Human Activity Recognition from Sensor Data

Sheshera Mysore
University of Massachusetts Amherst
smysore@cs.umass.edu

Anna Fariha
University of Massachusetts Amherst
afariha@cs.umass.edu

1 Introduction

Human Activity Recognition (HAR) refers to the problem of automatically identifying human activities by learning from data collected either from the individual or from the environment surrounding that individual. The form of the data might be inertial measurements, physiological signals from devices like wearable sensors or those such as image, audio and video data collected from environmental devices/sensors. HAR has been a well studied problem for its potential capability to mine human behavioral patterns and for monitoring human activity for security and health purposes.

One real-life scenario where human activity recognition would help is monitoring heart rate in patients of chronic heart disease as they perform different activities on a daily basis. The heart rate of a patient might need to be monitored to observe how his heart rate changes with different activities. It is typically cumbersome for the patient to report all his activities manually even as his heart rate is automatically monitored. An automatic activity recognition solution would be of value in a scenario of this nature.

In this project, we focus on wearable sensor based single user activity recognition and the application of transfer learning to this problem. The activity recognition problem can be formalized as follows: Given a set $\mathcal{W} = \{W_1, \ldots, W_m\}$ of m equally sized windows, a set of time series measured from k sensors, $S = \{S_i, \ldots, S_k\}$ where each S_i expands over several windows and a set of target activity class labels $\mathcal{A} = \{A_1, \ldots, A_n\}$, the task is to find a mapping function $f : \mathcal{W} \to \mathcal{A}$ such that $f(W_i)$ is as close as possible to the true activity performed during window W_i. Specifically, each window W_i consists of k time series chunks $S_{i,1}, \ldots, S_{i,k}$ that will be used as features to predict that window's activity class label.

In the context of activity recognition it is typically hard for conventional classification approaches to transfer a model learned on data from one individual to a recognition task on a different individual. For example, although sharing similarities, the sensor readings from an elderly woman performing a certain set of activities will be significantly different from that from a teenage girl performing the same set of activities. For this reason we also attempt a transfer of knowledge between two groups of different individuals. Transfer learning can be described as a set of methods which allow transfer of some form of knowledge learned from a source dataset to a target dataset which has significantly different properties from the source dataset.

The novelty of our project is two fold. First, we use the dataset[1] for classification of ambulatory activities (e.g., walking, sitting) using classifiers that have not so far been applied on this dataset and present a comparison across their performances. Second, we modify a transfer learning approach described by Chattopadhyay et al. [3] for application to the activity recognition problem. In general, HAR is an extremely challenging problem to generalize across datasets due to the variations in data collection setting, properties of individuals and data collection environment. Due to this fact, each dataset poses a different HAR problem.

We have experimented on the dataset using a number of different classifiers which make use of features extracted from accelerometer and gyroscope sensor data. The evaluation is done in two ways, (1) training and testing limited to one individual, and (2) training on a different group of individuals and testing on another individual. Using conventional supervised learning approach,

[1] Real world HAR: http://sensor.informatik.uni-mannheim.de/#dataset_realworld

we found gradient boosting performing best on achieving per individual prediction accuracy, while logistic regression performed best when the experimentation was leave one individual out cross validation. Additionally we have applied a transfer learning approach proposed by Chattopadhyay et al. [3] to our problem, proposed a modification to this algorithm and demonstrated it to perform better than the original method and a baseline method. Our experiments also indicate that domain adaptation only helps improve performance for the case where the source and target domains are significantly different; this is something which one would expect.

2 Related Work

In this section we first highlight few works by the authors of the dataset we use and some additional works that have been done using this dataset. Next we provide a literature survey of approaches to solve general human activity recognition problem. This is followed by a presentation of few domain adaptation/transfer learning approaches for HAR.

This project makes use of a dataset which was collected and made available by Sztyler et al. as part of work presented in the context of on-body sensor localization [12]. In this work the authors presented an approach to identify the position of the sensor on user's body, they also demonstrated that knowing the position of the sensor consistently improves the recognition performance. One recent work [11] have involved the proposal of an approach to reduce the effort needed to tune the performance of a trained model to a new individual. Here they proposed an approach which involves training a random forest classifier on individuals having similar physical characteristics to that of the target user and then continuing to train the model in an active manner as new data and feedback from the user becomes available. This solution is very similar in its goals to the goals of domain adaptation method we experiment with.

Pirttikangas et al. [6] worked on feature selection for activity recognition from wearable sensors on different parts of the body and found forward-backward sequential search algorithm most effective for best feature selection. Zhang et al. [14] proposed a human activity recognition framework based on feature selection where data sources were wearable multimodal sensors. The work proposed design of "physical features" and found through experimentation that those were among the top features in the proposed single-layer feature selection framework. Bao et al [2] proposed an offline method to identify user activity from data acquired from bi-axial accelerometers. Maurer et al. [5] proposed a real-time activity identification approach using time domain features collected from wearable sensors on different parts of the body. To highlight effect of different type of wearable sensors, Stisen et al. [9] investigated effects of heterogeneities among different sensors in smart-phones, smart-watches from different manufacturers on HAR problem.

Shoaib et al. [8] argued that smart-phone based sensors are not suitable for gesture based activity recognition (e.g., smoking, eating) and proposed a method that leverages combination of wrist-worn and pocket sensors to solve HAR problem more accurately. The authors also performed experimentation on different window sizes for observing accuracy of classifying activities. They made use of three classifiers – Naive Bayes, KNN and decision tree, for training and building model. They performed 10-fold cross-validation and measured performance using classification accuracy, precision and F-measure.

Awais et al. [1] proposed method using KNN to solve ADL (activities of daily living) problem from inertial sensor (accelerometer and gyroscope) data. Reyes-Ortiz et al. [7] proposed scheme to solve transition-aware HAR problem that takes transition between activities into consideration. Vail et al. [13] propose the use of Conditional Random Fields for the activity recognition problem and show it to perform better than Hidden Markov Model based approaches. Next we present some of the work which applies transfer learning to the HAR problem.

Cook et al. [4] broadly categorize the kinds of knowledge which can be transferred between source and target domains into 4 categories; (1) instance transfer – where data from the source domain is reused to train a classifier for the target domain, (2) feature-representation transfer – where differences in the source and target domain feature spaces are removed by some means, (3) parameter transfer – where parameters are shared between the source and target domains and (4) relational-knowledge transfer – where knowledge of relationships between samples in the source domain to the target domain. The transfer learning approaches we attempt here may roughly be placed in the instance and feature-representation transfer categories.

3 Dataset

The publicly available dataset by Sztyler et al. [10] was used for all our experiments. The dataset consists of measurements from 15 individuals performing 8 ambulatory tasks – walking, running, sitting, standing, lying, climbing upstairs, climbing downstairs and jumping. Due to irregular data for certain individuals we dropped data of 4 individuals (4,6,7,14) from the dataset, resulting in a dataset of 11 persons. Measurements are collected using 6 types of sensors – accelerometer, gyroscope, light, magnetic field, GPS and sound level. Measurements are collected simultaneously from the sensors placed on 7 body positions – chest, forearm, head, shin, thigh, upper arm and waist. In our experiments we only use accelerometer and gyroscope data measured from the thigh position. We decided to restrict ourselves to these measurements in order to simulate the readings as being obtained from a smart-phone in an individuals pocket. We also restrict ourselves to accelerometer and gyroscope readings since these two data sources were found by Zhang et al. [14] to be good for recognizing ambulatory tasks. Each of these two sensors contain 3 sets of readings along the x, y and z axis aligned to time.

Table 1: Number of data rows for each individual when window size = 2 seconds

Test Person	1	2	3	5	8	9	10	11	12	13	15
#Data rows	2174	2052	2242	2426	2416	2215	2069	2172	2091	2159	2195

Data Distribution: Table 1 shows the number of samples for each individual on application of a 2 second window to extract features. Each individual performed each activity for roughly 10 minutes (with the only exception being jumping for 1.7 minutes) this ensures that the class labels are equally distributed.

Table 2: Grouping individuals by gender, fitness and build

Build	Slim	Stocky	Slim	Stocky
Gender	Male		Female	
Fitness #1	14	9	6	
Fitness #2	2,4,7	10	13	
Fitness #3		3	1,12	
Fitness #4	5		11	
Fitness #5				8, 15

Subject Properties: Since the we attempt a transfer between subjects, we elaborate on the nature of these subjects from whom the data was collected. The subjects consisted of 8 males and 7 females, aged 31.9 ± 12.4, height 173.1 ± 6.9 cm and weighing 74.1 ± 13.8 kilograms. Since Sztyler et al. attempted a very similar cross-user learning on the same dataset in a recent paper [11] they also presented some more details about the subjects in their work. These details pertain to the physique of the subjects and are important in understanding the individuals who might have similar movement patterns for the same activity. This information is summarized in Table 3. Individuals close together in the table are expected to have similar movement patterns. We elaborate further on this aspect in section 6.

4 Methodology

In this section we describe the conventional supervised learning approach to solve the HAR problem, followed by the transfer learning based approach.

4.1 Conventional Approach

In this section, we describe how features were extracted from raw features which are time series data. We then describe few classifiers those were applied to build model based on the extracted features of the dataset.

4.1.1 Features

Windowing Strategy: In all of our feature extraction steps we made use of a 2 second window. However, we also experiment with different window sizes and reported the outcomes in section 5. For a given window, we consider samples from within that window to extract a given feature.

Features Extraction and Feature Engineering: We worked with 3 raw time-series readings (x_a, y_a, z_a) from the accelerometer and 3 raw time-series readings (x_g, y_g, z_g) from the gyroscope. Both of these sensors represent physical features. We also computed three composite features – (1) roll (r_a), (2) pitch (p_a) and (3) tilt angle (t_a) from these raw readings. As the raw data is time series data and that cannot be directly used as features, we extracted summary statistics over the time series. For the basic 6 time-series readings, $(x_a, y_a, z_a, x_g, y_g, z_g)$ we calculated 8 statistical properties – (1) Minimum, (2) Maximum, (3) Mean, (4) SMA (Signal Magnitude Area): Normalized integral of all values, (5) STD (Standard Deviation), (6) Variance, (7) Kurtosis and (8) Skew. For the three composite features we only calculate the mean. We also run experiments in 3 different feature settings – (1) only accelerometer data, (2) both accelerometer and gyroscope data and (3) adding 3 composite features.

4.1.2 Classifiers

We tried a number of different classifiers using different feature settings and evaluation settings. We mostly focused on one linear classifier – *logistic regression* and one ensemble method – *gradient boosting*. A short description of these two classifiers is provided below.

Logistic Regression: Logistic regression is a probabilistic discriminative classifier. For multiclass classification, logistic regression produces piece-wise linear decision boundary. Learning parameters for logistic regression requires numerical optimization. The equation of finding class label of a data object having value x is:

$$f_{LR}(x) = argmax_{c \in Y} P(Y = c|x)$$

Gradient Boosting: Gradient boosting is an ensemble of multiple weak classifiers. When using decision trees, gradient boosting simply incorporates all the predictions by individual decision trees and uses them up to generate its own prediction. However, as gradient boosting adds decision trees one by one, the weaknesses that it gathers from previous iteration are likely to be overcome when a new decision tree is added. As a result in each iteration, the classifier becomes better.

Some other classifiers that we tried are – ensemble methods (Ridge, AdaBoost, Bagging, Extra Trees, Random Forest) and other classifiers – Gaussian Naïve Bayes, K-Nearest Neighbors, Nearest Centroid, Multi-layer Perceptron, Decision Tree, Extra Tree, Support Vector Classifier, Gaussian Process, Passive Aggressive and Linear SVM (trained with Stochastic Gradient Descent).

4.2 Transfer Learning Approach

In this section we first describe the features we work with, the transfer learning approach we base our approach on, the approach which we adopted and finally our general pipeline.

Feature Extraction: Given the raw time series data of accelerometer and gyroscope readings, we extract the same set of features as described in the conventional approach above, with the exception of SMA and variance on 2 second windows.

Transfer Learning Method: We base our transfer learning approach on a method described by Chattopadhyay et al. [3] for knowledge transfer in identifying muscle fatigue from surface electromyography (SEMG) signals across multiple subjects. The method they describe is an Isomap based domain adaptation method which tries to maintain the topology of a manifold structure in the SEMG data across classes which is consistent across different users. Given data D^S from the source data, a small labelled subset of the target data D_l^T and unlabelled target data D_u^T, the method – Topology Preserving Domain Adaptation (TPDA), can be specified as:

- Step 1: Compute the low dimensional projection of D_l^T using Isomap.
- Step 2: Use the mapping to project D^S and D_u^T into the same mapped space.
- Step 3: In the mapped domain, compute the Euclidean distance from each sample in D_l^T to each of the data points in D_S belonging to the same class.

- Step 4: In the mapped domain, sort the distances in increasing order of the value and select k nearest points from D_S for each of samples in D_l^T to form $D_{selected}^S$.
- Step 5: Learn a classifier on $D_{selected}^S$ and D_l^T and compute the labels on D_u^T.
- Step 6: Compute new low dimensional projection again, using $D_{selected}^S$ and labeled test data D_l^T in the original space, and obtain the new mapping.
- Step 7: Go back to Step 2, for N number of iterations.
- Step 8: Assign majority class in N iterations as class for each sample in D_u^T.

A detail to note here is the use of a small amount of *labelled* target data which the above method makes use of; we expound on this in section 5. This aspect of the method was something we were uncomfortable with; it was also unclear why the procedure was repeated N times (Step 7). In our approach we avoid these steps and continue to use a small subset of the target data D_{usub}^T, but we do not require these to be labelled. The proposed approach, *modTPDA*, is described below:

- Step 1: Compute the low dimensional projection of D_l^T using Isomap.
- Step 2: Use the mapping to project D^S and D_u^T into the same mapped space.
- Step 3: In the mapped domain, compute the Euclidean distance from each sample in D_l^T to each of the data points in D_S.
- Step 4: In the mapped domain, sort the distances in increasing order of the value and select k nearest points from D_S for each of samples in D_l^T to form $D_{selected}^S$.
- Step 5: Learn a classifier on $D_{selected}^S$ and D_l^T and compute the labels on D_u^T.

The proposed method tries to build a training set from the source set such that the samples in the new training set are similar to the samples in the target domain. The classifier we chose to train in step 5 of TPDA and modTPDA is random forest classifier. We chose it because of its robustness to possibly irrelevant features and its general good performance on activity recognition tasks. We choose to follow recommendations by Sztyler et al [11] and picked the number of decision trees in the random forest to be 10 and set the dimensionality of the Isomap embedding to 15, roughly half the number of features.

5 Experimental Results

We describe our experimental results in two parts. The first part consists of our findings in conventional supervised learning approach described in section 4.1 and the later part contains findings from transfer learning approach.

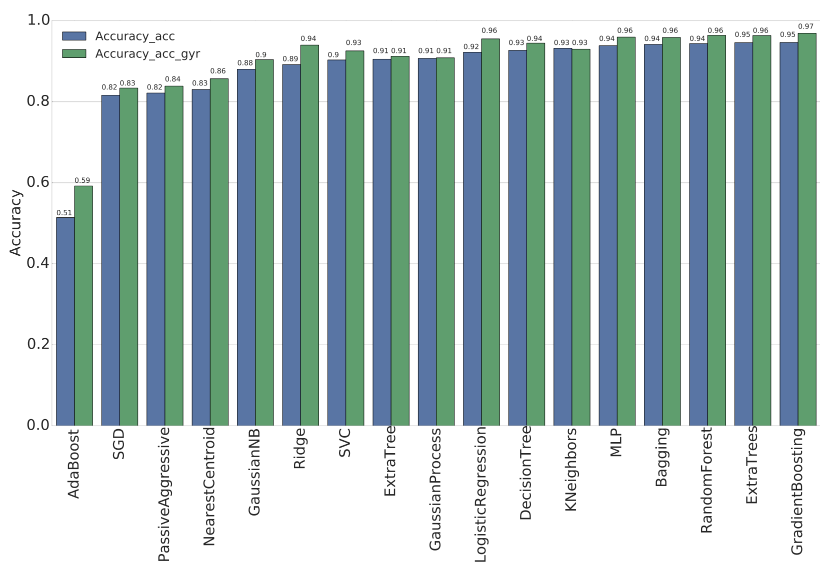

Figure 1: Average accuracy for different classifiers and different sensors per person

5.1 Conventional Approach

We applied a number of different classifiers with default hyperparameters for two different feature settings. The evaluation goal was to observe how the classifier performs when the training and test data both correspond to the same real world individual. This way we calculated accuracy per individual and computed the average of this accuracy over all individuals. Figure 1 shows our findings. It is clear that incorporating gyroscope sensor data significantly improves accuracy for all classifiers. We particularly found linear classifiers (Ridge and logistic regression) to do well compared to other simple classifiers. However, ensemble methods outperform these classifiers and gradient boosting performs best overall.

Table 3: Average accuracy for different window sizes using logistic regression in per person test

WinSz (s)	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
Acc(%)	95	95	96	96	96	96	97	96	96	97	97	96	97	97	96	96	96	96	96

We use a window size of 2 seconds in all our experiments. However, to understand how the window size affects accuracy, we performed an experiment with varying window size while keeping the classifier fixed to logistic regression. The result is shown in table 3 while varying window size from 2 seconds to 20 seconds. This does not affect accuracy to a great extent.

We performed another per-person evaluation on selected classifiers using three additional features (roll, pitch and tilt angle) and the for all classifier, prediction accuracy improved. The result of this experiment is shown in Table 4.

Table 4: Comparison between accuracy for per-person train test method using different feature sets

Classifier	Accuracy (36 features)	Accuracy (39 features)
KNeighborsClassifier	0.9297	0.9332
DecisionTreeClassifier	0.9447	0.9480
SVC	0.9255	0.9378
LogisticRegression	0.9551	0.9631
RidgeClassifier	0.9398	0.9608
MLPClassifier	0.9595	0.9642
BaggingClassifier	0.9586	0.9623
RandomForestClassifier	0.9638	0.9692
ExtraTreesClassifier	0.9634	0.9702
GradientBoostingClassifier	0.9690	0.9720

Next we move on to cross-person evaluation. In this evaluation, While building training set for person n, we have left out all data regarding that person. So the training data contains data of all persons except n-th person. The test data contains only data of n-th person. Table 5 shows accuracy obtained while classifying activities of each individual in leave one out methodology using logistic regression and ridge classifier.

Table 5: Accuracy for each person in leave one out experiment using logistic regression

Test Person	1	2	3	5	8	9	10	11	12	13	15	Avg
Acc (%)-LR	58.74	39.59	34.23	90.73	27.30	61.85	93.00	61.45	54.96	61.92	0.0	53.02
Acc (%)-Ridge	81.95	42.87	38.80	66.53	14.89	45.26	96.38	61.45	54.00	60.30	0.0	51.13

Finally, we have experimented leave one out cross validation over all the 11 persons data for different classifiers using 36 features and the result is reflected in Figure 2. Logistic regression performs best in this scenario even beating ensemble methods. All our implimentations were in `python` and used `scikit-learn`.

5.2 Transfer Learning Approach

5.2.1 Experiments

In our experiments we test the effectiveness of TPDA as a domain adaptation technique for learning a model on a set of users and transferring this knowledge to a target user, who is potentially very

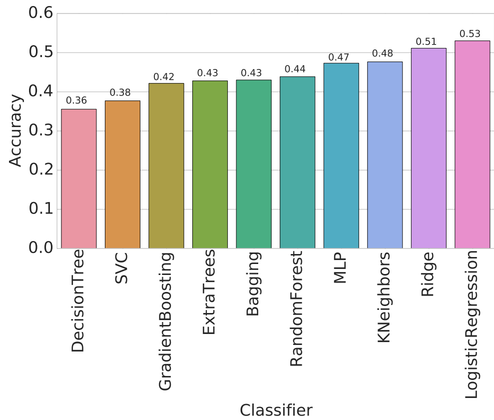

Figure 2: Average accuracy for different classifiers in leave one out cross validation

different from the users the model was trained on. Following this we attempt to demonstrate the effectiveness of modTPDA on the same problem. With these goals in mind we describe two sets of experiments, one set which compares the domain adaptation method TPDA to a baseline method (Baseline-1). The second set of experiments which we consider compares our modification of TPDA, modTPDA to a slightly different baseline (Baseline-2). Note here that we do not compare TPDA directly to modTPDA since this might be an unfair comparison since TPDA requires the use of a small subset of *labelled* target data while modTPDA only makes use of unlabelled target data. In all our experiments the subset of target data used at train time (labelled or un-labelled) is 10% of the target data. Next we describe our baseline methods and the way we set up our source and target group of individuals.

Baselines: Since we seek to demonstrate test the effectiveness of transfer learning approaches the baselines we compare against are conventional classifiers trained on the same data as the transfer learning approaches. Baseline-1 therefore is a simple random forest classifier with the same hyper-parameters as that used in step 5 of TPDA trained on the source data and the same subset of the labelled target data which TPDA makes use of. Baseline-2 is a random forest classifier with the same hyperparameters as that used in step 5 of modTPDA trained on the source data.

Table 6: Groups of people similar and different in movement patterns as source and target domains

	Source	Target
Different	5, 8, 11	2
	8, 11, 15	2
	9, 10, 13	8
	2, 9, 10	15
Similar	1, 10, 13	3
	8, 11, 15	5
	2, 3, 10	9
	3, 10, 13	12

Source and Target Groups: Sztyler et al [11], the publishers of this dataset, group the subjects in this experiment into groups of individuals who share similar movement patterns (Table 3). We make use of this grouping to form 8 sets of source and target groups of individuals. Of these 8 groups, 4 groups are ones where the source and target individuals are similar in their movement patterns while the remaining 4 groups consist of source and target individuals who are different in terms of their movement patterns. This grouping of individuals is summarized in Table 6. In all of our 8 groups the source set consists of data from 3 individuals and the target consists of data from 1 individual. We expect the domain adaptation techniques to perform better than a baseline method on groups where the source and target users are very different.

In all of our experiments we use 5 different sets of random seeds for each run of all our methods and present results averaged across each of these runs. All of the algorithms were implemented in `python` using routines from `pandas` and `scikit-learn`.

Table 7: Accuracies comparing baselines and domain adaptation methods averaged across 4 groups of each group type

	Baseline-1		TPDA		Baseline-2		modTPDA	
	Similar	Different	Similar	Different	Similar	Different	Similar	Different
Accuracy	0.926292	0.937689	0.343208	0.362291	0.568650	0.384043	0.483699	0.429902

5.2.2 Results

A summary of our experimental results are presented in Table 7.

TPDA vs Baseline-1: TPDA consistently seems to do significantly worse than the baseline method Baseline-1 for both types of groups of individuals in our experiments indicating that TPDA isn't ideally suited to our task. The results seem to indicate that if one had access to labelled target data it might be better to use the data to train a conventional classifier. But we do concede that this might be a result very specific to the kinds of experiments we perform. We do however highlight a particular detail in the work of Chattopadhyay et al. [3], where a baseline of this nature isn't considered in their work. They merely compare results to a conventional classifier which isn't given access to the labelled target data.

modTPDA vs Baseline-2: Our findings indicate that for the case of the source and target individuals being different from one another the proposed domain adaptation method modTPDA performs about 4.5% better than the baseline method, Baseline-2. The proposed method however does worse than Baseline-2 for the case of the source and target individuals being similar to one another. This might be expected since it might be very likely that in the formation of $D^S_{selected}$ some of the good samples from the source subset get discarded. We also highlight that modTPDA performs better than TPDA despite TPDA having the added advantage of having made use of labelled target data on both types of groups of individuals.

6 Discussion and Conclusions

In this project, we have approached the human activity recognition problem using conventional supervised learning and transfer learning. We have performed experimentation on a dataset consisting of 11 persons and found that logistic regression achieves best prediction accuracy among all the classifiers we tested in a leave one person out cross validation evaluation. In a per person evaluation strategy we found gradient boosting to have the best prediction accuracy.

In this work we also applied an Isomap based domain adaptation method, TPDA, proposed by Chattopadhyay et al. [3] and found it to perform worse than a baseline. Furthermore we applied a modified version of TPDA on our dataset and found it to perform better than TPDA for all 8 groups of source and target groups we tested over. The modified approach also performed better than our baseline for the case of the source and target individuals being different from one another. Our transfer learning experiments seem to indicate that domain adaptation helps improve performance when source and target datasets are actually different, something which one would expect. We also found our findings to be aligned with those of Sztyler et al. [11] in an active learning approach applied across individuals on the same dataset as ours.

A point about model selection we highlight in the context of transfer learning is as follows; Typically model selection could be performed by a cross-validation or by tuning parameters on a development set. In this case however, it is unclear how either of these are rational options; tuning parameters on a development set from the target data would be equivalent to learning from the test data, while a cross-validation for parameter selection run on the source (train) data would not be meaningful since the inherent problem we face is that of the source data being significantly different from the target (test) data. In this context therefore For the purposes of our experiments, where we demonstrate the effectiveness of the proposed approach, the models are sufficient since the same parameters are used in all the models being compared. However, a possible approach to tune the parameters of the proposed model could be to alternatively tuning the parameters of the Isomap step and that of the classifier on $D^S_{selected}$; however, this requires further experimentation and we did not perform this tuning in our experiments. This is an aspect of the work which we would like to explore further.

References

[1] M. Awais, L. Palmerini, and L. Chiari. Physical activity classification using body-worn inertial sensors in a multi-sensor setup. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–4, Sept 2016.

[2] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004.

[3] Rita Chattopadhyay, Narayanan Chatapuram Krishnan, and Sethuraman Panchanathan. Topology preserving domain adaptation for addressing subject based variability in SEMG signal. In *Computational Physiology, Papers from the 2011 AAAI Spring Symposium, Technical Report SS-11-04, Stanford, California, USA, March 21-23, 2011*, 2011.

[4] Diane Cook, Kyle D. Feuz, and Narayanan C. Krishnan. Transfer learning for activity recognition: a survey. *Knowledge and Information Systems*, 36(3):537–556, 2013.

[5] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pages 4 pp.–116, 2006.

[6] Susanna Pirttikangas, Kaori Fujinami, and Tatsuo Nakajima. *Feature Selection and Activity Recognition from Wearable Sensors*, pages 516–527. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[7] Jorge Luis Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. Transition-aware human activity recognition using smartphones. *Neurocomputing*, 171:754–767, 2016.

[8] Muhammad Shoaib, Stephan Bosch, Özlem Durmaz Incel, Hans Scholten, and Paul J. M. Havinga. Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors*, 16(4):426, 2016.

[9] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, SenSys '15, pages 127–140, 2015.

[10] T. Sztyler and H. Stuckenschmidt. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9, March 2016.

[11] Timo Sztyler and Heiner Stuckenschmidt. Online personalization of cross-subjects based activity recognition models on wearable devices. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 180–189. IEEE Computer Society, 2017.

[12] Timo Sztyler, Josep Carmona, Johanna Völker, and Heiner Stuckenschmidt. Self-tracking reloaded: Applying process mining to personalized health care from labeled sensor data. In *Transactions on Petri Nets and Other Models of Concurrency XI*, volume 9930, pages 160–180. Springer-Verlag Berlin Heidelberg, 2016.

[13] Douglas L. Vail, Manuela M. Veloso, and John D. Lafferty. Conditional random fields for activity recognition. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '07, pages 235:1–235:8, New York, NY, USA, 2007.

[14] Mi Zhang and Alexander A. Sawchuk. A feature selection-based framework for human activity recognition using wearable multimodal sensors. In *Proceedings of the 6th International Conference on Body Area Networks*, BodyNets '11, pages 92–98, 2011.