Transfer Learning in Body Sensor Networks using Ensembles of Randomised Trees

Pierluigi Casale
SPS Group-TU Eindhoven and Holst Center/IMEC
Eindhoven, The Netherlands
Email: pierluigi.casale@imec-nl.nl

Marco Altini
Holst Center/IMEC
Eindhoven, The Netherlands
Email: marco.altini@imec-nl.nl

Oliver Amft
University of Passau
Passau, Germany
olver.amft@uni-passau.de

***Abstract*—In this work we investigate the process of transferring the activity recognition models of the nodes of a Body Sensor Network and we proposed a methodology that supports and makes the transferring possible. The methodology, based on a collaborative training strategy, makes use of classifier ensembles of randomised trees that allow to generate activity recognition models able to be successfully transferred through the nodes of the network. Experimental results evaluated on 17 subjects with a network of 5 wearable nodes with 5 everyday life activities show that the recognition models can be transferred to a new untrained node replacing a node previously present in the network without a significant loss in the recognition performance. Moreover, the models achieve good recognition performance in nodes located in previously unknown positions.**

I. Introduction

Supervised Machine Learning techniques are widely used in the recognition process of physical activities using Wearable Sensors and Body Sensor Networks (BSN). Nevertheless, these data-driven methodologies can represent a limitation when the retraining of the activity recognition models of the network is required. For the aim of example, consider the situation where users need to replace a broken node or they want to relocate a node previously used on a specific position in a new position, e.g., from the chest to the waist. In both cases, the possibility to avoid a retraining still maintaining the recognition functionality of the node represents an appealing capability from the application point of view. In this work, we investigate the process of directly transfer the recognition models, i.e., the classifiers running in the nodes of a BSN, to a new node previously untrained in order to avoid the training phase after the deployment of the BSN. This transferring process, known as Transfer Learning [1], is accomplished in this work through the combination of a collaborative training strategy and the use of classifiers ensemble based on randomised trees. Using the collaborative training strategy, a limited amount of data shared between all the nodes of the BSN is used in combination to the data of the node for training a classifier ensemble before the deployment of the network. This ensemble, while still able to provide high recognition performance for the node, contains a degree of redundancy helpful during the transferring process. When algorithms based on randomized trees like Bagging [2], Random Forest [3] or Rotation Forest [4] are considered, the proposed strategy allows to learn ensembles that can be transferred through the nodes of the network and are able to recognize the activities sensed by nodes positioned at different locations. The random transformations that these algorithms apply to the training set are beneficial in the transferring process of the recognition models through the nodes placed at different position. Since the performance of the recognition models depend on the training data, the amount of data shared at training time is a quantity that needs to be taken into account in order to find a good trade-off between the performance of the node and the performance of the transferred classifiers.

The methodology has been applied in situations where a broken node needs to be replaced by a new node located in the same position (replacement scenario) and a node already present in the network is relocated to a previously unknown position (relocation scenario). A dataset collected using a BSN of 5 wearable nodes has been used for evaluating the methodology in both scenarios, using sensor data from 17 subjects. In both scenarios considered, experimental results show that the recognition model of a node can be transferred and high recognition performance can be obtained in the replacement scenario and good recognition performance are achieved in the relocation scenario. Results have been validated using several K-folds cross-validation protocols in order to test the performance of the methodology when different amount of data are shared between nodes.

This paper makes the following contributions:

1) We define a collaborative training strategy that, by sharing a limited amount of data between nodes, allows to generate classifier ensembles that, besides achieving high recognition performance in the node, contain a degree of redundancy useful in the transferring of the recognition models through the nodes of the network.
2) We use ensemble learning algorithms based on randomised trees that, using random transformations on the training set generated with the collaborative training strategy, allows the direct transferring of the recognition models through the nodes of the network.

II. Transfer Learning and its application in BSN

Classifier ensembles have been already considered as a mechanism for Transfer Learning ([5], [6], [7], [8]). Of particular interest, the work of Kamishima et al. [9] applied a bagging approach for transfer the learning capabilities of a model through different domains. In their work, an high number of trees was learned on data from both source and target domains and a pruned version of the final ensemble was used to predict examples of the target domain. The pruning step was used in order to avoid the decreasing of performance due to negative transfer. Although authors used training examples from both source and target domains, no experiments were conducted to quantify the amount of data needed to achieve a reasonable accuracy of the transferred classifier. Moreover, natural extensions of Bagging like Random Forest or Rotation Forest were not considered.

In activity recognition, Calatroni et al. [10] showed the feasibility of transfer learning in BSN and used a transferring approach to evaluate the classification performance of transferring to a new deployed sensor in the BSN. The transfer approach was based on a Teacher-Learner paradigm where all the pairs of nodes were considered as possible teachers and learners. The direct transferring of the classifiers in the nodes was compared to a *system-supervised* approach where the teacher node provides labels to the learner and a new classifier was thus trained in the node. Good results were obtained using the system-supervised approach specially when the transferring is done between nodes located in similar body part. Distance-based classifiers like Nearest Centroid Classifier, k-NN and SVM were used. Although we take into account the direct transfer of classifiers between the nodes, in our work we do not consider the possibility to train a classifier in the node. In order to avoid this training, we use the collaborative training strategy and ensemble methods that provide transformations of the feature space that are potentially beneficial in the classification of activities using nodes placed at different positions. Blanke and Schiele [11] applied transfer learning methodologies for the composition of complex activities from a time ordered sequence of events. In their work, the authors transferred the activity events that were shared between similar composite activities minimizing the training effort for new events. Results obtained show that their approach achieves good performance in the recognition of composite activities, also when different application domains are considered. Although the methodology does not use a direct transferring approach, the work shows that transfer learning is feasible when a level of abstractions is considered in the learning mechanism. This consideration is also reported in van Kasteren et al. [12] where a level of abstraction is used to map features with the aim of transferring activity recognition capabilities between different home scenarios.

III. Methodology

In this section, the proposed methodology is described. The learning algorithms considered are exclusively ensemble of randomized trees like Bagging, Random Forest and Rotation Forest. The trees composing the ensembles will be referred as *base classifiers*. The number of base classifiers in the ensemble T is a parameter defined at training time.

Given a BSN constituted by the set of sensor nodes

$$\mathbf{S} = \{S_i\},\ i = 1 \ldots N \tag{1}$$

the correspondent set of data

$$\mathbf{D_{in}} = \{{D_i}^I\},\ i = 1 \ldots N \tag{2}$$

is considered where each ${D_i}^I$ represents the data of node S_i. For each S_i, an additional set of data $\mathbf{D_i}^{\mathbf{O}}$ is defined as the composition of subsets of data coming from the other nodes of the network.

$$\mathbf{D_i^o} = \{{D'}_j \subset {D_j}^I\},\ \forall\ j = 1 \ldots N,\ j \neq i \tag{3}$$

For each node S_i, an ensemble H_i is trained using the dataset D_i as defined in Eq. 4, constituted by the composition of ${D_i}^I$ and $\mathbf{D_i^o}$.

$$D_i = {D_i}^I \bigcup \mathbf{D_i^o} \tag{4}$$

The classifier ensemble obtained in this step is used to classify activities on the node S_i during the normal operation of the node. Two considerations need to be outlined from the previous definitions. At first, the dataset ${D_j}'$ defined in Eq. 3 are strictly contained in ${D_j}^I$, i.e., only a subset of data from the other nodes should be considered for training the classifier ensemble for the node S_i. This fact ensures that the recognition performance will be kept high on S_i since H_i mostly learns from data coming from S_i. Secondly, the union defined in Eq. 4 does not refer to aggregate data at features level but to append a small amount of data coming from all the other nodes to the dataset gathered from S_i. In other words, the final training set of the classifier H_i contains a big amount of data coming from S_i and a small amount of *spurious* data coming from all the other nodes of the network. The aim of using this training set is to let the ensemble learning algorithm to learn some base classifiers that are barely able to classify data from the other nodes. These classifiers will not influence the final recognition performance of the classifier ensemble but they will provide a degree of redundancy that will be useful during the transferring. In case the node S_i has a failure and needs to be replaced, situation identified as *Replacement Scenario*, all the nodes of the network transfer their classifiers H_j to the new node ${S_i}^{new}$. The new ensemble for ${S_i}^{new}$ will be then constituted by the simple aggregation of all the ensembles H_j from the other nodes.

$${H_i}^{new} = \bigcup_{j \neq i} H_j\ ,\ j = 1 \ldots N \tag{5}$$

In case the node S_i is relocated to a different position, situation identified as *Relocation Scenario*, a model ${H_i}^{rel}$ assembled as defined in Eq. 5 can be used. Nevertheless, as noticed in [9], the use of a no pruned ensemble may generally yield to a performance decaying due to negative transfer. In this case, the ensemble ${H_i}^{rel}$ is generated by selecting only the base classifiers ${h_j}^t$ that perform better than a predefined accuracy α. In this case, the corresponding labels should also be transferred in order to check the performance of the single base classifiers transferred.

$${H_N}^{rel} = \bigcup_{j=1\ldots N, t=1\ldots T} {h_j}^t\ s.t.\ accuracy({h_j}^t) > \alpha \tag{6}$$

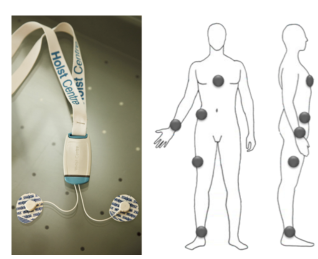

Fig. 1: ECG Necklace and on-body nodes positioning for the considered dataset.

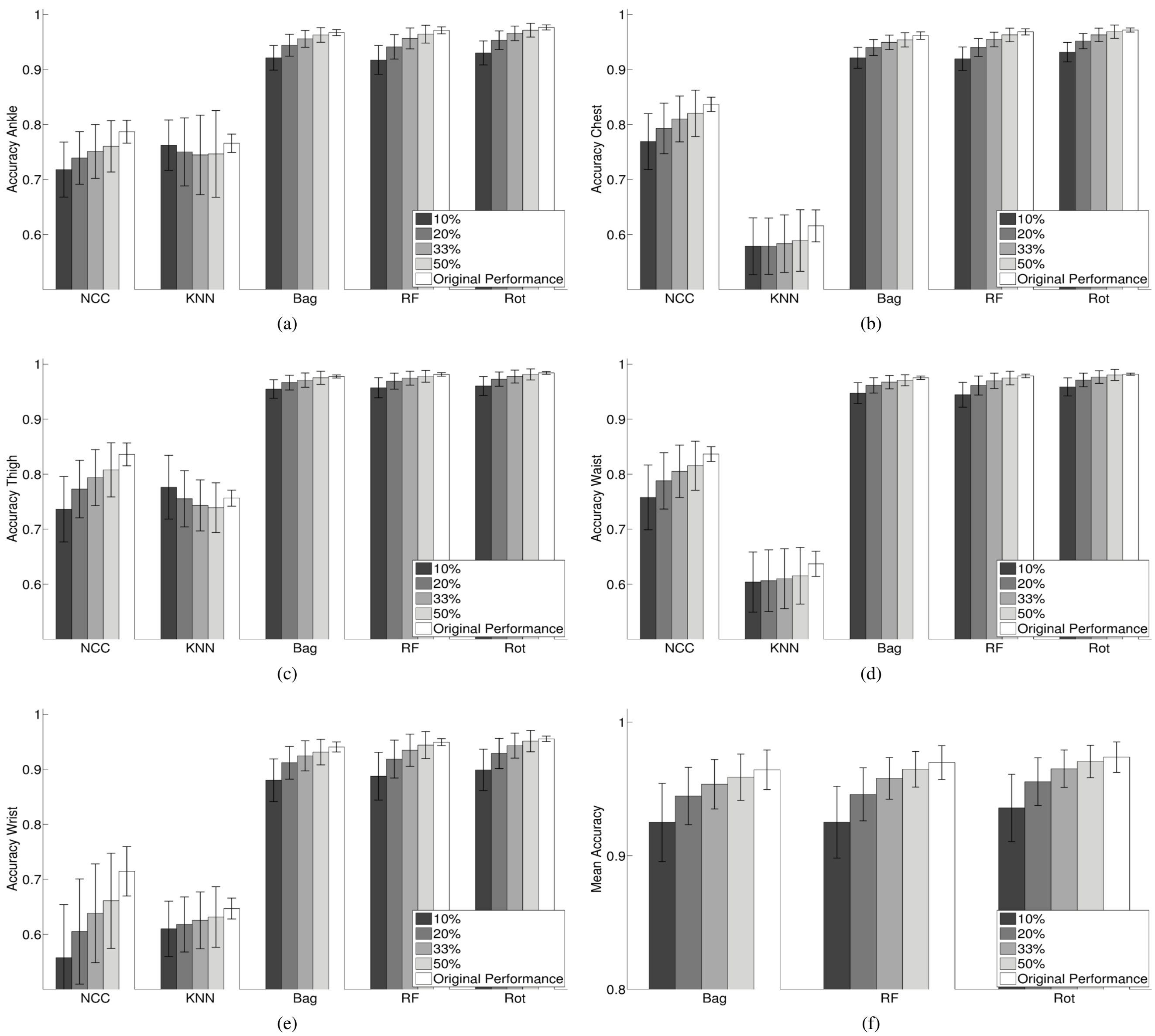

Fig. 2: Classification Accuracy obtained in the Replacement Scenario for the nodes on (a) Ankle, (b) Chest,(c) Thigh, (d) Waist and (e) Wrist, using different amounts of data shared between nodes. Results are averaged on all the subjects. In (f) the accuracy of Bag, RF and Rot averaged on subjects and nodes is shown. The recognition performance of classifier ensembles methods are within 5% from the original performance.

IV. Evaluation Dataset and Methods

Experiments have been conducted on a dataset gathered using a BSN constituted by 5 sensors nodes with ECG and accelerometer from 17 participants. The dataset has been previously used for research on energy expenditure using data from a single node [13]. A wide range of sedentary, lifestyle and sport activities have been considered and manually annotated on two levels of granularity. For the purpose of this work five macro-activities are considered, i.e., *Household Activities*, *Resting*, *Sitting*, *Sport* and *Walking*. The sensor platform used was the ECG Necklace. Nodes have been positioned on the chest, the dominant ankle, the dominant thigh, the dominant wrist and the waist at the right hip as shown in Fig. 1. The nodes have been attached to the body using elastic bands and synchronized over a wireless network. Accelerometers have been configured to acquire data at 64 Hz. A large set of features have been computed including statistics on the temporal domain and features from the spectral domain obtaining a 54-dimensional features vector for each node. The set of features includes mean, standard deviation, skewness and kurtosis for each acceleration axis and magnitude, correlations between each pairwise axis combination, entropy, signal spectral amplitude, power and frequency.

Four K-folds stratified cross-validation schemes have been considered, with $K = \{10, 5, 3, 2\}$. These schemes allow to test the methodology with subsets of 10%, 20%, 33% and 50% of data shared between nodes. For the Replacement Scenario, the training dataset D_i for sensor S_i has been constructed

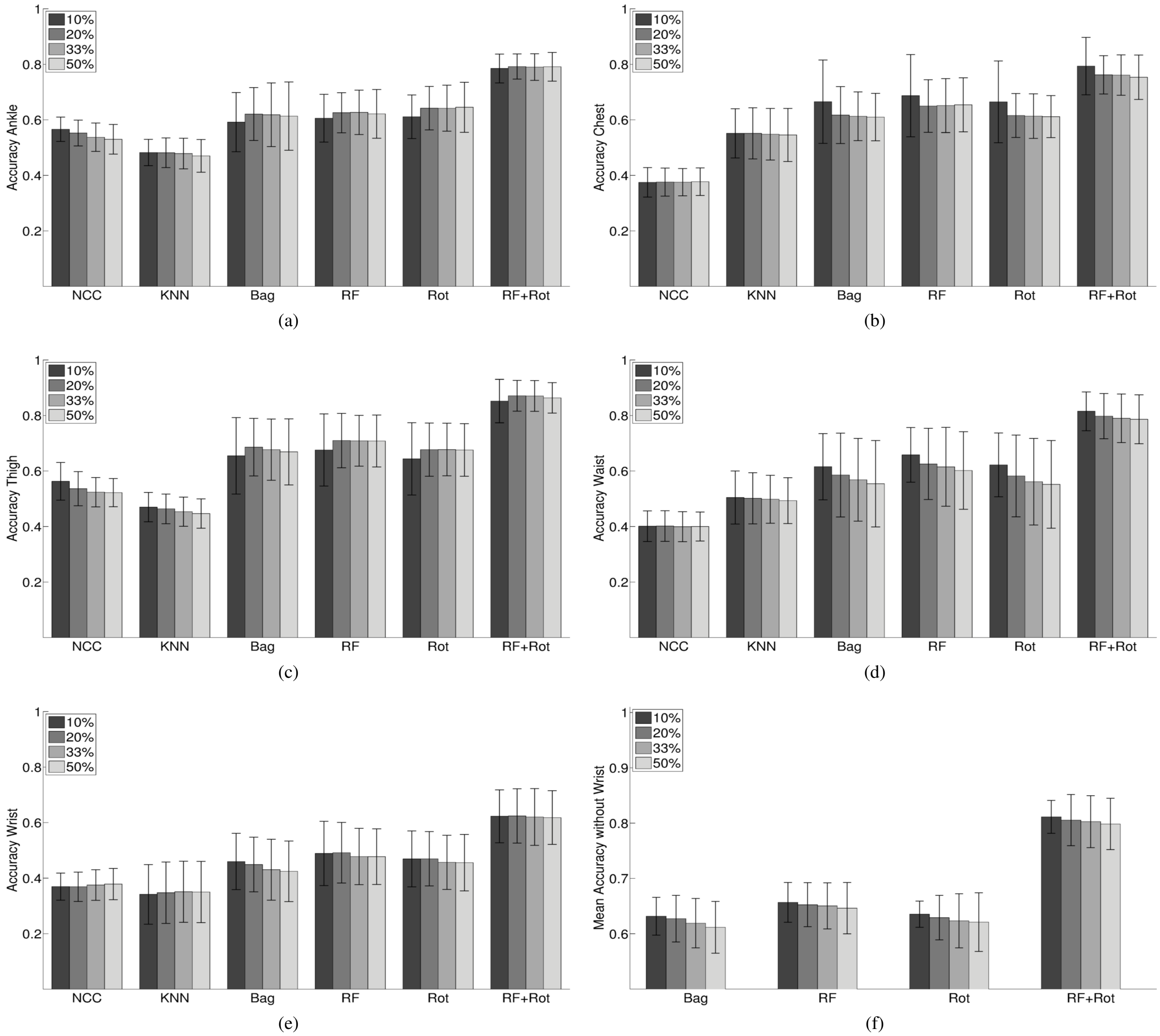

Fig. 3: Classification Accuracy obtained in the Relocation Scenario for the nodes on (a) Ankle, (b) Chest,(c) Thigh, (d) Waist and (e) Wrist using different amounts of data shared between nodes. Results are averaged on all the subjects. In (f), the accuracy averaged on all the nodes except Wrist is reported. Better recognition performance is achieved when a lower amount of data is shared.

using K-1 folds from sensor S_i and the K-th fold from all the other nodes. Test is then executed on the K-th fold for sensor S_i and the K-1 folds for the remaining nodes. For the Relocation Scenario, a Leave-One-Sensor-Out protocol has been used. Training data have been assembled using the same protocol as before. For each experiment, 2 runs of cross-validation have been considered. The methodology has been applied with Bagging (Bag), Random Forest (RF) and Rotation Forest (Rot) and compared with Nearest Centroid Classifier (NCC) and k-Nearest Neighbors (kNN), being the best performing methods in [10]. The number of trees in the ensembles has been arbitrarily set to 31 and the number of nearest neighbors in kNN set to 11. Classification accuracy has been adopted as performance measure for simplicity purposes. In order to avoid interpersonal variability between the activities performed, experiments have been conducted separately for each subject. Final results have been averaged on runs and subjects.

V. Experimental Results

Results reported in Fig. 2 show the accuracy obtained in the Replacement Scenario for all the nodes using different subsets of data shared. The original accuracy obtained on the nodes is also reported. Rot is the best performing method but all the ensemble methods provide results that are within the 5% from the original accuracy of the node. This fact is highlighted in Fig. 2(f) where the accuracy averaged on all the nodes is reported. As expected, performance improves as the amount of

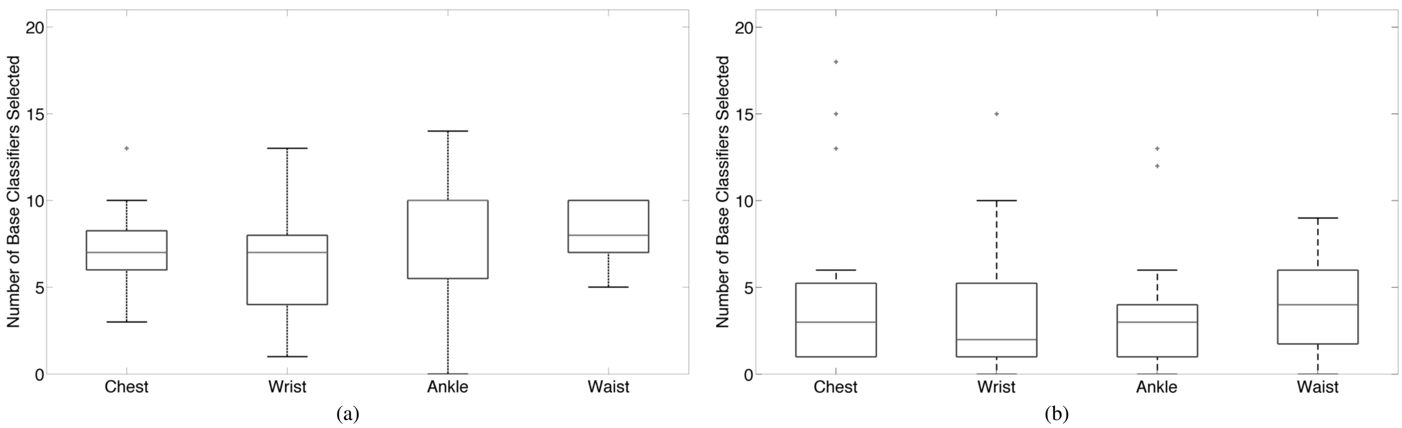

Fig. 4: Composition of the classifier ensemble in the Thigh node for (a) Replacement Scenario and (b) Relocation Scenario, computed on all the participants. The classifier ensemble in the Thigh node is equally composed by base classifiers coming from all the other nodes of the network.

data shared between nodes increases. Two interesting aspects are worth to be noted. With subsets of 33% of data shared, the transfer of Rot achieves recognition performance as good as the original performance achieved by Bag. Using a 50% of data shared, Rot achieve slightly better performance than Bag and results comparable with the original performance of RF in the node. Nevertheless, all the ensembles methods achieve very good recognition performance even when only a 10% of data is shared between the nodes. Considering the high values of accuracy obtained by these methods , these narrow differences in performance should be considered mainly due to the randomness of the cross-validation process.

Classification results for the Relocation Scenario are reported in Fig. 3. Although higher than NCC and KNN, the accuracy obtained by the transferred ensembles in this scenario does not provide a satisfactory result. Nevertheless, in the majority of the experiments, the accuracy obtained is approximately higher than 60%, much higher than the baseline performance of 20% provided by the random guess when 5 classes are considered. In this scenario, RF is the best performing method. An considerable improvement is provided when the final ensemble is obtained by combining base classifiers from both RF and Rot. This heterogeneous ensemble achieves in many cases a recognition accuracy of 80% or higher for the majority of the experiments considered. This fact is shown in Fig. 3(f) where the accuracy of the ensembles averaged on all the nodes is reported. The node of the Wrist has been not considered in the computation of the average. It is worth to note that, in most of the experiments, higher accuracy is achieved when a lower amount of data is shared between nodes. In this scenario, sharing bigger amount of data seems to be not beneficial to the transferring process.

In both scenarios, best results are obtained on the Thigh node. In order to verify if this effect is due to the presence of a node that show some similarity with the node positioned on the thigh, Fig. 4 plots statistics about the composition of the ensemble in this node for both scenarios. The boxplots, separated by nodes, are computed over all the participants. For both scenarios, the ensemble of the thigh node is constituted by base classifiers equally selected from the other nodes and no node seems to mainly contribute to the composition of the ensemble with a predominant number of base classifiers. This results is expected as a consequence of the collaborative training strategy. For each ensemble trained on the nodes of the network, there exist some base classifiers that learn from the combination of data coming from different nodes. These base classifiers do not depend on the particular position of the node and they are used to create the final ensemble. From the boxplot, an interesting situation also emerge. The average number of base classifiers in the Replacement scenario, shown in Fig. 4(a), is higher than the average number of base classifiers used in the Relocation scenario, shown in Fig. 4(b). This behavior is a consequence of the selection process defined in Eq. 6 since only the best performing classifiers are retained for creating the classifier ensemble in the Thigh node.

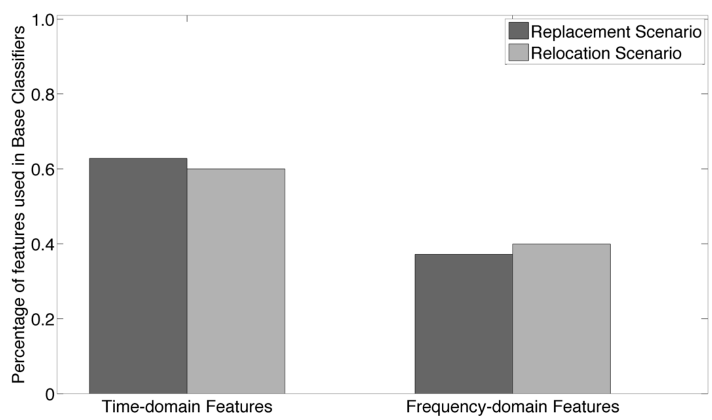

Fig. 5: Percentage of Features used as split points in the learning of base classifiers. Trees mostly use time-domain features in the learning process.

An in-depth look into the features that base classifiers use in the recognition process needs to be addressed. In particular, it is worth to verify if the good performance obtained after transferring may be due to the use of specific features that characterize some periodicity present in the set of activities used in the considered dataset. Fig. 5 show the percentage of features used in the learning process of the base classifiers, averaged on all the nodes and all the subjects, separated in time-domain and frequency-domain features. The percentage is shown for both Replacement and Relocation scenarios. In both scenarios, time-domain features are mainly used although a small increase of frequency-domain features is noticed in the Relocation Scenario. It turns out that the results of the transferring process is not biased by the use of a particular set of features that specifically describe the frequency content of the activities.

VI. Discussions

Although the collaborative training has been used for training NCC and kNN, these two classifiers do not provide the same level of performance achieved by Bag, RF and Rot. It turns out that the capability to successfully transfers classifiers through the nodes of the networks is due to the combination of sharing training data between the nodes and using ensemble of randomised trees. It is our intuition that the transformations performed on the features space, particularly by RF and Rot, provide an abstraction level that allow to model a set of activities disregarding the particular location of the nodes that sense the activity. In particular for the Replacing Scenario, the combination of different subsets of features and the use of rotated versions of them, as respectively provided by Random Forest and Rotation Forest, allows to generated a new ensemble that is able to provide acceptable recognition capabilities for the node sensing the activities in a completely new position.

For the Relocation Scenarios, labels should be transferred in order to check the performance of the classifiers on the new node. Although this process could look similar to the *system supervised* approach used in [10], there exists a substantial difference. In our approach, labels are transferred only for testing the base classifiers and no training is performed in the node. Since the base classifiers are constituted by decision trees, just a sequence of *if-then* statements need to be executed in the node. The transferring of labels can also be beneficial in the Replacement scenario, where labels are not required to be sent. In fact, in this scenario the final classifier is aggregated using all the base classifiers coming for the nodes of the network, generating a very large ensemble classifier. If labels are transferred, the ensemble can be pruned and only the best performing classifiers can be retained. Results obtained using pruned version of the classifier ensemble in the Replacement scenario show that there is no loss in the recognition performance in the node.

VII. Conclusions

In this work, the capability of transferring the activity recognition models in Body Sensor Networks has been investigated. In addition, a methodology that supports and makes the transferring possible has been proposed. The methodology is based on a collaborative training strategy and makes use of classifier ensembles of randomised trees. Both elements allow to generate recognition models that can be successfully transferred through the nodes of the Body Sensor Network placed on different body locations. Experimental results obtained on a dataset of 5 nodes with activities collected from 17 participants show that the methodology is able to generate classifier ensembles that, by direct transferring, can be used in a new untrained node that replaces a node previously present in the network without a significant loss in the recognition performance. Moreover, these classifier ensemble can also provide good recognition performance in nodes located in previously unknown positions. Results have also shown that the performance obtained does not depend on the presence of nodes that are correlated or the use of a specific set of features. The ensembles generated in the node after transferring are composed by an homogeneous combination of base classifiers coming from all the nodes present in the network. Future works plan to validate the approach on multiple datasets with a variable number of nodes and assessing the performance of the methodology taking into account the interpersonal variability in the activities considered. A study on the dimensionality of the ensemble needs also to be addressed as a trade-off between classification accuracy and redundancy needed for the transferring. Finally, being the amount of data shared at training time an important factor with strong influence in the practical application of the methodology, further studies should be addressed in order to assess the recognition performance that the methodology can achieve when very low amount of data ($< 10\%$) is shared between nodes at training time.

Acknowledgment

This work was supported by the EU FP7 Marie Curie Network iCareNet under grant number 264738.

References

[1] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[2] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[3] ——, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, “Rotation forest: A new classifier ensemble method,” *IEEE TPAMI*, vol. 28, no. 10, pp. 1619–1630, 2006.

[5] S. Mei, “Probability weighted ensemble transfer learning for predicting interactions between hiv-1 and human proteins.” *PLoS One*, vol. 8, no. 11, p. e79606, 2013.

[6] A. Acharya, E. R. Hruschka, J. Ghosh, and S. Acharyya, “Transfer learning with cluster ensembles,” in *ICML Unsupervised and Transfer Learning*, vol. 27, 2012, pp. 123–132.

[7] L. Jiang, K. Zhang, and G. Allen, “Transferred correlation learning: An incremental scheme for neural network ensembles.” in *IJCNN*, 2010, pp. 1–8.

[8] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, “Boosting for transfer learning,” ser. ICML ’07, 2007, pp. 193–200.

[9] T. Kamishima, M. Hamasaki, and S. Akaho, “Trbagg: A simple transfer learning method and its application to personalization in collaborative tagging,” in *ICDM*, 2009, pp. 219–228.

[10] A. Calatroni, D. Roggen, and G. Tröster, “Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion,” in *INSS’11)*, Jun. 2011.

[11] U. Blanke and B. Schiele, “Remember and transfer what you have learned - recognizing composite activities based on activity spotting,” in *ISWC’10*, 2010, pp. 1–8.

[12] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, “Transferring knowledge of activity recognition across sensor networks,” in *Pervasive’10*, 2010, pp. 283–300.

[13] S. Chen, J. Lach, O. Amft, M. Altini, and J. Penders, “Unsupervised activity clustering to estimate energy expenditure with a single body sensor.” in *BSN2013*, 2013, pp. 1–6.