Knowledge Infusion for Context-Aware Sensor-Based Human Activity Recognition

Luca Arrotta
Dept. of Computer Science
University of Milan
Milan, Italy
luca.arrotta@unimi.it

Gabriele Civitarese
Dept. of Computer Science
University of Milan
Milan, Italy
gabriele.civitarese@unimi.it

Claudio Bettini
Dept. of Computer Science
University of Milan
Milan, Italy
claudio.bettini@unimi.it

***Abstract*—Neuro-symbolic AI methods aim at integrating the capabilities of data-driven deep learning solutions with the ones of more traditional symbolic approaches. These techniques have been poorly explored in the sensor-based Human Activity Recognition (HAR) research field, even if they could lead to multiple benefits such as improving model interpretability and reducing the amount of labeled data that is necessary to reliably train the model. In this paper, we propose DUSTIN, a novel knowledge infusion approach for sensor-based HAR. DUSTIN concatenates the features automatically extracted by a CNN model from raw sensor data and high-level context data with the ones inferred by a knowledge-based reasoner. In particular, the symbolic features encode common-sense knowledge about the activities which are consistent with the context of the user, and they are infused within the model before the classification layer. We experimentally evaluated DUSTIN on a HAR dataset of mobile devices sensor data that includes 14 different activities performed by 26 users. Our results show that DUSTIN outperforms state-of-the-art neuro-symbolic approaches, with the advantage of requiring a limited amount of training data and training epochs to reach satisfying recognition rates.**

***Index Terms*—activity recognition, neuro-symbolic AI, knowledge infusion**

I. Introduction

Sensor-based Human Activity Recognition (HAR) is a well-established research area, thanks to its many applications ranging from healthcare to well-being [1]. The majority of the approaches that have been proposed in the literature are based on supervised deep learning solutions [2], [3].

Despite their undeniable achievements, purely data-driven models have several disadvantages. 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 data-driven 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 outdoor but less likely on rainy days) can significantly improve standard approaches only based on machine learning. However, to the best of our knowledge, no existing works proposed the integration of such knowledge in deep-learning HAR models.

Recently, neuro-symbolic AI methods have been proposed to enhance the capabilities of deep learning models with traditional symbolic AI approaches [5]. In these approaches, a symbolic module designed by domain experts through human knowledge is embedded in data-driven classification to reduce the amount of necessary labeled data. At the same time, a deep learning model whose decisions also rely on a reasoning module that operates on explicit symbolic representations of the application domain becomes inherently more transparent for humans.

Among the neuro-symbolic 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]. However, the effectiveness of such approaches for HAR is still an open research problem. Indeed, in existing neuro-symbolic solutions for HAR, external knowledge is only considered before [9] or after [10] the training process, and it is not infused into the deep learning model.

In this paper, we propose DUSTIN: a method for *knowleDge infUSion for human acTIvity recogNition*. DUSTIN 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. The considered symbolic features encode common-sense knowledge about the consistency of activities with the user's surrounding context. Our experiments on a dataset of 26 users performing 14 different activities show that DUSTIN outperforms existing state-of-the-art HAR neuro-symbolic solutions. Besides, we experimentally show that DUSTIN can be used to reduce the number of training epochs necessary to train an activity recognition classifier and to improve its recognition rates when a limited amount of labeled data is available.

The contributions of this paper are three-fold:

- We propose DUSTIN, a novel knowledge infusion method for HAR to improve the latent space representation of raw sensor data and high-level context data based on common-sense knowledge.
- Our experiments on a real HAR dataset show that DUSTIN outperforms state-of-the-art hybrid data-driven

2693-8340/22/$31.00 ©2022 IEEE

DOI 10.1109/SMARTCOMP55677.2022.00016

and knowledge-based HAR approaches.
- DUSTIN reduces the amount of labeled data and training epochs that are required to reach satisfying recognition rates, achieving an F1-score that is close to 90% when using only the 20% of the training set.

II. Related Work

A. Neuro-Symbolic AI and Knowledge Infusion

Neuro-symbolic AI aims at combining the strengths of *data-driven* and *knowledge-based* AI approaches [5]. The main advantages of data-driven approaches are their ability to automatically learn meaningful features from raw data and their robustness against uncertainty. However, these approaches require large amounts of training labeled data to build reliable models, and their opacity contrasts with the need for humans to understand the rationale behind each output. On the other hand, knowledge-based approaches (e.g., reasoning systems based on formal logic) are typically based on human domain knowledge and developed through an explicit symbolic representation [11]. This makes them inherently more transparent for humans. The main drawback of purely symbolic approaches is their rigidity which limits their scalability in real-world scenarios. Indeed, complex domains (like HAR) require significant manual efforts from domain experts and knowledge engineers in designing and implementing sophisticated rule-based models. The combination of these two worlds has several potential advantages [5], [12], [13]: (1) handling scenarios that are out of the training set distribution; (2) making the decisions of the system more interpretable and explainable to human users; (3) simplifying both the detection and the resolution of potential wrong decisions of the system; (4) making the system capable of learning from smaller training data and (5) more easily adaptable to different domains.

A promising Neuro-symbolic approach is the *Knowledge Infusion* paradigm, which aims at incorporating external structured knowledge within a deep neural network. For example, such knowledge can be obtained from knowledge graphs (KGs). Since the infusion of knowledge can occur at different levels of depth, a recent survey performed a categorization of Knowledge Infusion approaches as shallow, semi-deep, or deep [6].

In shallow infusion, the data-driven component is directly fed or coupled with a pre-trained model or weight vectors that encode the external knowledge. The use of pre-trained models based on word embeddings (e.g., Word2Vec, GloVe) is the most common form of shallow infusion. Given large text corpora, these models are trained in an unsupervised manner to capture the domain-specific meanings of words, which are then represented as n-dimensional vectors. Such vectors can then be used to feed a deep neural network for a specific classification task [14].

In semi-deep infusion, learnable knowledge constraints are used to guide the learning process of the data-driven model. In [15], knowledge-based constraints are used to guide the adversarial training process of generative models. Such an approach has been experimented on the sentence generation application domain to force the generative model to produce sentences that match a given text template.

Finally, deep knowledge infusion aims at integrating external knowledge within the hidden layers of the deep neural network, combining the representation of the knowledge concepts with the latent representation of data. For instance, in [16] a *Knowledge Infusion Layer* (K-IL) has been specifically designed to merge the output of the last hidden layers of an LSTM with the output of a layer that encodes the external knowledge provided by KGs.

Given this taxonomy, DUSTIN is positioned as a semi-deep knowledge infusion approach, since the constraints from a symbolic reasoning module are integrated into the deep learning classifier to drive the learning process.

In general, knowledge infusion has been mainly applied to the *Natural Language Processing* (NLP) domain [7], [8] and computer vision [17]. Knowledge Infusion has also been used in reinforcement learning tasks in [13] to guide the agent's decisions when it has little experience, relying on a set of knowledge-based functions defined by domain experts.

B. Knowledge Infusion for sensor-based HAR

In the literature, only a few knowledge infusion methods for HAR have been proposed. The majority of the approaches are based on *shallow knowledge infusion*, taking advantage of semantic reasoning only before or after the data-driven training process [18], [19].

In [20], *shallow infusion* is used to associate frequent patterns extracted from an unlabeled dataset using data mining techniques with the corresponding activity. However, this approach can be only applied to environmental sensors in smart-home environments. Other shallow knowledge infusion approaches of HAR take advantage of knowledge-based reasoning before the learning process. In [9], knowledge-driven reasoning is adopted to infer an initial activity model which is then fine-tuned using data-driven techniques, in order to adapt it to the user's habits. However, these approaches can only be applied to environmental sensors data in smart-home environments, and not to inertial sensors data from wearable/mobile devices like we propose in this work.

Considering wearable/mobile device data, a shallow approach has been proposed in [4]. In this work, the probability distribution over the possible activities derived by the classifier is refined by common-sense knowledge constraints to exclude unlikely activities.

Differently from existing works, DUSTIN takes advantage of semantic reasoning to guide the learning phase of a deep learning HAR classifier. Hence, the resulting model incorporates such knowledge to improve the recognition rate.

III. Methodology

In this section, we present DUSTIN, our knowledge infusion approach for sensor-based HAR. The main idea of DUSTIN is that common-sense knowledge constraints in the HAR domain (e.g., lying is more likely performed in indoor environments) have the potential of driving the learning process of a classifier based on deep learning. With respect to standard data-driven approaches that only consider sensor data, DUSTIN learns the correlations between input sensor data and the set of activities that are consistent with the user's context based on symbolic reasoning. These correlations are particularly useful in the classification step since the information about consistent activities positively *"revises"* decisions purely based on sensor data.

Besides an improved recognition rate, this mechanism has the objective of reaching high recognition rates even with limited labeled data and with a lower number of epochs. Indeed, learning the knowledge constraints without symbolic reasoning would require a significantly large dataset of activities performed in a high number of different context conditions. Hence, symbolic reasoning makes it possible to reduce the need for training data.

Moreover, since the learning process is driven by knowledge, DUSTIN generates a model with improved interpretability. Indeed, the common-sense knowledge used to train the classifier can also be used to partially understand the rationale behind each classification output.

A. Overall architecture

In the following, we describe the overall architecture of DUSTIN, which is depicted in Figure 1. The user's physical movements are continuously monitored thanks to the sensor data stream generated by the inertial sensors of the user's mobile devices (e.g., smartphone and smartwatch) as well as raw context data that describes the environment that surrounds the user (e.g., GPS data).

First, the raw context data are provided to the CONTEXT AGGREGATION module that is in charge of generating high-level context data. For instance, given the GPS position, this module returns the semantic position by interacting with a dedicated web service.

Similarly to recent works in the literature [4], high-level context data are provided to a SYMBOLIC REASONING module that uses common-sense knowledge to determine the activities that are *consistent* with the current context. For instance, the *running* activity is not consistent when the user is in the office during working hours.

The high-level context and inertial sensor data are then provided as input (in separate channels) to the AUTOMATIC FEATURE EXTRACTION modules of our DEEP LEARNING CLASSIFIER. The output of the symbolic reasoner is translated into a feature vector that is provided to the deep learning model.

The KNOWLEDGE INFUSION module is in charge of infusing (in the latent space) the context semantic constraints encoded in the symbolic features into the features extracted from inertial sensors and high-level context data. During the training phase, this module learns the correlation between the features learned by the data-driven classifier and external knowledge. The symbolic features are also used during the classification phase (i.e., when using the trained model to recognize activities).

Finally, the output of the KNOWLEDGE INFUSION module is forwarded to the ACTIVITY CLASSIFICATION module to predict the current activity performed by the subject.

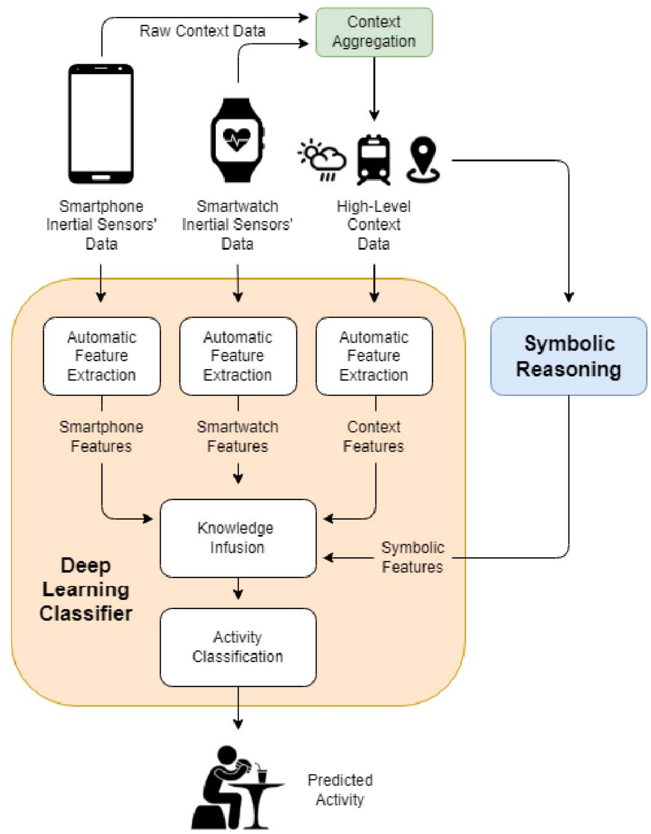

Fig. 1. Overall architecture of DUSTIN

B. Data-driven feature extraction

The streams of raw sensor data continuously collected by each mobile device are temporally aligned and pre-processed before they are provided to the DEEP LEARNING CLASSIFIER. DUSTIN considers both inertial sensors data as well as high-level context data.

Taking into account inertial sensors, we consider common sensing devices installed on nowadays mobile devices (i.e., accelerometer, gyroscope, and magnetometer). In order to reduce the intrinsic noise of the signals, we apply a median filter to each stream.

High-level context data describe the environment which surrounds the user, such as her current semantic location, whether she is in an indoor or an outdoor setting, her proximity to transportation routes, the current weather, and information about the day of the week, the month, and the season. The mobile devices collect such information both from built-in sensors (e.g., GPS) as well as publicly available web services (e.g., transportation service). These data are one-hot encoded into a vector.

We apply a fixed-size segmentation on the pre-processed data (without overlap). In our experimental setup, we considered segments of 4 seconds to detect both simple (e.g., lying) and complex (e.g., brushing teeth) activities. The inertial and high-level context data segments generated as described are provided as separate input flows to the AUTOMATIC FEATURE EXTRACTION modules. In particular, our classifier has one channel for each mobile device to receive inertial data, while one single channel to receive all high-level context data derived by all the considered devices. Each module is based on convolutional layers in charge of deriving meaningful features from each data stream.

More details about the hyper-parameters related to the convolutional layers used in our experiments are reported in Section IV-B.

C. *Knowledge-based symbolic features inference*

The SYMBOLIC REASONING module of DUSTIN (running locally on the mobile devices) analyzes the user's surrounding context to infer symbolic features that encode common-sense knowledge about the context-consistent activities. This module relies on an ontology that models high-level relationships between high-level context and activities. Specifically, thanks to ontological reasoning (i.e., consistency check), it is possible to infer the activities that can be performed considering the user's context.

For this work, we considered the ontology proposed in [4]. This ontology models several categories of context data: user's semantic place, user's presence in an indoor or outdoor setting, user's speed, weather conditions, user's proximity to public transportation stops and routes, user's height variations, environment's noise and light levels, and temporal context (i.e., day of the week, month, and season). Figure 2 shows a small portion of the context data using the Protégé tool[1].

The ontology explicitly states the necessary conditions which make an activity possible in a given context, taking into account domain knowledge. For instance, as shown in Figure 3, the activity *brushing teeth* should take place in an indoor location (i.e., a building), while the user should have null speed and null height variation.

To check whether an activity A is context-consistent, DUSTIN adds to the terminological part of the ontology an axiom representing an instance of Person which identifies the user. Then, available context data are represented as ontological concepts. Finally, DUSTIN relies on context reasoning to check if A is consistent with the user's context.

The high-level context data generated by the CONTEXT AGGREGATION module are automatically mapped to ontological concepts. Most of the context data we considered have a one-to-one mapping with ontological entities. For instance, the user's semantic location provided by public web services is automatically mapped to the corresponding ontological fact,

[1]https://protege.stanford.edu/

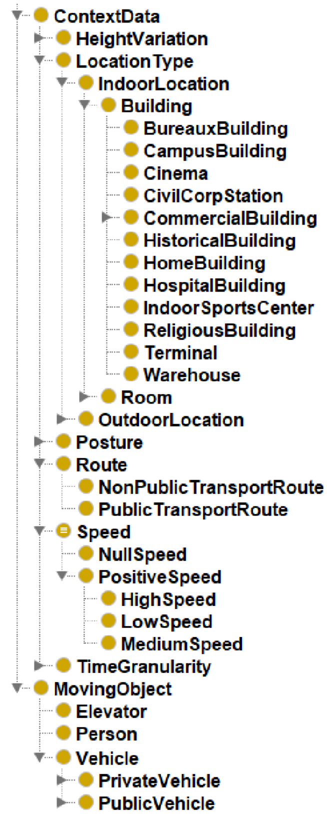

Fig. 2. An excerpt of the context hierarchy in the reference ontology

as described in Example 1. Raw context data available as scalar values are discretized by the CONTEXT AGGREGATION module. For instance, each user's speed value is mapped to one of the following ontological concepts: *NullSpeed*, *LowSpeed*, *MediumSpeed*, and *HighSpeed*. The rules used to discretize scalar values rely on ranges of values designed by knowledge engineers (e.g., speed values greater than 0 km/h and lower than 8 km/h are mapped to *LowSpeed*).

The output of ontological reasoning (i.e., the context-consistent activities) is encoded through a vector, in which each position represents one of the activities that DUSTIN can detect. The value of an element of this vector is 1 if the corresponding activity is consistent with the user's context, 0 otherwise.

Example 1: Alice is using a system that relies on DUSTIN. When the SYMBOLIC REASONING module is triggered to generate the symbolic features to be infused into the data-driven module, *Person(Alice)* is added as a fact. High-level context data are then processed to expand the set of facts. Suppose that the CONTEXT AGGREGATION module of DUSTIN derives that Alice is in a park and that the speed value provided by the GPS sensor of her smartphone is 2 km/h. These context data are automatically instantiated in the ontology with two individuals: *Park(place)* and *LowSpeed(speed)*. Then, existing relationships between Alice and the available context data are

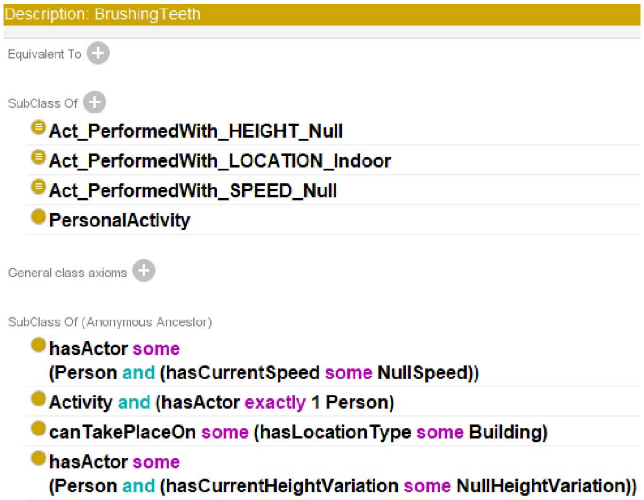

Fig. 3. Definition of the activity "Brushing Teeth" in our ontology

added as facts: *hasCurrentSymbolicLocation(Alice, place)* and *hasCurrentSpeed(Alice, speed)*. Finally, to check whether the activity *walking* is consistent with Alice's context, the symbolic module adds two other axioms: *Walking(currentActivity)* and *isPerforming(Alice, currentActivity)*. The consistency between the set of facts and the domain knowledge will determine if *walking* is consistent with the current Alice's context. This last step is repeated for each activity that the system aims at detecting, in order to create the symbolic features vector that will be infused into the data-driven module.

D. Knowledge infusion and classification

The KNOWLEDGE INFUSION module combines the feature automatically extracted by the convolutional layers from inertial sensors and high-level context data with the symbolic features provided by the symbolic module.

The KNOWLEDGE INFUSION module works in two steps. First, the features are concatenated into a single feature vector in the latent space to infuse the knowledge with sensor and high-level context data. Then, the resulting feature vector is provided to a sequence of fully-connected layers that learn the correlations between input data and context-consistent activities.

Finally, the ACTIVITY CLASSIFICATION module consists of a Softmax layer that is in charge of providing as output the activity currently performed by the subject.

IV. RESULTS

A. Dataset Description

We evaluated DUSTIN on a HAR dataset proposed in a recent work [4]. The main advantage of this dataset is that includes a wide variety of context-dependent activities.

Overall, the dataset includes data from 26 subjects that carried a smartphone in their pocket and a smartwatch on their dominant hand's wrist. For both the mobile devices, the dataset includes raw data about inertial sensors (accelerometer, gyroscope, and magnetometer).

At the same time, the dataset includes context data gathered combining the smartphone's built-in sensors and public web services. Taking into account the smartphone's sensors, the luminosity sensor and the microphone are used to measure the environment's brightness and noise levels, respectively. Instead, the barometer and the GPS quantify the users' height and speed variations. At the same time, the dataset includes the output of different web services: (1) *Google's Places API* provides the semantic places closest to the user; (2) *OpenWeatherMap* provides current local weather conditions (e.g., rainy, sunny); (3) *Bing's Traffic API* provides the nearby traffic situation, and (4) *Transitland* provides transportation routes and stops that are close to the user. The dataset also includes temporal context information, such as the moment of the day (e.g., evening), the day of the week, the season, and so on.

The high-level context data provided as input to the DEEP LEARNING CLASSIFIER cover all the context data presented in the dataset. However, only a subset of context types is used by the SYMBOLIC REASONING module to infer which are the context-consistent activities. Indeed, the environment's brightness and noise levels do not provide useful information to infer meaningful symbolic features. At the same time, without knowing the habits of the users involved in the dataset collection campaign, the SYMBOLIC REASONING module cannot rely on temporal context information like the day of the week.

Overall, 14 different activities are included in the dataset: *walking*, *running*, *standing*, *lying*, *sitting*, *stairs up*, *stairs down*, *elevator up*, *elevator down*, *cycling*, *moving by car*, *sitting on transport*, *standing on transport* and *brushing teeth*. Overall, the dataset contains almost 9 hours of labeled data (≈ 350 activities instances).

B. Experimental setup

In the following, we describe the specific setup of our experiments. The two AUTOMATIC FEATURE EXTRACTION modules related to the inertial sensors data provided by mobile devices are composed of two *convolutional* layers with 8 3×3 and 64 2×2 filters, respectively, separated by a 2×2 *max pooling* layer. After the second *convolutional* layer, we added another 2×2 *max pooling* layer, followed by a *flatten* layer, and, finally, a *fully connected* layer with 128 neurons. The AUTOMATIC FEATURE EXTRACTION module that extracts features from high-level context data is instead composed of a single *fully connected* layer with 8 neurons.

The KNOWLEDGE INFUSION module is composed of a Concatenation layer to combine inertial, context, and symbolic features, followed by a *dropout* layer with a dropout rate of 0.1 and a *fully connected* layer with 256 neurons to extract meaningful information from these concatenated features. Finally, the ACTIVITY CLASSIFICATION module consists of a *softmax* layer for the final classification.

The evaluation has been carried out by splitting the dataset as follows: 70% for the training set, 10% for the validation set, and 20% for the test set. We considered 200 training *epochs*, with a *batch size* of 32 samples. We chose the *adam* optimizer and the *categorical crossentropy* as loss function. Furthermore, we implemented the *early stopping* technique to stop the learning process when the loss accuracy computed on the validation set did not improve for 5 consecutive epochs.

C. Baselines

We compared DUSTIN considering three different baselines.

- *Inertial only*. This baseline represents a standard HAR classifier that does not use context data. Hence, we consider this baseline to evaluate the recognition rate without the use of context data.
- DUSTIN *without symbolic features*. This baseline is used to show the performance of the classifier that processes high-level context data without infusing symbolic reasoning.
- DUSTIN *without context features*. This baseline represents how the infused knowledge impacts the recognition rate when high-level context data is not provided to the network.

For DUSTIN and all the baselines, we also evaluate the application of a state-of-the-art *context refinement* method [4] which discards from the probability distribution emitted as output by the data-driven module those activities which are not consistent with the user's context (*context refinement*). This shallow knowledge infusion approach for HAR is orthogonal with respect to DUSTIN.

D. Experimental Evaluation

In the following, we report the results of the effectiveness of DUSTIN on the dataset described above.

1) Overall results: Figure 4 shows the overall F-1 score of DUSTIN compared to the baselines. Also, we show the effectiveness of each approach with and without context refinement. DUSTIN without context refinement outperforms *Inertial only* by $+46\%$, DUSTIN *without symbolic features* by $+4\%$ and DUSTIN *without context features* by 1%. Consistently with the literature [4], the use of context data significantly outperforms solutions that are purely based on inertial sensors. Overall, the results confirm the advantage of considering symbolic features to improve the model accuracy. Note that the very limited improvement obtained by DUSTIN with respect to DUSTIN *without context features* data is that symbolic features implicitly encode context. Hence considering both information is sometimes redundant for the classifier.

Even when *context refinement* is considered, DUSTIN still outperforms the baselines. Note that *Inertial only* with *context refinement* is exactly the knowledge infusion method proposed in [4]. Moreover, *context refinement* does not have any positive impact on the recognition rate of DUSTIN. This is because knowledge infusion generates accurate predictions that do not benefit from further refinements. Indeed, the more the output probability distribution of the model is reliable, the more the context refinement step is accurate.

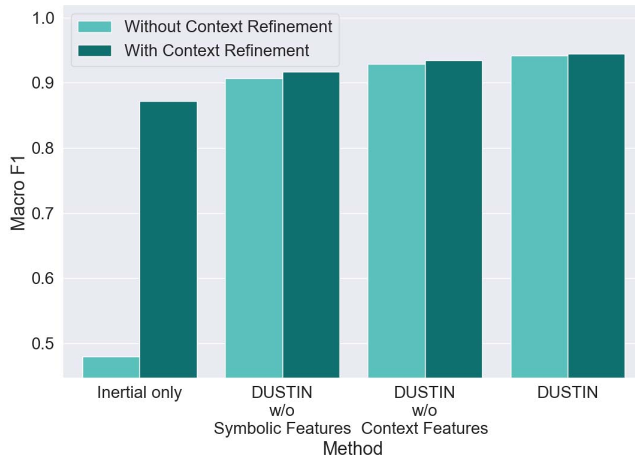

Fig. 4. Overall results

2) Activity-level results: Figure 5 provides the results for each activity in the dataset. As expected, the use of context data has a positive impact on those activities that are strictly related to the user's surroundings. For instance, the proximity of the user to public transportation routes is essential to recognize *sitting/standing on transport*, while *elevator up/down* and *stairs up/down* benefit from information about height variations. The solutions that are based on symbolic features significantly outperform DUSTIN *without symbolic features* considering the activities *cycling* and *moving by car*. Since

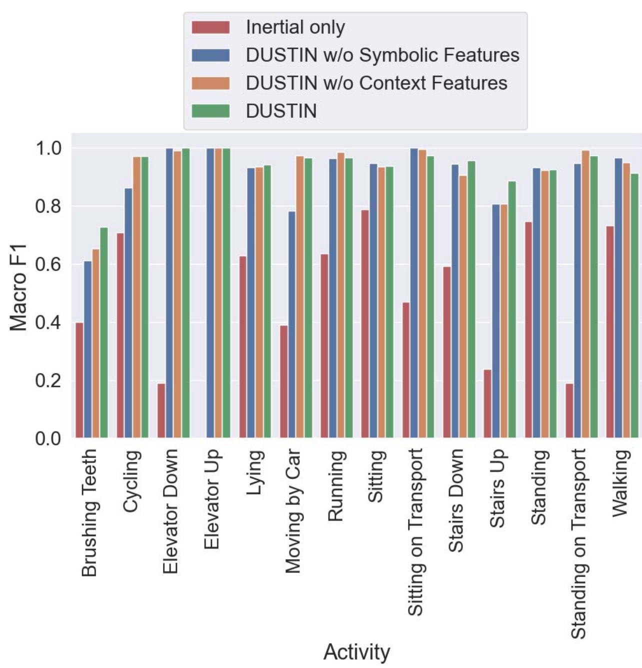

Fig. 5. Activities results

these activities share similar context data, learning these patterns from high-level context data leads to worse recognition rates. Moreover, only a few symbolic features are used to represent context-consistency, hence the CNN focuses more on inertial sensor data patterns (that are significantly different considering the two activities).

Some activities do not exhibit an improvement when symbolic features are considered. For instance, both *elevator up* and *elevator down* are easily recognized by all the methods that involve context data. This happens since this activity is simply characterized by simple context information (e.g., height and speed variations).

We also want to point out that there are a few activities that are slightly negatively affected by symbolic features. This phenomenon occurs considering those activities that are consistent with many contexts (e.g., *walking*). Hence, in such cases, symbolic features may confuse the classifier.

Finally, these results confirm that the advantage of DUSTIN with respect to DUSTIN *without context features* is limited and focused only on some activities (e.g., *brushing teeth*).

3) Results with low labeled data availability: Figure 6 shows how the percentage of available labeled training data affects the recognition rate. In these experiments, we only considered DUSTIN and the baselines that use context since the *Inertial only* approach reaches poor recognition rates even with fully labeled data availability. Considering a very small number of training samples (10% of the training set of the considered dataset) all the considered approaches perform poorly. In particular, the approaches that involve symbolic features are the ones that perform worse. However, when more training samples (20%) are available, the approaches based on symbolic features show a significant boost in recognition rate (that is close to 90% in terms of F1-score). Specifically, DUSTIN exhibits an improvement of about $+9\%$ with respect to DUSTIN *without symbolic features*. Besides, DUSTIN also outperforms DUSTIN *without context features* by 2%. By increasing the number of available labeled training samples, the trend of the performances presented by these three methods reflects the results previously described in Figure 4.

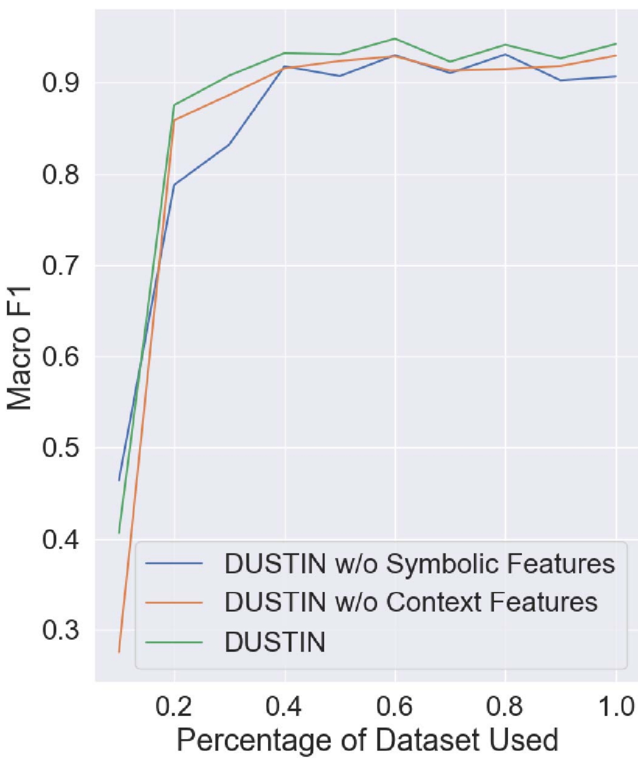

Fig. 6. Overall performance while varying the available training data

These results suggest that symbolic features have the potential of enhancing the recognition rate of the classification model when there are not enough available training data to reliably extract meaningful features from high-level context data.

4) Impact on the number of epochs: Figure 7 shows the evolution of the recognition rate during the training phase (i.e., at each epoch) considering the approaches based on context data. We observe that DUSTIN significantly speeds up the convergence of the deep learning model with respect to other approaches, quickly reaching high recognition rates. DUSTIN *without context features* requires slightly fewer epochs, but it under-performs DUSTIN during the whole learning process. Finally, the number of required epochs is significantly higher when symbolic features are not considered.

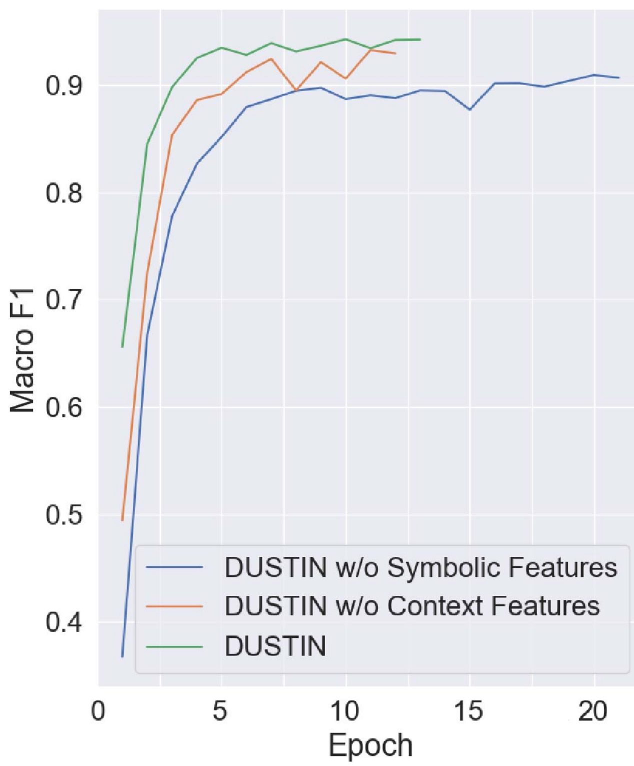

Fig. 7. F1-score trend during training

V. Conclusion

In this paper, we presented DUSTIN, our novel knowledge infusion method for sensor-based HAR. This neuro-symbolic approach relies on ontology reasoning to infer 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.

Even though our preliminary results are promising, DUSTIN still has several limitations that we will tackle in future work. First, the ontology design and implementation require significant manual work from knowledge engineers and domain experts. It is questionable if such a model can generalize over all the possible context conditions and activities [21]. Hence, we will evaluate DUSTIN on other datasets (e.g., the ExtraSensory dataset [22]) and we will also study semi-automatic approaches to obtain such knowledge from external sources (e.g., text, videos, and images on the web). We will also investigate how to introduce probabilistic reasoning in knowledge-based reasoning to improve flexibility.

Another significant limitation of DUSTIN is that ontological reasoning is required during both training and classification. This setting may be not suitable for real-world deployments on mobile devices due to the computational complexity of ontologies. In future work, we plan to design alternative knowledge infusion approaches where the deep learning classifier learns the common-sense knowledge constraints without requiring symbolic reasoning also during the classification phase. For example, this could be achieved by designing a semantic loss function to guide the learning process through knowledge. An alternative solution is to train another neural network that is specialized in mapping high-level context data to symbolic features.

We will also investigate how to practically collect context data in real-world deployments, since there may be problems related to energy consumption, network delay, inference time, QoS, etcetera.

Finally, our future efforts will also focus on analyzing the interpretability of DUSTIN. Since the predictions of DUSTIN 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. Also, we will study whether explanations obtained through eXplainable AI (XAI) methods applied to the predictions of DUSTIN are consistent with the common-sense knowledge.

References

[1] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

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

[3] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

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

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

[6] A. Sheth, M. Gaur, U. Kursuncu, and R. Wickramarachchi, "Shades of knowledge-infused learning for enhancing deep learning," *IEEE Internet Computing*, vol. 23, no. 6, pp. 54–63, 2019.

[7] K. Annervaz, P. K. Nath, and 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] A. S. A. Sukor, A. Zakaria, N. A. Rahim, L. M. Kamarudin, R. Setchi, and H. Nishizaki, "A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4177–4188, 2019.

[10] C. Bettini, G. Civitarese, D. Giancane, and R. Presotto, "Procaviar: Hybrid data-driven and probabilistic knowledge-based activity recognition," *IEEE Access*, vol. 8, pp. 146876–146886, 2020.

[11] G. Civitarese, T. Sztyler, D. Riboni, C. Bettini, and H. Stuckenschmidt, "Polaris: Probabilistic and ontological activity recognition in smart-homes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 209–223, 2019.

[12] M. Gaur, K. Faldu, and A. Sheth, "Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?," *IEEE Internet Computing*, vol. 25, no. 1, pp. 51–59, 2021.

[13] J. Fürst, M. F. Argerich, B. Cheng, and E. Kovacs, "Towards knowledge infusion for robust and transferable machine learning in iot," *Open Journal of Internet of Things (OJIOT)*, vol. 6, no. 1, pp. 24–34, 2020.

[14] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, B. Qin, *et al.*, "Learning sentiment-specific word embedding for twitter sentiment classification.," in *ACL (1)*, pp. 1555–1565, Citeseer, 2014.

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

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

[17] G. Chhablani, A. Sharma, H. Pandey, and T. Dash, "Superpixel-based domain-knowledge infusion in computer vision," *arXiv preprint arXiv:2105.09448*, 2021.

[18] D. Riboni and C. Bettini, "Cosar: hybrid reasoning for context-aware activity recognition," *Personal and Ubiquitous Computing*, vol. 15, no. 3, pp. 271–289, 2011.

[19] W. Van Woensel, P. C. Roy, S. S. R. Abidi, and 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*, vol. 108, p. 101931, 2020.

[20] G. Azkune and A. Almeida, "A scalable hybrid activity recognition approach for intelligent environments," *IEEE Access*, vol. 6, pp. 41745–41759, 2018.

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

[22] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.