Adaptation and Selection Techniques Based on Deep Learning for Human Activity Recognition Using Inertial Sensors †

Manuel Gil-Martín *, José Antúnez-Durango and Rubén San-Segundo

Speech Technology Group, Information Processing and Telecomunications Center, E.T.S.I. Telecomunicación UPM;

* Correspondence: manuel.gilmartin@upm.es; Tel.: +34-91-067-25-00

† Presented at the 7th Electronic Conference on Sensors and Applications, 15–30 November 2020; Available online: https://ecsa-7.sciforum.net/.

Abstract: Deep learning techniques have been widely applied to Human Activity Recognition (HAR), but a specific challenge appears when HAR systems are trained and tested with different subjects. This paper describes and evaluates several techniques based on deep learning algorithms for adapting and selecting the training data used to generate a HAR system using accelerometer recordings. This paper proposes two alternatives: autoencoders and Generative Adversarial Networks (GANs). Both alternatives are based on deep neural networks including convolutional layers for feature extraction and fully-connected layers for classification. Fast Fourier Transform (FFT) is used as a transformation of acceleration data to provide an appropriate input data to the deep neural network. This study has used acceleration recordings from hand, chest and ankle sensors included in the PAMAP2 dataset. This is a public dataset including recordings from nine subjects while performing 12 activities such as walking, running, sitting, ascending stairs or ironing. The evaluation has been performed using a Leave-One-Subject-Out cross-validation: all recordings from a subject are used as testing subset and recordings from the rest subjects are used as training subset. The obtained results suggest that strategies using autoencoders to adapt training data to testing data improve some users' performance. Moreover, training data selection algorithms with autoencoders provide significant improvements. The GAN approach, using the generator or discriminator module, also provides improvement in selection experiments.

Keywords: Human Activity Recognition; adaptation; feature selection; autoencoders; Generative Adversarial Network; deep learning; PAMAP2

1. Introduction

Human Activity Recognition (HAR) using wearable sensors is a widely studied field [1,2] with a high number of applications such as physical activity classification [3,4], health [5,6] or biometrics [7]. In this field, daily life activities performed by humans are modelled by a system and could be recognized afterwards. However, each user presents specific and unique patterns that are not easily generalizable by a model. In this sense, adaptation approaches could adjust training data to testing domain to generate a model according to the target subject. Moreover, these approaches could also process testing data to generate a version adapted to a model generated with training data. Therefore, user adaptation techniques allow to modify user-independent models into user-dependent models to improve the HAR performance.

In a recent previous work [8], the authors developed and evaluated three adaptation techniques for HAR: data augmentation, feature matching and confusion maximization. They concluded that performing unsupervised domain adaptation and domain generalization for HAR in wearing

diversity problem is a very complicated task and they proposed some practical considerations for unsupervised domain adaptation. One of these considerations is applying other unsupervised approaches. Another previous work [9] proposed a user adaptation based on Convolutional Neural Networks (CNNs). In this case, the authors included special user-dependent layers in the architecture that are estimated using a small amount of data from a specific user.

The purpose of this work is presenting autoencoders and Generative Adversarial Networks (GANs) as two new adaptation techniques for HAR.

2. Materials and Methods

This section provides a description of the dataset, the signal processing, the deep learning adaptation approaches, and the cross-validation methodology used in this study. The workflow of the methodology consists of the extraction of activity recordings from different sensors of the dataset, the signal processing module to generate the spectra representation and the deep learning approaches. Once the signals are processed, they feed different deep learning architectures in parallel to analyze the performance of the proposed unsupervised adaptation and selection techniques: autoencoders and GANs. In this work, GNU Octave software is used for processing the signals of the dataset and Keras framework (written in Python) with Tensorflow backend for developing the deep learning approaches.

2.1. Dataset

For this work, we used the PAMAP2 dataset [10], which contains recordings of 12 different physical activities (lying, sitting, standing, walking, running, cycling, Nordic walking, ascending stairs, descending stairs, vacuum cleaning, ironing, and rope jumping), performed by nine subjects. These subjects wore three Inertial Measurement Units (IMUs) with tri-axial accelerometer, gyroscope, and magnetometer. These sensors were placed on the dominant hand, chest and ankle and they collected data sampling at 100 Hz. In this work, we used the accelerometer signals to model the physical activities.

2.2. Signal Processing

We implemented a signal processing module to generate windows of physical activity. First, we dealt with wireless data loss performing a linear interpolation of raw sensory data. Second, we used Hamming windowing to divide the data into 5.12-second windows (512 samples per window) separated by 1 s. We labelled each window using the mode of the classes included in the data sequence. Third, we computed the module of the Fast Fourier Transform (FFT) to transform data from time to frequency domain. We obtained a 256-point representation of the spectrum in the 0–50 Hz frequency band. As physical activity concentrates its energy in low frequencies, we selected the first 128 points of the spectrum corresponding approximately to the 0–25 Hz frequency range. Figure 1 represents the signal processing performed to the acceleration signals.
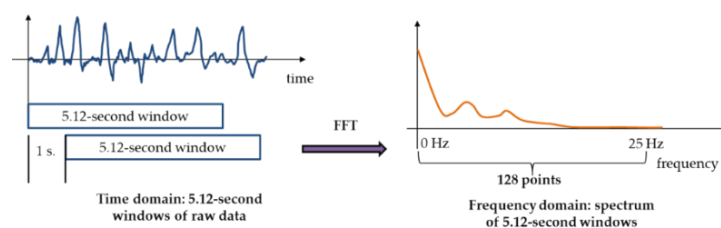
Figure 1. Signal processing performed to the inertial signals.

2.3. Deep Learning Adaptation Approaches

As baseline system, we used a deep learning architecture with four convolutional layers and two max-pooling layers for learning relevant features from signals spectra and three fully-connected

layers for classifying activities, as shown in Figure 2. In this work, we used nine signals with 128 samples of the spectrum as input of the architecture.
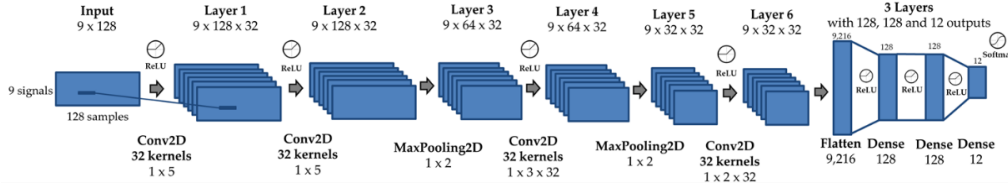

Figure 2. Deep learning architecture used in this work.

In this work, we included dropout layers (with a fraction of 0.3) after convolutional and fully-connected layers to avoid overfitting. ReLU is the activation function in intermediate layers to reduce the impact of gradient vanishing effect and softmax is the activation function in the last layer to perform the classification task. We used the training set (using part of the training set for validation) to tune number of epochs (5) and batch_size (50). The optimizer was fixed to the root-mean-square propagation method [11] with a learning rate of 0.001.

2.3.1. Autoencoders

An autoencoder is a type of neural network which is trained to reproduce the input at the output through unsupervised learning. Autoencoders are composed of two main modules: encoder, which transforms input data to an advanced feature representation, and decoder, which decompresses this representation to the original data. In this work, autoencoders approach is used for adapting the training data to testing data and vice versa and for selecting the most appropriate training data. The autoencoder architecture used in this work has an encoder composed of three convolutional layers and a max-pooling layer and a decoder with three convolutional layers and an up-sampling layer. The autoencoder uses 32 kernels with (1,5) size for convolutional layers and a pooling kernel of (1,2).

In the adaptation approach, training data is adapted to the testing domain or vice versa. The procedure to adapt training data to testing data is as follows. First, the autoencoder is trained using training data to capture the main features of the training subjects. Second, the decoder module of the autoencoder is replaced by a new decoder and the encoder module is configured to avoid training the encoder part. Third, the new decoder is trained using testing data. After this adaptation, we have generated an autoencoder able to transform data from training domain to testing domain (train2test autoencoder). The autoencoder was trained using 10 epochs and a root-mean-square optimizer with a learning rate of 0.001. This procedure could be also performed to generate an autoencoder able to transform data from testing domain to training domain (test2train autoencoder). After processing original training data with the train2test autoencoder, we use the baseline deep learning architecture to model and classify the physical activities. Figure 3 represents the training to testing adaptation process using autoencoders.
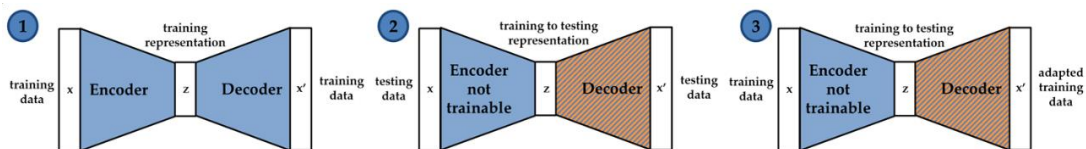

Figure 3. Training to testing adaptation using autoencoders.

In the selection approach, we compute the Root Mean Square (RMS) error between original training data and adapted training data obtained in the adaptation approach. We remove the windows from the original training data where the adapted data is greater than a threshold based on the RMS mean and standard deviation of the adapted training data. The following equation represents the computation of this threshold:

$$\text{value}_{RMS-adapted\ data} >= \mu_{RMS-adapted\ data} + \gamma * \sigma_{RMS-adapted\ data} \tag{1}$$

where γ is a double value optimized during experimentation (γ = 2.75). After this selection of the most appropriate training data, we use the baseline deep learning architecture to model and classify the physical activities.

2.3.2. Generative Adversarial Networks

A GAN is a system where two neural network modules compete in a game where one's victory is another's loss. These two modules are called generator and discriminator. The generator module uses latent space and tries to generate samples close to ones of the true distribution. The discriminator tries to distinguish candidates created by the generator module from real ones. Both modules are trained through backpropagation so that the generator produces better data, while the discriminator becomes more skilled at detecting fake data. In this work, the generator module has the same architecture than the autoencoder and the decoder has the same architecture than the baseline system.

In the adaptation approach, training data is adapted to the testing domain or vice versa. In this adaptation procedure, GAN architecture is trained to transform training data (fake samples) into data as similar as possible to testing data (real samples). First, the discriminator is trained with a random subset of testing data, fixing its output to one because these are real samples. Second, a random subset of training data is processed through the generator and its output is processed through the discriminator, fixing its output to zero because these are fake examples. Third, the GAN is configured to avoid training the discriminator part and the generator part is trained with training data, fixing its output to one because we want that the generator cheats the discriminator with generated examples that are similar to the testing data. After processing original training data with the generator, we use the baseline deep learning architecture to model and classify the physical activities. Figure 4 represents the training to testing adaptation process using GANs.
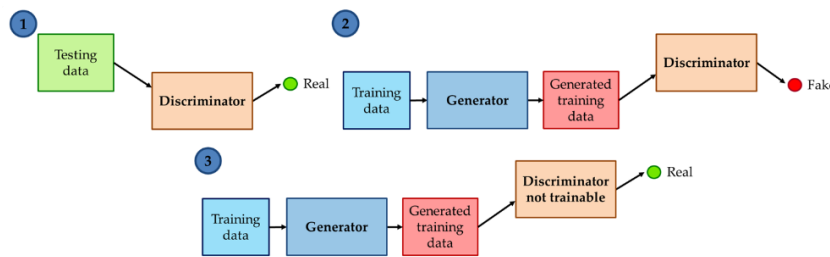
Figure 4. Training to testing adaptation using GANs.

Regarding selection with GANs, it is possible to use generator or discriminator to select the most appropriate training data.

In the selection approach with the generator, we compute the RMS error between original training data and adapted training data obtained in the adaptation approach. After that, we perform a similar selection procedure than in autoencoder selection approach with Equation (1), fixing γ to 3.5 in this case.

In the selection approach with the discriminator, once the GAN architecture is trained to adapt training data to testing data, we predict the capability of training data to impersonate testing data. We process the training data through the discriminator to obtain this measure between 0 and 1. After that, we remove the windows from the original training data where the output of the discriminator is lower than a threshold based on the mean and standard deviation of the discriminator output. The following equation represents the computation of this threshold:

$$\text{value}_{\text{discriminator output}} < \mu_{\text{discriminator output}} - \gamma * \sigma_{\text{discriminator output}} \tag{2}$$

where γ is a double value optimized during experimentation (γ = 2.5).

2.4. Leave-One-Subject-Out Cross-Validation

In this study, we used a Leave-One-Subject-Out (LOSO) Cross-Validation (CV). This evaluation technique allows to test a system with data from a subject that has not seen during training process. We performed round-robin strategy, training the system with data from eight out of nine subjects and testing the system with data from the remaining one. We used a 10% of the training data to validate the system. This configuration has been repeated nine times and the results in this work are average values obtained throughout the ninefold cross validation procedure. This way, recordings from the same subject did not appear in training and testing subsets at the same experiment. This strategy supports the need of adapting training and testing domains because the system is tested with data from a subject that is not included during training process.

3. Results

Table 1 summarizes the main results of this paper. Activity classification accuracy and confidence intervals (CI) using a 95% of significance level are included in the table. Accuracy is the ratio between the number of correctly classified examples and the number of total examples. For each adaptation experiment, the significant improvement is highlighted in green, the significant deterioration is highlighted in red and no color is included when there is no statistical difference between results. As shown in the table, the adaptation approaches reach higher number of significant improvements than deteriorations. It could be observed that selection approaches using autoencoders and using GAN generator reach a general significant improvement in accuracy. Moreover, other approaches such as adapting training data to testing data with autoencoders and selection of training data using GAN discriminator improve the performance for some users.

Table 1. Activity classification accuracy of baseline and adaptation approaches.

Experiment		Subject									
		101	102	103	104	105	106	107	108	109	All
Baseline	Acc (%)	86.34	87.93	85.91	89.76	87.69	85.15	92.04	87.61	96.88	**87.85**
	CI (±%)	1.35	1.24	1.63	1.23	1.23	1.39	1.10	1.26	4.26	0.46
Adapt. Auto.	Acc (%)	86.26	82.88	89.23	93.35	87.36	86.23	91.91	90.62	93.75	**88.36**
	CI (±%)	1.35	1.44	1.45	1.02	1.25	1.35	1.11	1.12	5.93	0.45
Select. Auto.	Acc (%)	86.50	87.17	88.95	91.45	88.06	85.83	92.65	92.30	87.50	**89.06**
	CI (±%)	1.34	1.28	1.47	1.14	1.22	1.37	1.06	1.02	8.10	0.44
Select. GAN Gen.	Acc (%)	85.58	85.84	90.38	90.15	88.39	88.04	92.04	91.69	98.44	**88.94**
	CI (±%)	1.38	1.33	1.38	1.21	1.20	1.27	1.10	1.06	3.04	0.44
Select. GAN Discr.	Acc (%)	87.46	88.99	87.57	84.45	89.75	86.79	93.08	90.62	96.88	**88.68**
	CI (±%)	1.30	1.19	1.55	1.48	1.14	1.33	1.03	1.12	4.26	0.45

4. Discussion and Conclusions

The adaptation techniques proposed in this work allow adapting the training data to testing data and vice versa and selecting the most appropriate training data once adaptation procedure is completed. Results suggest that performing adaptation for HAR is a very difficult task. However, selection approaches could remove from training data the examples that differ from testing domain and improve the general performance of the system. As shown in the results, there exist more subjects than significantly improve their activity classification accuracy than subjects than significantly deteriorate it. This fact implies that adaptation techniques for HAR could benefit individual classification performance or both individual and total classification performances when adapting training data to testing data.

The unsupervised learning adaptation approaches described in this work could pave the way for future research on domain adaptation in HAR. It would be interesting to focus the research on these techniques to achieve fruitful HAR adaptation systems.

References

1. Wang, J.D.; Chen, Y.Q.; Hao, S.J.; Peng, X.H.; Hu, L.S. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11, doi:10.1016/j.patrec.2018.02.010.
2. Slim, S.O.; Atia, A.; Elfattah, M.M.A.; Mostafa, M.-S.M. Survey on Human Activity Recognition based on Acceleration Data. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 84–98.
3. Gil-Martin, M.; San-Segundo, R.; Fernandez-Martinez, F.; Ferreiros-Lopez, J. Improving physical activity recognition using a new deep learning architecture and post-processing techniques. *Eng. Appl. Artif. Intell.* **2020**, *92*, doi:10.1016/j.engappai.2020.103679.
4. Gil-Martín, M.; San-Segundo, R.; Fernández-Martínez, F.; de Córdoba, R. Human activity recognition adapted to the type of movement. *Comput. Electr. Eng.* **2020**, *88*, 106822, doi:10.1016/j.compeleceng.2020.106822.
5. Gil-Martin, M.; Montero, J.M.; San-Segundo, R. Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks. *Electronics* **2019**, *8*, 907, doi:10.3390/electronics8080907.
6. San-Segundo, R.; Navarro-Hellin, H.; Torres-Sanchez, R.; Hodgins, J.; De la Torre, F. Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease. *Electronics* **2019**, *8*, 119, doi:10.3390/electronics8020119.
7. Gil-Martin, M.; San-Segundo, R.; de Cordoba, R.; Manuel Pardo, J. Robust Biometrics from Motion Wearable Sensors Using a D-vector Approach. *Neural Process. Lett.* **2020**, doi:10.1007/s11063-020-10339-z.
8. Chang, Y.; Mathur, A.; Isopoussu, A.; Song, J.; Kawsar, F. A Systematic Study of Unsupervised Domain Adaptation for Robust Human-Activity Recognition. *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 39, doi:10.1145/3380985.
9. Matsui, S.; Inoue, N.; Akagi, Y.; Nagino, G.; Shinoda, K. User adaptation of convolutional neural network for human activity recognition. In Proceedings of 2017 25th European Signal Processing Conference (EUSIPCO), Kos Island, Greece, 28 August–2 September 2017; pp. 753–757.
10. Reiss, A.; Stricker, D.; Ieee. Introducing a New Benchmarked Dataset for Activity Monitoring. *2012 16th Int. Symp. On Wearable Comput. (ISWC)* **2012**, 108–109, doi:10.1109/iswc.2012.13.
11. Weiss, N.A. *Introductory Statistics*; Pearson: 2017.