Personalized Deep Learning in Human Activity Recognition from Inertial Signals: a Preliminary Study on its Effectiveness

Anna Ferrari[a], Daniela Micucci[a], Marco Mobilio[a] and Paolo Napoletano[a]

[a]*Department of Systems, Information and Communications, University of Milano - Bicocca, viale Sarca 336, Milano, Italy*

Abstract

In the recent years there has been a growing interest in techniques able to automatically recognize activities performed by people. This field is known as Human Activity recognition (HAR). HAR can be crucial in monitoring the wellbeing of the people, with special regard to the elder population and those people affected by degenerative conditions. One of the main challenges concerns the *population diversity* problem, that is, the natural differences between users' activity patterns, which implies that executions of the same activity performed by different people are different. Previous experiments have shown that personalization based on similarity between subjects and signals can increase the accuracy of recognition models of human activities obtained by traditional machine learning techniques. In this article, we investigate whether personalization applied to deep learning techniques can lead to more accurate models with respect to those obtained both by applying personalization to machine learning models, and to traditional deep learning models. In particular, the experiments have been done on two public domain datasets and using the AdaBoost classifier and two Convolutional Neural Networks. Preliminary results show that, on average, traditional deep learning outperforms both personalized deep learning and personalized machine learning techniques.

Keywords

Human activity recognition, personalization, ADL, machine learning, deep learning

1. Introduction

Nowadays smartphones are able to acquire, store, share, and elaborate huge amount of data in a very short time. This technological advancement has attracted the interest of many research fields, including the one dealing with Human Activity Recognition (HAR). Using a smartphone to detect activities, identify potential risks such as falls, and highlight behavioral changes, leads to many advantages, including pervasiveness and low realization costs. Moreover, the increased computational power makes possible to consider not only traditional machine learning, but also more complex deep learning techniques.

Traditional machine learning methods (ML) are low cost in terms of time consumption, data availability, and complexity, however the dependency on expert knowledge in the features

Workshop on Artificial Intelligence for an Ageing Society (AIxAS 2020)

a.ferrari@campus.unimib.it (A. Ferrari); daniela.micucci@unimib.it (D. Micucci); marco.mobilio@unimib.it (M. Mobilio); paolo.napoletano@unimib.it (P. Napoletano)

0000-0001-5718-630X (A. Ferrari); 0000-0001-7116-9338 (D. Micucci); 0000-0002-9421-8566 (M. Mobilio); 0000-0002-9421-8566 (P. Napoletano)

extraction phase often generates weak models difficult to compare [1, 2, 3]. On the other side, deep learning methods (DL) rely on a (mainly) automated feature extraction procedure, but the training phase requires more data, and, consequently, it is either very time consuming or requires expensive hardware [4, 5].

Regardless of the underlying learning method (either classic machine learning or deep learning), real-world HAR systems achieve non satisfying recognition accuracy in real world applications mostly because HAR techniques struggle to generalize to new users and/or new environments [6, 7]. One of the most relevant difficulty to face with new situations is due to the population diversity problem [8], that is, the natural differences between users when they perform the same activities. According to Zunino et al. [9], two factors cause the same activity to be performed differently.

- *Inter-subject variability*, which refers to anthropometric differences of body parts or to incongruous personal styles in accomplishing the scheduled action.
- *Intra-subject variability*, which represents the random nature of a single action class and reflects the fact that the same subject never performs an action in the same way.

To face subjects variability, algorithms should be trained on a representative number of subjects and on as many cases as possible. The number of subjects present in the dataset does not just impact the quality and robustness of the induced model, but also the ability to evaluate the consistency of results across subjects [10]. Nevertheless, in the sensor-based HAR community, datasets are in a low number.

Another way to face variability is to consider *similarity* as a key factor to obtain more robust recognition models. Indeed, previous experiments have shown that personalization based on similarity between subjects and signals can increase the accuracy of recognition models of human activities obtained by traditional machine learning techniques [11].

The rationale behind similarity-based personalizations is based on two considerations.

1. Users with different physical characteristics, such as age or weight, walk or run in a different way. This results in a different accelerometer signal. We refer to this aspect as *physical-based similarity.*
2. Independently from similarities based on physical characteristics, accelerometer signals from two different users may be more similar with respect to other users performing the same activity. We refer to this aspect as *signal-based (or sensor-based) similarity.*

In this article, we investigate whether personalization applied to deep learning techniques can lead to more accurate models than the ones obtained by applying personalization to traditional machine learning models. Moreover, we investigate whether personalized deep learning techniques are more effective with respect to non-personalized deep learning techniques. The evaluation has been performed on two public domain datasets [12, 13] and using AdaBoost as a traditional machine learning classifier and two Convolutional Neural Networks as deep learning techniques.

Preliminary results show that personalization applied to CNNs leads to more accurate models with respect to the ones obtained by applying personalization to AdaBoost only in one dataset, namely Motion Sense. Moreover, traditional CNN in average obtained better results in most of the configurations used.

The paper is organized as follows. Section 2 discusses similarity and specifies how it is employed in traditional machine and deep learning techniques; Section 3 describes the set up of our experiments; Section 4 presents the results of the experiments; finally, Section 5 presents the conclusions and outlines future research on personalization.

2. Proposed Methods

To take into account the population diversity, we introduce the concept of similarity between subjects. The similarity between subjects is used to weight the training data in order to give more importance to data that are more similar to the data of the user under test.

Each subject i can be described with a feature vector $\mathbf{g}_i = \{g_1, \dots, g_K\}$. Similarity between two subjects i and j is defined as follows.

$$\text{sim}(i, j) = e^{-\gamma d(i,j)} \tag{1}$$

where γ is a scale parameter and $d(i, j)$ is the Euclidean distance between the feature vectors of two subjects:

$$d(i, j) = \sqrt{\sum_{k=1}^{K} (g_{k,i} - g_{k,j})^2} \tag{2}$$

The resulting similarity value ranges from 0 to 1, where 0 means that the two subjects are dissimilar, and 1 means that the two subjects are equal. The idea is to take advantage of the similarity between subjects in machine learning and deep learning engines as follows.

- **Personalized Machine Learning (PML)**. Given a subject i under test, all the training data are weighted by using the similarity between the user i and the rest of the users. We can define three types of similarity: *physical-based* ($\text{sim}^{physical}$), *sensor-based* (sim^{sensor}), and *physical combined with sensor-based similarity* ($\text{sim}^{physical+sensor}$). *Physical-based* similarity exploits age, weight, and height of the subjects. The choice of these characteristics is inspired by the literature and it is subject to the availability of the metadata within the public data sets. Details about the types of similarity considered in this study can be found in [11].
- **Personalized Deep Learning (PDL)**. Starting from a minimum value m we select the most m similar subjects, with respect to the test subject. The network is trained with the samples related to these m subjects. We experimented several m values starting from 10 to the maximum number of subjects available in the dataset with a step of 5.

These two methods have been compared with a traditional end-to-end deep learning methods (**DL**).

3. Experimental setup

The PML technique used is an Adaboost classifier, while both the PDL and the DL techniques used are Convolutional Neural Networks.

In particular, as for PML we adopt the AdaBoost classifier as described in [11]. As for PDL and DL we adopted a Residual Network (ResNet) based on the ResNet proposed in [4] and [5]. The input size of the network is $1 \times 128 \times 3$, that corresponds to 3 segments along the three axes x, y, and z. The network architecture is made of an initial convolutional block, 3 residual stages, each containing a variable number n of residual blocks, average pooling layer, fully connected layer, and softmax layer. A convolutional block is made of three layers: convolutional, batch normalization, and ReLu. A residual block is made of 2 subsequent convolutional blocks and an additional operator that sums the input of the residual block with the output of the residual block itself. Each convolutional layer is $1 \times 3 \times f_{maps}$, where f_{maps} is the number of feature maps of the filter. For each dataset, the best values for n and f_{maps} have been found by following a grid search approach: n ranged between 3 and 21, while f_{maps} ranged between 10 and 200.

Data has been split according to two different configurations [11]: *subject-independent* (**SI**) and *hybrid* (**HYB**). The SI data split configuration does not use the end user data for the development of the activity recognition model, that is, the classification model is trained on the data of the users except the end user. The HYB data split configuration uses the end user data and the data of the other users for the development of the activity recognition model, that is, the classification model is trained both on the data of the users and on a part of the data of the end user.

Two public datasets containing accelerometer signals of Activities of Daily Living (ADLs) and Falls ave been used in the experimentation.

- **UniMiB-SHAR** [12] contains tri-axial acceleration data organized in 3s windows around the peak. The dataset contains 17 different activities (both ADLs and Falls) performed by 30 subjects. Sex, age, weight, and height of each subject are known. The original sampling rate is 50Hz. We have chosen segments of 3 seconds for this dataset. The subjects placed the smartphone used for the acquisition (a Samsung Galaxy Nexus I9250) half of the times in the left trouser pocket and the remaining times in the right one.

- **Motion Sense** [13] contains time-series data generated by the accelerometers in an iPhone 6s worn by 24 participants. Sex, age, weight, and height of each subject are known. Each of the subjects performed 6 activities (only ADLs). The smartphone were kept in the participant's front pocket. The original sampling rate is 50Hz. We have chosen segments of 5 seconds for this dataset.

4. Results

Table 1 shows results in terms of macro average accuracy (i.e., the average across subjects, splits, and m selection of subjects when deep learning technique is considered).

Comparison between PML and PDL leads to contrasting conclusions. In UniMiB-SHAR dataset, PML shows better results. In particular, for UniMiB-SHAR the best-performing model

Table 1
Experimental Results - accuracy of PDL and of PML compared on averange with DL.

	UniMiB-SHAR		**Motion Sense**	
	PDL - PML	DL	PDL - PML	DL
SI-physical	30.00 - **57.39**		**76.57** - 72.45	
SI-sensor	42.08 - **57.00**		**77.51** - 74.03	
SI-physical sensor	42.09 - **56.93**		**77.51** - 73.85	
average	38.06 - 57.11	**58.88**	77.20 - 73.44	**81.03**
HYB-physical	44.42 - **85.44**		**79.06** - 77.76	
HYB-sensor	46.62 - **84.71**		**79.65** - 78.06	
HYB-physical sensor	46.27 - **84.87**		**79.81** - 77.86	
average	45.77 - **85.00**	69.72	79.51 - 77.89	**85.75**

is the one with the hybrid split and related to the physical similarity. In Motion Sense it is the personalized deep learning (PDL) that achieves the best performance of 79.81%. In general, the hybrid model is the best-performing one.

Finally, comparison with traditional deep learning techniques (DL) leads, with these datasets, to assert that on average DLs achieve better results. The only exception is when dataset UniMiB-SHAR with a hybrid data split is used.

This preliminary study seems to favor the traditional deep learning techniques. However, this study does not favor in a clear way one personalized method with respect to the other (PML vs PDL).

A further investigation is required. In particular it is worth to deepen how similarity of the subjects is distributed within the dataset. Moreover, other datasets should be experimented.

5. Conclusion

Over last decades, HAR has been a very active field. Nevertheless the lack of availability of large datasets prevent the traditional algorithms to generalize in real world situation.

Personalized machine learning and deep learning techniques are becoming more and more popular because of their promising results.

In this study we showed that traditional deep learning outperform personalized technique in most of the cases. Although, results on UniMiB-SHAR still confirm that personalized machine learning can yield better results.

Given the contrasting results obtained with UniMiB-SHAR and Motion Sense datasets, we planned further investigation using other datasets, such as, for instance, MobiAct [14].

References

[1] R. Zhu, Z. Xiao, Y. Li, M. Yang, Y. Tan, L. Zhou, S. Lin, H. Wen, Efficient human activity recognition solving the confusing activities via deep ensemble learning, IEEE Access 7 (2019) 75490–75499.

[2] N. H. Friday, M. A. Al-garadi, G. Mujtaba, U. R. Alo, A. Waqas, Deep learning fusion conceptual frameworks for complex human activity recognition using mobile and wearable sensors, in: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), IEEE, 2018, pp. 1–7.

[3] T. Yu, J. Chen, N. Yan, X. Liu, A multi-layer parallel lstm network for human activity recognition with smartphone sensors, in: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), IEEE, 2018, pp. 1–6.

[4] A. Ferrari, D. Micucci, M. Marco, P. Napoletano, Hand-crafted features vs residual networks for human activities recognition using accelerometer, in: Proceedings of the IEEE International Symposium on Consumer Technologies (ISCT), 2019.

[5] A. Ferrari, D. Micucci, M. Mobilio, P. Napoletano, Human activities recognition using accelerometer and gyroscope, in: European Conference on Ambient Intelligence, Springer, 2019, pp. 357–362.

[6] J.-H. Hong, J. Ramos, A. K. Dey, Toward personalized activity recognition systems with a semipopulation approach, IEEE Transactions on Human-Machine Systems 46 (2016) 101–112.

[7] R. Igual, C. Medrano, I. Plaza, A comparison of public datasets for acceleration-based fall detection, Medical engineering & physics 37 (2015) 870–878.

[8] N. D. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. T. Campbell, F. Zhao, Enabling large-scale human activity inference on smartphones using community similarity networks (csn), in: Proceedings of the International Conference on Ubiquitous Computing (UbiComp), 2011.

[9] A. Zunino, J. Cavazza, V. Murino, Revisiting human action recognition: Personalization vs. generalization, in: International Conference on Image Analysis and Processing, Springer, 2017, pp. 469–480.

[10] J. W. Lockhart, G. M. Weiss, Limitations with activity recognition methodology & data sets, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, 2014, pp. 747–756.

[11] A. Ferrari, D. Micucci, M. Mobilio, P. Napoletano, On the personalization of classification models for human activity recognition, IEEE Access 8 (2020) 32066–32079.

[12] D. Micucci, M. Mobilio, P. Napoletano, Unimib shar: A dataset for human activity recognition using acceleration data from smartphones, Applied Sciences 7 (2017) 1101.

[13] M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi, Protecting sensory data against sensitive inferences, in: Proceedings of the Workshop on Privacy by Design in Distributed Systems (W-P2DS18), 2018.

[14] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, M. Tsiknakis, The mobiact dataset: Recognition of activities of daily living using smartphones., in: Proceedings of Information and Communication Technologies for Ageing Well and e-Health (ICT4AgeingWell16), 2016.