# Unsupervised Domain Adaptation for Human Activity Recognition*

Paulo Barbosa[1], Kemilly Dearo Garcia[2,4], João Mendes-Moreira[1,3][0000−0002−2471−2833], and André C.P.L.F. de Carvalho[2][0000−0002−4765−6459]

[1] Faculty of Engineering, University of Porto, Portugal, up201607828@g.uporto.pt
[2] ICMC, University of São Paulo, Brazil, andre@icmc.usp.br
[3] LIAAD-INESC TEC, Porto, Portugal, joao.mendes.moreira@inesctec.pt
[4] University of Twente, Netherlands, kemilly.dearo@usp.br

**Abstract.** Human Activity Recognition has been primarily investigated as a machine learning classification task forcing it to handle with two main limitations. First, it must assume that the testing data has an equal distribution with the training sample. However, the inherent structure of an activity recognition systems is fertile in distribution changes over time, for instance, a specific person can perform physical activities differently from others, and even sensors are prone to misfunction. Secondly, to model the pattern of activities carried out by each user, a significant amount of data is needed. This is impractical especially in the actual era of Big Data with effortless access to public repositories. In order to deal with these problems, this paper investigates the use of Transfer Learning, specifically Unsupervised Domain Adaptation, within human activity recognition systems. The yielded experiment results reveal a useful transfer of knowledge and more importantly the convenience of transfer learning within human activity recognition. Apart from the delineated experiments, our work also contributes to the field of transfer learning in general through an exhaustive survey on transfer learning for human activity recognition based on wearables.

**Keywords:** Human Activity Recognition · Transfer Learning · Unsupervised Domain Adaptation.

## 1 Introduction

Human Activity Recognition (HAR) is a machine learning task that induces human activity models able to classify human activities. Though, HAR has relevant applications, including in health care, it still has several challenges that

* This work was partially funded by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-016883.

need to be overcome, including the subject sensitivity and model adaptability. The predictive performance of a classification model in human activities is largely affected by the physiological features of each person. For instance, young people perform ambulation activities different from elder or disabled people [1]. Although a specialised model could be developed for each individual, the cost would be much higher than to induce more generic models, with high predictive performance for different user profiles.

When HAR systems are used for classification of human activities, they usually assume the data distribution is stationary. However this is not necessarily true. Users tend to perform activities differently over time, as they get older or face temporary movement restrictions. Besides, sensors may need frequent calibration and are prone to misfunction.

In this work, we provide a literature revision in Transfer Learning (TL) to leverage the performance of HAR systems. TL is a Machine Learning (ML) sub-area that focuses on the re-utilisation of knowledge between ML tasks, by allowing the dataset distributions and even their training and testing sets be different.

The primary goal we expect to achieve by applying TL in HAR is through different, but related, datasets, boost the system's performance of the target domain. As a side effect of achieving this goal, we expect to reduce the effort of capturing labelled data increasing the efficiency of a HAR system.

## 2 Transfer Learning in HAR

HAR relies on the deployment of pervasive but non-intrusive sensors in the environment surrounding humans. HAR techniques are able to help understanding the context in which a person is involved, especially when help is needed, as in health care systems. Most typical ML models assume that both training and test datasets have the same distribution and feature space. This assumption is indeed a significant limitation since it's proven [2] that real problems suffer distribution mismatch over time and the test error is in proportion to the given difference.

### 2.1 Transfer Learning

The learning paradigm of TL uses past-knowledge, whether through datasets or model parameters, to leverage the performance of a different, but related, problem. In fact, this paradigm is inspired on our biological learning process, that the more experience we have better prepared are we when confronting novel conditions [3].

Essentially, a TL problem is composed of two categories of datasets, source and target. The primary goal is to, using a source dataset, leverage the performance in a target datasets. These two datasets have to be different whether regarding the source and target domains or both tasks. The knowledge of the particular cause of the difference is crucial, since it dictates which TL approach can be applied. A dataset domain consists of two elements, a feature space and a marginal probability distribution, while its task is composed of two elements, a label space and an objective predictive function. A distribution mismatch between datasets is considered a dataset-shift problem.

### 2.2 Dataset shift in HAR

According to a survey of the related literature, the most likely shift-cause agents within HAR are the sensor and the user.

Sensor misplacement is likely to happen since it can sometimes be intrusive to the user causing him to change the position voluntarily. Furthermore, if even the sensor is placed in a comfortable position, it can be misplaced involuntarily during the activities, especially the ambulant ones. For this specific problem, we emphasize the works by Calatroni et al. [4], Khan et al. [5], Rokni [6], Krishnan [7], Kurz et al. [8], Roggen et al. [9], and Zhao et al. [10].

Regarding the user, the age and the physiology are the main causes for different activity executions and must be thought of when training a HAR classification model, to prevent training a model with users different from the target user. For user-related shift problems, various works have been proposed, such as the works of Deng et al. [11], Zhao et al. [12], Chen et al. [13], Hachiya et al. [14], Diethe et al. [15], and Fallahzadeh [16].

There are other shift causes, such as the environment itself, which are typical in smart homes HAR, where models are trained with laboratory data that do not accurately represent a real-world scenario. As research works have shown, laboratory environments generally restrict and influence subject activity patterns [17].

## 3 Unsupervised Domain Adaptation

Unsupervised Domain Adaptation is one specific scenario of TL recurrent in HAR contexts because it assumes the availability of labelled data only from the source domain and unlabelled data from the target domain, the same feature space and label space, and a distribution mismatch (e.g. different users, sensor's positions, surrounding environments). We firstly notice four different approaches to tackle this field: instance-weighting; self-labelling, feature representation; and cluster-based [19]. An instance weighting approach assigns weights to the source samples depending on their similarity with the target samples. A self-labelling approach focuses on adapting classifiers trained from the source samples to the new target domain. An approach of feature representation constructs an abstract representation of the data, while a cluster-based one assumes that from high-density regions we can perceive the similarity between the two domains.

### 3.1 Feature-Representation Employed Techniques

Transfer Component Analysis (TCA) [20] recurs to a dimensionality reduction method, Maximum Mean Discrepancy Embedding (MMDE)[21], that learns a low-dimensional latent space common to both domains. This way, it embeds both source and target domains into a shared low-dimensional latent space using a non-linear mapping function to learn then a corresponding kernel matrix.The kernel matrix is defined on all data by minimising the distance, measured with the Maximum Mean Discrepancy metric, between the projected two domains while maximising the embedded data variance. Having found a transformation matrix, a classifier can finally be trained with it to predict the target domain accurately.

Subspace Alignment (SA) [22] first generates the subspaces by representing the source and target domains into subspaces spanned by eigenvectors. Then, it focuses on the alignment of the two subspaces to decrease the discrepancy between both the source and target domains.

### 3.2 Instance-Weighting Employed Techniques

Nearest Neighbour Weighting (NNeW) [23] applies a Voronoi tessellation of the space to help to determine the weights for each source sample. Hence, the weight on each sample is dependent on the number of target neighbour samples. As the final step, it is then applied Laplace smoothing for regularising the weights to avoid some getting exaggeratedly biased towards the test set.

Kernel Mean Matching (KMM) [24] directly estimates source weights without performing density estimation, as most algorithms in this scenario. Maximum Mean Discrepancy is the technique used to define the weights in a way that the divergence between the source and target domain is minimised.

## 4 Experiments

Employing two HAR datasets, PAMAP2 [25] and ANSAMO [26], we evaluated the four TL techniques (TCA, SA, KMM, NNeW) in various HAR scenarios having different mismatches of distribution between the source and target datasets. In total, we delineated five different experiment scenarios (AA-AGE, AA-POS, AP-ENV, PA-ENV, PP-POS), where we explore the user's age, sensor misplacement, and environment as shift-causes.

As for the employed techniques, we adopted three different learning approaches, TL, supervised and semi-supervised for comparison purposes. The two latter approaches are necessary since both serve as baseline methods helping to identify negative transfer in each TL approach.

Each TL technique had to be linked with a classifier, since the TL techniques' responsibility was only adjoining the source and target distributions, leaving classification task to the classifier. Therefore we incorporated each TL technique with four different classifiers (Logistic Regression, Bernoulli Naive Bayes, Decision Tree, Linear Support Vector Machine) to find the highest performance combination. Having found the best combinations, we posteriorly implemented a majority voting ENSemble (ENS) composed of KMM with Decision Tree, NNeW with Logistic Regression, and SA with Logistic Regression.

**Table 1.** Average accuracy obtained in each experiment scenario per technique.

| | **AA-AGE** | **AA-POS** | **AP-ENV** | **PA-ENV** | **PP-POS** |
|---|---|---|---|---|---|
| **SUP** | 0,52 | 0,25 | 0,65 | 0,71 | <u>0,57</u> |
| **SEMI** | 0,30 | 0,24 | 0,49 | 0,46 | 0,33 |
| **KMM** | 0,58 | 0,35 | <u>0,72</u> | 0,77 | 0,50 |
| **NNeW** | 0,57 | 0,34 | 0,68 | 0,78 | 0,55 |
| **SA** | 0,54 | 0,24 | 0,68 | 0,76 | <u>0,57</u> |
| **TCA** | 0,35 | 0,27 | 0,27 | 0,25 | 0,19 |
| **ENS** | <u>0,60</u> | <u>0,36</u> | 0,65 | <u>0,83</u> | 0,54 |

Regarding the semi-supervised approaches, we used two different iterative techniques, Label Propagation [27] and Label Spreading [28], where a model is trained with both the labelled source and unlabelled target samples.

As for the supervised approaches, we adopted the same four classifiers of TL, where a model is taught only with the original source dataset and tested with the target sample. These two approaches are necessary to help identify negative transfer in the TL results.

### 4.1 Age, Position, and Environment as Shift-Causes

The AA-AGE scenario focuses on exploring the users' age as the leading cause for distribution mismatch between datasets. Using only the ANSAMO dataset, the younger users compose this scenario's source dataset to help classify the activities of the older people unlabelled target dataset. In total we implemented 3 experiments, each trying to classify from different positions (ankle, chest, and wrist) 6 activities (bending, hopping, sitting, ascending stairs, and walking).

Regarding the AA-POS and PP-POS scenarios, the goal was to explore the misplacement of sensors as the leading cause for distribution mismatch between datasets. Each experiment consists of using a source dataset with data of a particular position (e.g. wrist) to help identify the activities of a target dataset captured from a different position (e.g. ankle). Two scenarios were devised, one using the ANSAMO, named AA-POS, and the other the PAMAP2 dataset, named PP-POS. Since the ANSAMO dataset had data captured from four different positions (ankle, chest, waist, wrist), and the PAMAP from three (ankle, chest, wrist), we devised 18 unique experiments. As for the activities, the AA-POS scenario had 7 (bending, hopping, running, sitting, ascending stairs, descending stairs, and walking) and PP-POS had 8 (lying, rope jumping, running, sitting, ascending stairs, descending stairs, standing, and walking).

As for the AP-ENV and PA-ENV scenarios, the goal was to explore the environment as the primary cause for distribution mismatch between datasets. Two scenarios were delineated, AP-ENV and PA-ENV, in which the former uses the ANSAMO dataset and PAMAP2 as source and target datasets, respectively. The latter employs the PAMAP2 dataset to leverage the activity recognition in the ANSAMO dataset. Every implemented experiment classifies 5 activities (running, sitting, ascending stairs, descending stairs, and walking) in 3 different positions (ankle, chest, wrist).

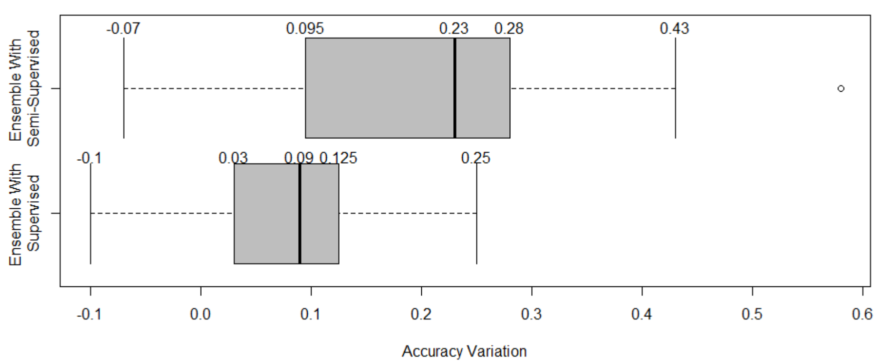
### 4.2 Results & Discussion

Altogether, we tested 22 different approaches in 5 experiment scenarios. As depicted in Table 1, in 3 of the 5 scenarios, the TL approaches had a significant transfer of knowledge particularly in the age and environment shift-cause scenarios. The ensemble approach, as expected, obtained the highest score gaining 8% in AA-AGE, 11% in AA-POS, and 12% in PA-ENV comparing with the supervised approach.

Furthermore, by comparing all the yielded results between the ensemble and baseline methods through Friedman Sum Rank test we are able to observe a significant difference between the three methods for the significance level of 1% (p-value = 6.707e-07, p-value $<$ 0.01). Additionally, with a Nemenyi post-hoc test [29] we are also able to observe that there is a highly significant difference between the ensemble and the semi-supervised approach and between the ensemble and the supervised approach for a significance level of 5% (p = 3.3e-07 and p = 0.022, respectively).

Moreover, Figure 1 has the comparative performance between ensemble and supervised models, and between ensemble and semi-supervised models. In overall, the results show the usefulness of applying TL in scenarios where there is a considerable distribution mismatch. While the ensemble has a median positive variance relative to the semi-supervised of 23%, it has a minor variation relative to the supervised approach with 9%.

**Fig. 1.** Accuracy's variation between the ensemble technique and the supervised approach, and between the ensemble technique and the semi-supervised approach.



## 5 Conclusions

Overall, we accomplished the delineated objectives. Our work not only contributes to the unsupervised domain adaptation field but to TL as well through the extensive performed survey. The yielded results indicate a highly significant difference between the TL ensemble approach with the semi-supervised approach and a less but significant difference with the supervised approach.

As for future work, we envision the study of this same domain but with an online learning approach. Having obtained in experiments negative transfer by some TL techniques, especially with TCA, it would be better employing multiple source datasets, which would likely have higher success since it increases the chance of discovering useful transferable knowledge between the source and target datasets.

## References


1. Malbut-Shennan, Katie, and Archie Young. "The physiology of physical performance and training in old age." Coronary artery disease 10.1 (1999): 37-42.
2. Sun, Baochen, Jiashi Feng, and Kate Saenko. "Return of frustratingly easy domain adaptation." AAAI. Vol. 6. No. 7. 2016.
3. Thrun, Sebastian, and Lorien Pratt, eds. Learning to learn. Springer Science & Business Media, 2012.
4. Calatroni, Alberto, Daniel Roggen, and Gerhard Tröster. "Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion." 8th International Conference on Networked Sensing Systems, 2011.
5. Khan, Md Abdullah Al Hafiz, and Nirmalya Roy. "TransAct: Transfer learning enabled activity recognition." 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (2017) 545-550.
6. Rokni, Seyed Ali, and Hassan Ghasemzadeh. "Autonomous Training of Activity Recognition Algorithms in Mobile Sensors: A Transfer Learning Approach in Context-Invariant Views." IEEE Transactions on Mobile Computing, 17.8 (2018): 1764-1777.
7. Krishnan, Narayanan Chatapuram. A computational framework for wearable accelerometer-based. Diss. PhD thesis, Arizona State University, 2010.
8. Kurz, Marc, et al. "Real-time transfer and evaluation of activity recognition capabilities in an opportunistic system." The Third International Conference on Adaptive and Self-Adaptive Systems and Applications (2011): 73-78.
9. Roggen, Daniel, et al. "The adarc pattern analysis architecture for adaptive human activity recognition systems." Journal of Ambient Intelligence and Humanized Computing 4.2 (2013): 169-186.
10. Zhao, Zhongtang, et al. "Cross-mobile elm based activity recognition." International Journal of Engineering and Industries 1.1 (2010): 30-38.
11. Deng, Wan-Yu, Qing-Hua Zheng, and Zhong-Min Wang. "Cross-person activity recognition using reduced kernel extreme learning machine." Neural Networks 53 (2014): 1-7.
12. Zhao, Zhongtang, et al. "Cross-people mobile-phone based activity recognition." IJCAI. 11.3 (2011).

13. Chen, Hailin, Shengping Cui, and Sebastian Li. "Application of Transfer Learning Approaches in Multimodal Wearable Human Activity Recognition." arXiv preprint arXiv:1707.02412 (2017).
14. Hachiya, Hirotaka, Masashi Sugiyama, and Naonori Ueda. "Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition." Neurocomputing 80 (2012): 93-101.
15. Diethe, Tom, Niall Twomey, and Peter Flach. "Active transfer learning for activity recognition." European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2016.
16. Fallahzadeh, Ramin, and Hassan Ghasemzadeh. "Personalization without user interruption: boosting activity recognition in new subjects using unlabelled data." Proceedings of the 8th International Conference on Cyber-Physical Systems. ACM, 2017.
17. Laasonen, Kari, Mika Raento, and Hannu Toivonen. "Adaptive on-device location recognition." International Conference on Pervasive Computing. Springer, Berlin, Heidelberg, 2004.
18. Blanke, Ulf, and Bernt Schiele. "Remember and transfer what you have learned-recognizing composite activities based on activity spotting." Wearable Computers (ISWC), 2010 International Symposium on. IEEE, 2010.
19. Margolis, Anna. "A literature review of domain adaptation with unlabeled data." Tec. Report (2011): 1-42.
20. Pan, Sinno Jialin, et al. "Domain adaptation via transfer component analysis." IEEE Transactions on Neural Networks 22.2 (2011): 199-210.
21. Pan, Sinno Jialin, James T. Kwok, and Qiang Yang. "Transfer Learning via Dimensionality Reduction." AAAI. Vol. 8. 2008.
22. Fernando, Basura, et al. "Subspace alignment for domain adaptation." arXiv preprint arXiv:1409.5241 (2014).
23. Loog, Marco. "Nearest neighbor-based importance weighting." Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on. IEEE, 2012.
24. Huang, Jiayuan, et al. "Correcting sample selection bias by unlabeled data." Advances in neural information processing systems. 2007.
25. Santoyo-Ramón, José Antonio, Eduardo Casilari, and José Manuel Cano-García. "Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection with supervised learning." Sensors 18.4 (2018): 1155.
26. Reiss, Attila, and Didier Stricker. "Creating and benchmarking a new dataset for physical activity monitoring." Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2012.
27. Zhu, Xiaojin, and Zoubin Ghahramani. "Learning from labeled and unlabeled data with label propagation." (2002): 1.
28. Zhou, Denny, et al. "Learning with local and global consistency." Advances in neural information processing systems. 2004.
29. Demšar, Janez. "Statistical comparisons of classifiers over multiple data sets." Journal of Machine learning research 7.Jan (2006): 1-30.