considering the autonomous driving domain, reasoning on high-level context data may help in improving the decisions made by analyzing the sensors equipped in the smart car. As another domain example, risk assessment and/or security applications may benefit from context reasoning to improve their decisions.

7. Conclusion and future work

In this paper, we presented NIMBUS, our knowledge infusion framework for sensor-based HAR. This neuro-symbolic approach relies on symbolic reasoning to compute symbolic features from the user's context data. Such features are infused in the latent layers of a deep neural network through their concatenation with the features automatically extracted by convolutional layers from raw inertial sensor and high-level context data. Our results indicate that the generation of symbolic features by using probabilistic reasoning leads to high recognition rates even in labeled data scarcity scenarios.

Even though our preliminary results are promising, NIMBUS still has several limitations that we will tackle in future work besides the ones discussed in Section 6. First, there are no comprehensive public ontologies for this domain, and the ontology design and implementation require significant work by knowledge engineers and domain experts. It is questionable if such a manual approach can generalize over all the possible context conditions and activities [33]. However, promising semi-automatic approaches exist to obtain such knowledge from external sources (e.g., text, videos, and images on the Web, as well as existing knowledge graphs).

Our future efforts will also focus on analyzing the interpretability of NIMBUS. Since the predictions of NIMBUS rely on common-sense knowledge, they are inherently more interpretable than fully data-driven approaches. Hence, we will study how to design user-based experiments to investigate this aspect.

Acknowledgement

This work was partially supported by the project "MUSA - Multilayered Urban Sustainability Action", NextGeneration EU, MUR PNRR.

References

[1] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, Z. Yu, Sensor-based activity recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42 (6) (2012) 790–808.

[2] J. Wang, Y. Chen, S. Hao, X. Peng, L. Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognition Letters 119 (2019) 3–11.

[3] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, Y. Liu, Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities, ACM Computing Surveys (CSUR) 54 (4) (2021) 1–40.

[4] C. Bettini, G. Civitarese, R. Presotto, Caviar: Context-driven active and incremental activity recognition, Knowledge-Based Systems 196 (2020) 105816.

[5] M. K. Sarker, L. Zhou, A. Eberhart, P. Hitzler, Neuro-symbolic artificial intelligence: Current trends, arXiv preprint arXiv:2105.05330 (2021).

[6] A. Sheth, M. Gaur, U. Kursuncu, R. Wickramarachchi, Shades of knowledge-infused learning for enhancing deep learning, IEEE Internet Computing 23 (6) (2019) 54–63.

[7] K. Annervaz, P. K. Nath, A. Dukkipati, Actknow: Active external knowledge infusion learning for question answering in low data regime, arXiv preprint arXiv:2112.09423 (2021).

[8] R. Wang, D. Tang, N. Duan, Z. Wei, X. Huang, G. Cao, D. Jiang, M. Zhou, et al., K-adapter: Infusing knowledge into pre-trained models with adapters, arXiv preprint arXiv:2002.01808 (2020).

[9] L. Arrotta, G. Civitarese, C. Bettini, Knowledge infusion for context-aware sensor-based human activity recognition, in: 2022 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2022, pp. 1–8.

[10] G. Civitarese, T. Sztyler, D. Riboni, C. Bettini, H. Stuckenschmidt, Polaris: Probabilistic and ontological activity recognition in smart-homes, IEEE Transactions on Knowledge and Data Engineering 33 (1) (2019) 209–223.

[11] M. Gaur, K. Faldu, A. Sheth, Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?, IEEE Internet Computing 25 (1) (2021) 51–59.

[12] J. Fürst, M. F. Argerich, B. Cheng, E. Kovacs, Towards knowledge infusion for robust and transferable machine learning in iot, Open Journal of Internet of Things (OJIOT) 6 (1) (2020) 24–34.

[13] Z. Hu, Z. Yang, R. R. Salakhutdinov, L. Qin, X. Liang, H. Dong, E. P. Xing, Deep generative models with learnable knowledge constraints, Advances in Neural Information Processing Systems 31 (2018).

[14] U. Kursuncu, M. Gaur, A. Sheth, Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning, arXiv preprint arXiv:1912.00512 (2019).

[15] G. Chhablani, A. Sharma, H. Pandey, T. Dash, Superpixel-based domain-knowledge infusion in computer vision, arXiv preprint arXiv:2105.09448 (2021).

[16] D. Riboni, C. Bettini, Cosar: hybrid reasoning for context-aware activity recognition, Personal and Ubiquitous Computing 15 (3) (2011) 271–289.

[17] W. Van Woensel, P. C. Roy, S. S. R. Abidi, S. R. Abidi, Indoor location identification of patients for directing virtual care: An ai approach using machine learning and knowledge-based methods, Artificial Intelligence in Medicine 108 (2020) 101931.

[18] G. Azkune, A. Almeida, A scalable hybrid activity recognition approach for intelligent environments, IEEE Access 6 (2018) 41745–41759.

[19] A. S. A. Sukor, A. Zakaria, N. A. Rahim, L. M. Kamarudin, R. Setchi, H. Nishizaki, A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes, Journal of Intelligent & Fuzzy Systems 36 (5) (2019) 4177–4188.

[20] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, D. Riboni, A survey of context modelling and reasoning techniques, Pervasive and mobile computing 6 (2) (2010) 161–180.

[21] D. Riboni, C. Bettini, Owl 2 modeling and reasoning with complex human activities, Pervasive and Mobile Computing 7 (3) (2011) 379–395.

[22] F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, D. Nardi, et al., The description logic handbook: Theory, implementation and applications, Cambridge university press, 2003.

[23] L. Chen, C. Nugent, Ontology-based activity recognition in intelligent pervasive environments, International Journal of Web Information Systems (2009).

[24] C. Bettini, G. Civitarese, D. Giancane, R. Presotto, Procaviar: Hybrid data-driven and probabilistic knowledge-based activity recognition, IEEE Access 8 (2020) 146876–146886.

[25] O. Banos, C. Villalonga, J. Bang, T. Hur, D. Kang, S. Park, T. Huynh-The, V. Le-Ba, M. B. Amin, M. A. Razzaq, et al., Human behavior analysis by means of multimodal context mining, Sensors 16 (8) (2016) 1264.

[26] M. Richardson, P. Domingos, Markov logic networks, Machine learning 62 (1) (2006) 107–136.

[27] R. Helaoui, D. Riboni, H. Stuckenschmidt, A probabilistic ontological framework for the recognition of multilevel human activities, in: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, 2013, pp. 345–354.

[28] K. Gayathri, K. Easwarakumar, S. Elias, Probabilistic ontology based activity recognition in smart homes using markov logic network, Knowledge-Based Systems 121 (2017) 173–184.

[29] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, Y. Katz, Pellet: A practical owl-dl reasoner, Journal of Web Semantics 5 (2) (2007) 51–53.

[30] J. Noessner, M. Niepert, Elog: a probabilistic reasoner for owl el, in: International Conference on Web Reasoning and Rule Systems, Springer, 2011, pp. 281–286.

[31] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[32] V. Petsiuk, A. Das, K. Saenko, Rise: Randomized input sampling for explanation of black-box models, arXiv preprint arXiv:1806.07421 (2018).

[33] G. Civitarese, C. Bettini, T. Sztyler, D. Riboni, H. Stuckenschmidt, newnectar: Collaborative active learning for knowledge-based probabilistic activity recognition, Pervasive and Mobile Computing 56 (2019) 88–105.