Smart Health 19 (2021) 100174

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Activity recognition in wearables using adversarial multi-source domain adaptation

Avijoy Chakma [a,*], Abu Zaher Md Faridee [a], Md Abdullah Al Hafiz Khan [b], Nirmalya Roy [a]

[a] Department of Informations Systems, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250, USA
[b] Philips Research Americas, 222 Jacobs St., Cambridge, MA, 02141, USA

A B S T R A C T

Human activity recognition (HAR) from wearable sensors data has become ubiquitous due to the widespread proliferation of IoT and wearable devices. However, recognizing human activity in heterogeneous environments, for example, with sensors of different models and make, across different persons and their on-body sensor placements introduces wide range discrepancies in the data distributions, and therefore, leads to an increased error margin. Transductive transfer learning techniques such as domain adaptation have been quite successful in mitigating the domain discrepancies between the source and target domain distributions without the costly target domain data annotations. However, little exploration has been done when multiple distinct source domains are present, and the optimum mapping to the target domain from each source is not apparent. In this paper, we propose a deep Multi-Source Adversarial Domain Adaptation (*MSADA*) framework that opportunistically helps select the most relevant feature representations from multiple source domains and establish such mappings to the target domain by learning the *perplexity scores*. We showcase that the learned mappings can actually reflect our prior knowledge on the semantic relationships between the domains, indicating that *MSADA* can be employed as a powerful tool for exploratory activity data analysis. We empirically demonstrate that our proposed multi-source domain adaptation approach achieves 2% improvement with OPPORTUNITY dataset (cross-person heterogeneity, 4 ADLs), whereas 13% improvement on DSADS dataset (cross-position heterogeneity, 10 ADLs and sports activities).

1. Introduction

The widespread availability and popularity of wearables and IoT devices have helped to collect and analyze a large volume of human activity data streams seamlessly from inbuilt accelerometer, gyroscope, and magnetometer sensor on devices. Moreover, the rapid advances in machine learning, especially deep learning architectures, have equipped us with powerful data analytics tools to build innovative applications that help to exploit the latent information prevalent in these collective heterogeneous data streams. For example, the novel application of human activity recognition with IoT and wearable devices have penetrated smart home environments (Wen et al., 2016), sports analytics (Hossain et al., 2017), dance choreographies (Faridee et al., 2018), and smart health such as cognitive and mobility health assessment of older adults. However, several pervasive challenges have hindered the large scale adoption of such statistical and machine learning-based HAR models. One of the most prominent impediments occurs in the form of a discrepancy between the distributions of data during the training and the inference phase of these algorithms. Machine learning models are often trained on carefully curated data collected in a controlled environment; as such, the models often lack the capacity to adapt when encountering relevant data streams containing heterogeneous distributions. *Device heterogeneity* (Stisen et al., 2015) is one of the most common precursor of such *distribution shift*.

* Corresponding author.
E-mail addresses: achakma1@umbc.edu (A. Chakma), faridee1@umbc.edu (A.Z.M. Faridee).

https://doi.org/10.1016/j.smhl.2020.100174

Available online 14 December 2020

In addition to the heterogeneity arising from each device's inherent characteristics, the heterogeneity can also resurface during the data collection process. Not only the maximum sampling frequency can vary between devices, but they can often be set purposely different due to the inherent power and resource constraints on the devices. Even with devices of the same model, make, and configuration, the divergence between the data distributions can still occur as devices are placed on different parts of the body. From smart jewelry placements in ears and neck to smart shoes on the feet, the ubiquitous nature of the off-the-shelf wearable consumer devices (e.g., Apple watch,[1] Fitbit,[2] Oura Ring[3] etc.) transpire an ever-increasing number of options about the on-body placements. As these wearable sensors are expected to capture the deeper local context of specific human body positions/limbs, the data distributions across such positions start diverging, even if the performed activity remain constant and wearable of interest does not change. Such a disparity is shown in Fig. 1 on PAMAP2 dataset (Reiss & Stricker, 2012, pp. 108–109), where three Colibri wireless IMU sensors are placed on the neck, wrist and ankle, respectively for the same *standing* activity but exhibits three different data distributions. Therefore, machine learning models trained on one specific wearable device face challenges in classifying activities from its counter ones unless special pre-processing and/or post-processing measures are postulated and integrated in the HAR pipeline well to minimize those discrepancies.

Most of the traditional machine learning algorithms are based on the fundamental assumption that both the training and test data arise from the same distribution, which leads to performance degradation as the distributions diverge. In the case of supervised learning (especially deep learning), achieving good performance hinges on training on a large number of labeled samples that cover the diverse distributions, which are often costly or downright implausible. Unsupervised domain adaptation techniques circumvent the requirement for labeled data by reducing the feature discrepancy between the source and the target domain (the literature generally refers to the labeled dataset as the source domain and the unlabeled dataset as the target domain). Recently, several unsupervised domain adaptation techniques (Akbari & Jafari, 2019; Khan et al., 2018; Wang et al., 2018a) have shown success in classifying human activities without any labeled samples in the target domain. However, little exploration has been done in the case where multiple distinct source domains are present, and the optimum mapping to the target domain from each source domain is unknown. For example, we might have labeled data available for several users where the sensors are placed on both at the neck and ankle, but we are tasked with classifying the activity of a user whose data was collected with a single sensor placed at the wrist. Given such predicament, it is not obvious whether we choose the neck or the ankle as the primary source domain as the domain similarity can vary drastically depending on the body position, let alone the activity label or device type. As shown in Fig. 1, if we consider the wrist as the target domain and the neck and ankle as the two source domains for standing activity, the data from the neck might constitute more similarity with the wrist compared to ankle, whereas for a different activity such as biking, the ankle might provide more relevant features. Existing single source domain adaptation techniques have inherent limitations in addressing such issues. As such, there is a need for an activity recognition framework that can find the most relevant feature representations between multiple source domains based on the body position, activity performed, or device chosen and thus learn optimum mapping to the target domain. In this work, we propose a deep adversarial learning approach to the multi-source domain adaptation problem and focus on cross-user and cross-position heterogeneity in detail. Our deep architecture is designed to extract low-level domain invariant features in the presence of multiple source domains and unlabeled target domain. The feature extractor acts as a generator against per source-specific discriminators. We postulate a novel perplexity score to dynamically select which of the source domains are close to a given target sample, which helps to define the final classification labels as a weighted mean of the source classifiers. To our knowledge, this is the first work that systemically explores adversarial multi-source domain adaption in activity recognition literature.

The key contributions of our paper are summarized as follows:

- **Deep adversarial multi-source domain adaptation for activity recognition:** The sensors placed on different body positions contribute significantly differently in the activity training, recognition, and inference pipeline. The traditional machine learning models are built for recognizing one activity using a specific on-body position sensor data streams when tested using the same sensor data streams for the same activity, but from another new body position, the classifier performance degrades drastically. Motivated by this, we propose a deep learning-based framework - Multi-Source Adversarial Domain Adaptation (*MSADA*) to map the most relevant feature representations from multiple source domains to one target domain. Our proposed *MSADA* approach helps mitigate the confusion in label annotations arisen from heterogeneous body positions. We also investigate the different deep learning settings to assess the impact of our proposed deep network design.
- **Empirical evaluation with multiple publicly available datasets:** We considered three publicly available datasets - OPPORTUNITY, DSADS, and PAMAP2 to demonstrate the superior performance of the proposed *MSADA* over other alternative approaches. Considering multiple body positions as the sources and target domain, we experimented on two heterogeneous data distribution settings from these three datasets. Experimentation depicts that *MSADA* generates better classification accuracy over the transfer learning-based methods and deep learning-based approach.
- **Reinforcing the most relevant source to the target domain:** Our proposed *MSADA* framework helps to find out the most relevant source domain using an adversarial learning framework from the multiple source domains. We identify the most pertinent source domain in terms of the classification accuracy in the target domain. To demonstrate, we considered a cross-person data distribution heterogeneity for a particular body position. Analyzing the resulting confusion matrices, we show that with the help of

[1] https://www.apple.com/watch/.
[2] https://www.fitbit.com/charge3.
[3] https://ouraring.com/.

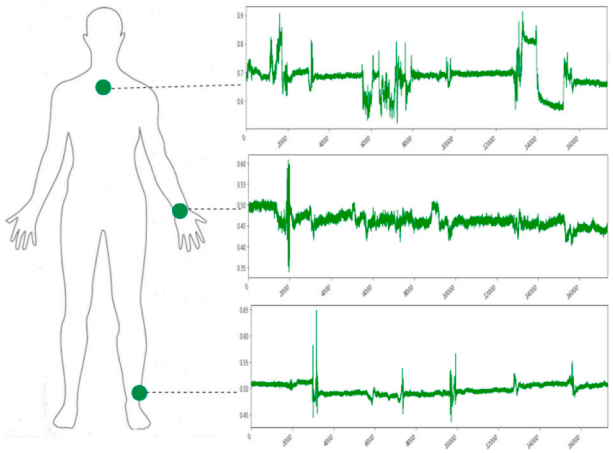

Fig. 1. Three accelerometers' readings from three different body positions of same user for standing activity. Three readings follow three different data distributions.

the novel, perplexity scores *MSADA* opportunistically select the most relevant domain in classifying the unlabeled target domain samples.

The rest of the paper is organized as follows. In section 2, we review the current literature on the domain adaptation in activity recognition. Section 3 introduces our proposed approach, the *MSADA* framework. Section 3.3 details the various algorithmic component of the framework. In section 4, we articulate the design choices of our experiment, describe the datasets and discuss the baseline machine learning approaches. In section 5, we discuss the potential application of the multi-source domain adaptation. Finally, we conclude and discuss the future work in section 6.

2. Related works

Domain adaptation (DA) is the particular branch of the transfer learning that studies the effective measures against data distribution heterogeneity (Cook et al., 2013; Wang et al., 2018b). Depending on the technique, domain adaptation approaches are primarily categorized among instance-based (Rokni & Ghasemzadeh, 2017), classifier-based (Karbalayghareh et al., 2018), feature representation-based (Haeusser et al., 2017) methods. The feature representation-based approach aims to keep the features from all the domains as close as possible while maintaining the class distribution probabilities. More generally, the goal of these approaches is to minimize the marginal distribution among heterogeneous data distributions. Such approaches can further be divided into discrepancy-based (Sun & Saenko, 2016), and generative-adversarial (Sankaranarayanan et al., 2018) approaches depending on the learning mechanism. Two major components of the generative-adversarial network are the generator and the discriminator. Generator aims to generate samples closer to some ground-truth data distribution to fool the discriminator as the discriminator receives the real and generated samples and aims to predict the sample origin. Both components striving towards their own goal, modulate the learning process of each other. On the other hand, based on the availability of sample annotations, the learning approaches can be categorized between supervised, unsupervised, and semi-supervised adaptation. Given such a hierarchy, our proposed framework falls under the unsupervised adversarial domain adaptation umbrella.

2.1. Domain adaptation in activity recognition

There have been several successful attempts to apply domain adaptation methodologies to tackle distribution heterogeneity in activity recognition literature. Among them, discrepancy-based DA approaches use different learning metrics to alleviate the distribution divergence. For example, Wang et al. (Wang et al., 2018a) proposed a stratified transfer learning (STL) approach that employed the pseudo-leveling concept on the unlabeled target data. The approach tackles the distribution heterogeneity by minimizing maximum mean discrepancy (MMD) between the feature spaces for the source domain data and the pseudo-labeled target data. Khan et al. (Khan et al., 2018) investigated a deep learning-based approach, HDCNN, that leverages a pre-trained network consist of feature extractor and classifier where the pre-trained network is trained with the labeled source domain samples. The strategy that HDCNN applied was to transfer the feature extractor weights to a target domain-dedicated feature extractor through the minimization of the Kullback-Leibler (KL) divergence of the activations between the pre-trained network and the target domain dedicated-feature extractor. Faridee et al. (Faridee et al., 2019) explored a semi-supervised deep learning-based framework, AugToAct, that operates on a limited number of source and target domain samples. To achieve domain adaptation on a limited number of source domain samples, AugToAct tailored the data augmentation technique and semantic preservation computing component to acquire a pre-trained network. Following a pre-trained network, AugToAct similarly performs domain adaptation as HDCNN, except it, employs JSD divergence to minimize the activation discrepancy between source and target domain networks instead of (KL) divergence. The

approach proposed by Akbari et al. (Akbari & Jafari, 2019) follows the same network architectural design as HDCNN. The significant difference in the proposed approach is that the features are extracted in a probabilistic manner by employing variational auto-encoder instead of deterministic feature extraction. In all these approaches, the model considers only a single source domain for domain adaptation task.

2.2. Multi-source domain adaptation

In the sensor-based activity recognition, there is a very handful amount of works that consider multiple sources for the domain adaptation purpose. Some approaches focus on explicitly identifying the most relevant source domain among the multiple source domains with the target domain based on similarity measurement such as *cosine similarity*. Another group of approaches combines all the available source domains data and projects into lower-dimensional space, which is further processed by the classifier. Wang et al. (Wang et al., 2018b) have considered explicitly selecting the most relevant domain from the multiple source domains based on the cosine similarity and used the selected domain for domain adaptation. Recently Jeyakumar et al. (Jeyakumar et al., 2019) proposed *SensHAR*, a multiple-sensor data-fusion-based approach to mitigate the heterogeneous data distribution and assign label the unlabeled data. The significant difference between our proposed approach with the mentioned approaches is that we consider multiple source domain data simultaneously, and our network process individual source domain data independently. Such consideration allows the approach to capture the uncertainty within the classification tasks and dynamically select the most relevant source without fixing it explicitly. The proposed methodology also aligns with the theoretical motivation of Multi-Source Domain Adaptation that assumes the availability of multiple labeled sources adaptation process. Mansour et al. (Mansour et al., 2009) postulate that the hypothesis for the target domain data can be constructed from the combination of the source domain hypotheses. In this paper, we propose an adversarial-based technique to achieve domain adaptation considering multiple source domains simultaneously and identify the most relevant source domain with the target domain for the classification task.

In the next section, we discuss the multi-source domain adaptation problem formulation and describe our proposed framework and learning mechanism in detail.

3. Multi source adversarial domain adaptation (*MSADA*)

In this paper, we propose an adversarial deep learning framework, *MSADA* for domain adaptation to assign labels for the unlabeled target domain data in the presence of multiple labeled source domain data. The proposed framework consists of four components - feature extractor (F), domain discriminator (D), label classifier (C), and label decider. Feature extractor extracts the domain invariant features among multiple modalities. The domain discriminator, similar to other adversarial approaches, attempts to maximize its origin-identifying capability of the incoming data (discussed in detail section 3.2.2). The classifier serves to classify the incoming features into different classes. The final component, label decider, process incoming the classification results from the classifiers and novel perplexity scores from the domain discriminators to decide labels to the unlabeled target domain data. In the following, we describe the multi-source domain adaptation problem formulation, the proposed framework details, and the domain adaptation mechanism.

3.1. Problem formulation

In our multi-source domain adaptation problem formulation, we assume that there are N source domains with labels and one target domain without any labeling information. The source domains samples are drawn from N-different data distributions $P_{s_j}(x, y)_{j=1}^{N}$. These

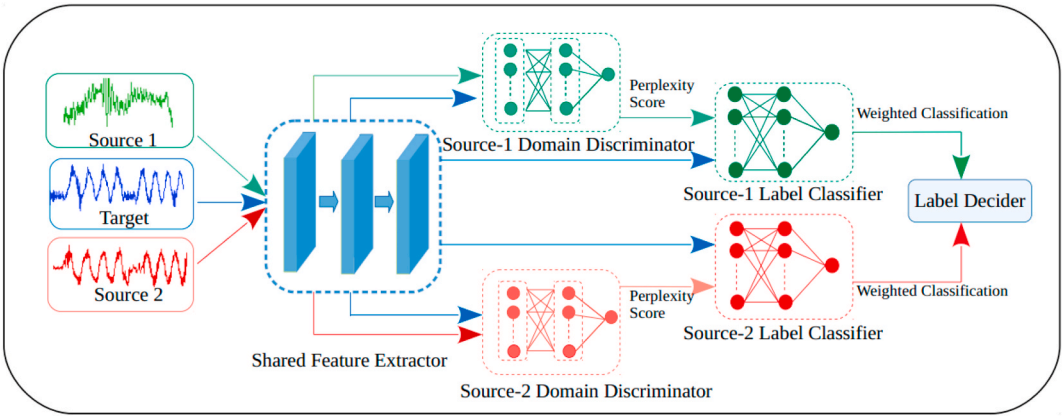

Fig. 2. Overview of our proposed *MSADA* framework that employs pre-trained feature extractor and classifiers. The framework receives labeled source and unlabeled target domain samples. Through an adversarial learning mechanism, it aligns the marginal probability of the distributions. *MSADA* leverages the aligned data distributions, and the novel, perplexity scores classify the unlabeled target domain samples.

distributions generate the labeled source domain samples, $(X_{s_j}, Y_{s_j})_{j=1}^{N}$ where $X_{s_j} = \{x_{s_{ji}}\}_{i=1}^{|X_{s_j}|}$ belongs to the source j and $Y_{s_j} = \{y_{s_{ji}}\}_{i=1}^{|X_{s_j}|}$ is the ground truth for the corresponding sample from the source j. We consider $P_t(x,y)$ as the probability distribution that generates the target domain samples, $X_t = \{x_{t_i}\}_{i=1}^{|X_t|}$ without the label space Y_t. We assume that both the source domains and the target domain have homogeneous label space.

3.2. Proposed architecture

The overall architecture of *MSADA* is shown in Fig. 2. In the following sections, we discuss the four major components in detail.

1) *Feature Extractor (F):* The feature extractor module consists of three two-dimensional convolution layers, followed by a max-pooling and a *SELU* (Klambauer et al., 2017) activation function. We employ three convolutional layers for the sake of capturing the generalizable features across multiple datasets with multiple activities. The max-pooling layers help to down-sample the input features into lower dimensions to provide scale-invariant feature representations. In the proposed framework, the same feature extractor is shared between the *Domain Discriminators* (section 3.2.2) and *Label Classifiers* (section 3.2.3) as shown in Fig. 2 with an aim to capture the domain invariant features. Through learning the domain invariant feature across multiple domains, we aim to minimize the marginal data distribution heterogeneity.
2) *Domain Discriminator (D):* Domain discriminator module consists of the *N*-source domain-specific discriminators, $\{D_{s_j}\}_{j=1}^{N}$ each tries to predict the origin of the incoming activation from the respective source or target feature distributions. In the framework, each dedicated source-specific domain discriminator D_{s_j} serves as a regular adversarial unit of an adversarial learning mechanism.

In the adversarial learning mechanism, the generator aims to generate domain-invariant features of the source and target domain samples, so that the discriminator can not reach its goal of predicting the origin of the incoming features. The overall mechanism resembles a min-max game, and over the process, both generator and discriminator becomes better at achieving the corresponding goal. In addition to extracting the domain invariant features in the proposed approach, the feature extractor serves the role of a feature generator for the domain discriminators during the adversarial learning phase.

In *MSADA*, each of the source-specific discriminators consists of two layers of a fully connected layer with *SELU* in the first layer and a sigmoid activation in the final layer. Each dedicated source-specific domain discriminator process the incoming features from the target domain, $F(x_t)$ and a specific source domain, $F(x_{s_j})$ samples. Besides, providing feedback through an adversarial loss to the feature extractor of its domain-invariant feature generation, each of the domain discriminators serves an additional role of providing a perplexity score for the target domain data to the classifiers. The *perplexity score* acts as a measure of the closeness of the target domain with a specific source domain. Perplexity score serves as a classification *weight* of the corresponding dedicated source domain classifier. The perplexity score is calculated using the following formula:

$$P_{s_j}\left(x_t; F, D_{s_j}\right) = -log\left(1 - D_{s_j}(F(x_t))\right) \tag{1}$$

3) *Label Classifier (C):* Label classifier consists of *N* source domain-specific multi-class classifiers $\{C_{s_j}\}_{j=1}^{N}$. Classifier design is similar to the domain discriminator except for the softmax activation in the final layer, which is configured according to the label space of the corresponding source domain *j*. Each classifier C_{s_j}, along with the shared feature extractor, are pre-trained with labeled data from corresponding *j* source domains.
4) *Label Decider:* The label decider determines the label for the unlabeled target data. It receives the perplexity scores provided by the source-domain specific discriminators, the classification results, $C_{s_j}(F(x_t))$ provided by the source-domain specific *classifiers*. Using the perplexity score label decider re-weights the classification scores and performs a weighted classification to assign a label to the target data. The confidence of class *c* for a target data x_t is calculated as -

$$Confidence\left(c|x_t\right) = \sum_{j=1}^{N} \frac{P_{s_j}\left(x_t; F, D_{s_j}\right)}{\sum_{j=1}^{N} P_{s_j}\left(x_t; F, D_{s_j}\right)} C_{s_j}(c|F(x_t)) \tag{2}$$

x_t is categorized into the class with the highest confidence.

3.3. Model learning

MSADA learns the model leveraging the labeled and unlabeled datasets from both source and target domains, respectively. Feature extractor and the source-domain specific classifiers are pre-trained using the labeled source-domain data. We adapt the pre-trained feature extractor by employing the adversarial objective from the domain discriminators to the feature extractor to align the domain invariant features.

The adversarial objective can be constructed as -

$$\min_{F}\max_{D} V\left(F, D; \overline{C}\right) = L_{adv}(F, D) + L_{cls}\left(F, \overline{C}\right) \tag{3}$$

$$L_{adv}(F,D) = \frac{1}{N}\sum_{j}^{N} \mathbb{E}_{x\sim X_{s_j}}\left[\log D_{s_j}(F(x))\right] + \mathbb{E}_{x_t\sim X_t}\left[\log\left(1 - D_{d_j}(F(x_t))\right)\right] \tag{4}$$

where the first term of the equation (3) defines the adversarial mechanism (Tzeng et al., 2017) and the second term denotes the multi-class classification loss. We keep the classifiers unchanged to provide stable gradients for the feature alignment. Based on equation (4), the optimization works well for D but not F. As the feature extractor continuously processes the multiple data distributions, these heterogeneous data distributions cause an adversary in the feature extractor learning process. To mitigate this issue, as suggested by Tzeng et al. (Tzeng et al., 2017), we replace the adversarial objective with the domain confusion loss, which enables stable learning for the feature extractor. We replace the adversarial loss in equation (3) with the confusion loss as follows -

$$L_{adv}(F,D) = \frac{1}{N}\sum_{j}^{N} \mathbb{E}_{x\sim X_{s_j}} L_{cf}\left(x; F, D_{s_j}\right) + \mathbb{E}_{x_t\sim X_t} L_{cf}\left(x_t; F, D_{s_j}\right) \tag{5}$$

Here the first and second term determines the domain confusion loss for the source samples and target samples, respectively. Domain confusion loss for both source and target domain samples is calculated using the equation (6).

$$L_{cf}\left(sample; F, D_{s_j}\right) = \frac{1}{2}\log D_{s_j}(F(sample)) + \frac{1}{2}\log\left(1 - Ds_j(F(sample))\right) \tag{6}$$

In a summery, we train the domain discriminator, D using equation (4) whereas the feature extractor is trained using equation (3) with the adversarial loss is replaced with a confusion loss calculated through equation (5) for a stable learning process.

Algorithm 1: Mini-batch Learning via online hard domain batch mining

Input: Mini-batch of M samples from sources and target domains; pre-trained feature extractor F and label classifier $\overline{C}$; domain discriminator D.
Output: Updated F'
1: Select the source-specific discriminator where the corresponding adversarial loss for M samples is maximum using Eq. (4)
2: Compute the domain confusion loss for the selected source-specific discriminator in Step-1 using Eq. (6) (Using only M samples of selected source and target samples)
3: Replace L_{adv} loss in Eq. (3) with the calculated confusion loss in Step-2
4: Update F by Eq. (3)
5 return $F' = F$

Another challenge in multi-source domain adaptation is the ratio of the source domain samples against the target domain samples as it impacts the feature extractor learning process. To alleviate this issue, for each batch of the target samples, we select a particular source domain batch for which the adversarial loss is maximum. The higher adversarial loss serves as an indicator that a particular domain discriminator incurs higher loss when it does not perform well in predicting the origin of incoming activations. Algorithm 1 describes the online source domain batch selection process, and algorithm 2 summarizes the overall *MSADA* approach.

Algorithm 2: Learning Algorithm for *MSADA*

Input: N source labeled datasets $\{X_{s_j}, Y_{s_j}\}_{j=1}^{N}$; unlabeled target dataset X_t; pre-trained feature extractor F and label classifier $\overline{C}$; domain discriminator D; adversarial iteration threshold β
Output: well-trained feature extractor F^*, domain discriminator D^*.
1: Pre-train $\overline{C}$ and F
2: while not converged do
3: for 1: β do
4: Sample mini-batch from $\{X_{s_j}\}_{j=1}^{N}$ and X_t
5: Update D by equation (4);
6: Update F by algorithm 1;
7: end for
8: end while
9: return $F^* = F$; $D^* = D$.

4. Experiments and evaluation

In this section, we discuss the experimental details and evaluation of *MSADA* on three publicly available datasets, compare the baseline methods, and present the effectiveness of the proposed novel perplexity scores.

4.1. Datasets

We validate *MSADA* performance with OPPORTUNITY (Roggen et al., 2010), Daily and Sports Activities (DSADS) (Barshan & Yüksek, 2014) and PAMAP2 (Reiss & Stricker, 2012, pp. 108–109) dataset. OPPORTUNITY and PAMAP2 dataset contain daily living activities, whereas the DSADS contains a combination of sports activities and daily living activities.

Activities in the OPPORTUNITY dataset are - {‘ Standing’, ‘Walking’, ‘Sitting’, ‘Lying’}, collected from 4 participants using Inertial Measurement Units (IMU) sensors configured at a sampling frequency of 32 Hz. Dataset was collected by placing IMU sensors at seven different body positions from each participant. In the experiment, we considered data from five body positions - BACK, Right Upper Arm (RUA), Right Left Arm (RLA), Left Upper Arm (LUA), Left Lower Arm (LLA) from all four participants.

DSADS dataset contains a combination of daily and sports activities, collected from 8 participants. Sensors were placed on five different body positions - TORSO, Right Arm (RA), Left Arm (LA), Right Leg (RL), and Left Leg (LL). Each sensor consisted of a 3-axis accelerometer (Acc), gyroscope (Gyr) and magnetometer (Mag), and was configured with a sampling frequency of 25 Hz. DSADS induced a substantial inter-user variation as the participant naturally performed the activities. We considered 10 activities from DSADS dataset - {‘standing’, ‘lying-back’, ‘ascending’, ‘walking-parking-lot’, ‘treadmill-running’, ‘stepper-exercise’, ‘cross-trainer-exercise’, ‘rowing’, ‘jumping’, ‘playing-basketball’} in our experiments.

PAMAP2 dataset contains 18 activities from 9 participants. Dataset was collected using 3 Colibri wireless IMU sensors configured at a sampling frequency of 100 Hz from three different body positions - Dominant Arm (DA), Chest, Dominant Side’s Ankle (DA). We considered 11 activities from three body-positions - {‘lying’, ‘sitting’, ‘standing’, ‘walking’, ‘running’, ‘cycling’, ‘ascending’, ‘descending’, ‘vacuum’, ‘ironing’, ‘rope_jumping’}.

Table 1 summarizes the characteristics of these datasets. In the experiment, we use only the accelerometer data.

4.2. Baseline methods

We compare *MSADA’s* performance with dimension reduction-based technique, transfer learning-based approach, and deep learning-based methodology. We include principal component analysis (PCA) as a dimension reduction-based approach. Transfer learning-based baseline approaches cover Balanced Distribution Adaptation (BDA) (Wang et al., 2017), and Correlation Alignment (CORAL) (Sun et al., 2017, pp. 153–171), Stratified Transfer Learning(STL) (Wang et al., 2018a). We include the RevGrad (Ganin & Lempitsky, 2014) as the deep learning-based baseline method.

CORAL focuses on aligning the source and target domain feature distribution without projecting into lower-dimensional subspaces. BDA considers balancing between the marginal and conditional distribution during the domain adaptation process. Whereas, STL incorporates the idea of pseudo-labeling for the target domain samples and reduces the class-specific MMD distance between the source and target domain samples. On the other hand, RevGrad focused on reducing the marginal distribution between the source and target domain with the gradient reversal layer’s help.

4.3. Performance metrics

In the experiments, we have computed micro confusion matrix and observed that the accuracy score and F1 values are similar at the fourth digit fractional value. Without overwhelming the result section, we report the classification accuracy on the target domain samples as the evaluation metric, which is also widely used in existing domain adaptation methods (Wang et al., 2018a).

$$Accuracy = \frac{|x : x \in X_t \wedge \widehat{y}(x) = y(x)|}{x : x \in X_t} \tag{7}$$

where $y(x)$ and $\widehat{y}(x)$ are the truth and predicted labels, respectively.

4.4. Implementation details

We implement the proposed *MSADA* framework and deep learning-based RevGrad using the open-source deep-learning library PyTorch (Paszke et al., 2017). We adopted the authors-provided MATLAB code of Stratified Transfer Learning (STL).[4] The rest of the baselines methods are implemented using Python, and Python-based library Sci-kit (Pedregosa et al., 2011) Learn Tool. We execute the experiments on a Linux Server (Ubuntu 18.04) running on Intel Core i7-6850K CPU and 64 GB DDR4 RAM, with 4 Nvidia 1080Ti Graphics cards with 44 GB VRAM.

In the preprocessing of each dataset, we initially extract the body position-wise accelerometer activity data from each participant, followed by standardization. We remove the `NaN` (Not a Number) entries from the dataset and further split the position-wise extracted dataset into training, validation, and testing set in the standard ratio of 60-20-20% in such a way that each activity contributes in the mentioned ratio. For example, consider the Right-Upper-Arm (RUA) body position from the OPPORTUNITY dataset that contains data for four activities for each participant. Lets hypothetically consider that for a participant, RUA extracted dataset has 10000 entries with equal activity distribution that is each activity will have 2500 entries. In this scenario, we split activity-wise 2500 entries in the ratio of 60-20-20%. Then we combine individual training splits for each activity for each participant into a single and repeat the same process for validation and testing splits. Finally, as the final preprocessing for the proposed *MSADA* framework and RevGrad baseline, we have applied the windowing technique over the splits and empirically found that 90% overlap generates superior accuracy over 80% and 70% overlapping. As part of the data preprocessing, we leveraged data augmentation in sensor-collected data as augmented data is found to help the network achieve better generalizability and improve network performance (Faridee et al., 2019; Um et al., 2017). We

[4] https://github.com/jindongwang/activityrecognition/tree/master/code.

Table 1
Datasets overview.

Factor	Opportunity	PAMAP2	DSADS
Sensors	Acc, Gyr, Mag	Acc, Gyr, Mag	Acc, Gyr, Mag
Positions	BACK, RUA, RLA, LUA, LLA, 2 sensors on Shoes	DA, Chest, DL	TORSO, LA, RA, LL, RL
Sampling Frequency (Hz)	32	100	25
Dataset Size	2551	3850505	9120
Subject	4	9	8

used jitter, magnitude warping, and time warping for the data augmentation and used (Faridee et al., 2019) suggested parameter values for these augmentation methods.

To the best of our knowledge, there are no such approaches in sensor-based activity recognition that considers at least three different domains simultaneously. The input design-consideration of the baseline methods was different from *MSADA* because the baselines considered only a single labeled source domain. We slightly modified the input for the baselines to compare with *MSADA*. We combined the corresponding training splits of source domains, trained the baseline, tested the target domain, and derived the baseline performances. Among the baseline methods, we adjust the convolutional layers filter size and input-output channels of the RevGrad method, which is initially proposed for image data. In the case of STL, the authors calculated 81 features - 27 features for accelerometer, gyroscope, and magnetometer sensor modality. To keep competency with our approach, we considered only accelerometer-focused 27 features. For 81 features, the authors used a reduced dimension of 30; we adopted a reduced dimension of 10 for 27 accelerometer features.

Table 2 specifies the hyper-parameters that we used in our proposed architecture. We considered employing three layers of the convolutional layer to capture the generalized feature. The max-pooling layer follows the initial two convolutional layers, whereas we avoided using after the third convolutional layer to provide sufficient features to the discriminators and classifiers. The classifier network is almost identical to the discriminator network, where the sole difference is the output of the final fully connected layer - discriminator uses a sigmoid layer and provides only a single output. The output of the classifiers varies depending on the number of classes in the considered dataset. As the optimizer, we have experimented with ADAM (Kingma et al., 2014) as advised by the experts.[5] For the other hyper-parameters - learning rate, epochs, batch size, beta values, we have used the standard hyper-parameter values.

4.5. Experimental settings

We considered two scenarios where data distribution heterogeneity can arise due to: (i) cross-person heterogeneity, (ii) body-position heterogeneity. To align with the multi-source domain adaptation assumption, we notion the term *domain* to refer individual persons and body-positional data.

1) *Cross-person Heterogeneity:* In cross-person heterogeneous settings, we considered a fixed body position and three participants data for that particular body position with an assumption of labeled data availability from two participants. The motivation behind this setting is that a participant might have activity execution similarity for different activities that of activities from different individuals. In the experiment, each participant's labeled data acts as a source domain and participant with unlabeled data as the target domain. In a given dataset, we considered all the available body-positions 4.1, and within each body-position, we experimented with all possible unique combinations of available participants from the total participants.
2) *Cross-position Heterogeneity:* Under cross-position heterogeneous settings, we considered the heterogeneity of different body-positions within a participant. In the experiment, we evaluated *MSADA* with three different body positions assuming that the labeled data availability from two body positions and unavailability for the third position. We refer to the labeled body-positional data as the source domain and the unlabeled body-positional data as the target domain. We consider all possible unique body-positions combinations of three from all the available body-positions for all the participants.

4.6. Results

We report the experimental results and findings in terms of accuracy under different heterogeneous data distribution settings for OPPORTUNITY, PAMAP2, and DSADS dataset.

1) *OPPORTUNITY:* Fig. 3 plots the detailed experimental results on the OPPORTUNITY dataset. The left sub-figure presents the cross-user heterogeneity for different body positions and whereas, the right sub-figure details the cross-position heterogeneity for different users. We report the results at a granular level in terms of individual user and position with the baseline methods.

[5] https://github.com/soumith/ganhacks.

Table 2
Hyper-parameters of *MSADA* framework.

Hyper-parameters	Values
No. of conv. Layers	3
No. of filters in conv. layers	32, 64, 128
Conv. filter dimension	1×9, 1×9, 1×9
No. of fully connected layers in Domain Discriminator	2
No. of units in fully connected layers Domain Discriminator	128, 1
Batch size	32
ADAM (Kingma et al., 2014) optimizer parameter	0.9, 0.99
Learning rate	0.0001
Pre-training epoch	150
MSADA training epoch (β)	200

For cross-person heterogeneity, *MSADA* outperforms all the baselines except the deep learning-based Rev-Grad. It is important to note that in this setting, we consider multiple participants the same body-positional data. To adapt the baselines, we combine two source domain data and use it as a single source domain. As a deep learning framework, RevGrad observes more variation in terms of *participant data*. Such variation gives a slight advantage to the baseline methods. Even it is unfair to *MSADA*, *MSADA* performs better than the baseline methods.

In the case of the cross-position heterogeneity, we consider the different unique combinations of body-positions of a participant. For each participant, we find the unique positional combinations of $\binom{5}{3}$ where two positions act as the source domain and the other as the target domain. Within each unique combination of three positions, we consider the permutations of three positions so that each position can be considered as the target domain once. We repeat the same procedure for all the participants. *MSADA* performs better than baselines.

2) *PAMAP2*: Fig. 4 plots the experimental results over two different settings for the PAMAP2 dataset. In the PAMAP2 dataset, we found only 4 participants with 11 activity data from all three body positions. In a person-heterogeneous environment, *MSADA* performs on par with RevGrad and BDA, whereas STL performs better among all the methods. We observe that CORAL performs the worst (10% accuracy) as well as in other datasets. In the case of the position-heterogeneous setting, all the methods except BDA perform poorly, including the Deep learning-based approaches performing nearly 13–20% accuracy. It is realizable that cross-position heterogeneity is harder than cross-person heterogeneity, but the performance drop in respect to that of the OPPORTUNITY dataset was huge. Such results lead us to further investigation on the PAMAP2 dataset using LDA visual representation. The investigation reveals exploratory findings that we discuss in the discussion 5 section.
3) *DSADS:* Fig. 5 presents the experimental result on the DSADS dataset. We considered several sports activities in the DSADS dataset that we mentioned in 4.1. In the cross-person heterogeneous settings, among five different positions, *MSADA* performs on par with the other baselines. We observe an interesting result in the cross-position experimental settings where *MSADA* outperforms the baseline methods by a significant margin, given the nature of the considered activities. In this setting, we observe the instability of STL to generate results successfully and, henceforth, excluded as a baseline.

4.7. *Effects of perplexity scores*

In this section, we analyze the perplexity scores' effectiveness, which estimates the relevance of the target domain data with the two source domain data. For the analysis, we consider a user-heterogeneity case from the OPPORTUNITY dataset.

We consider a body position (Right Upper Arm) for this analysis and three users - User-2, User-3, User-4, Where User-2 and User-4

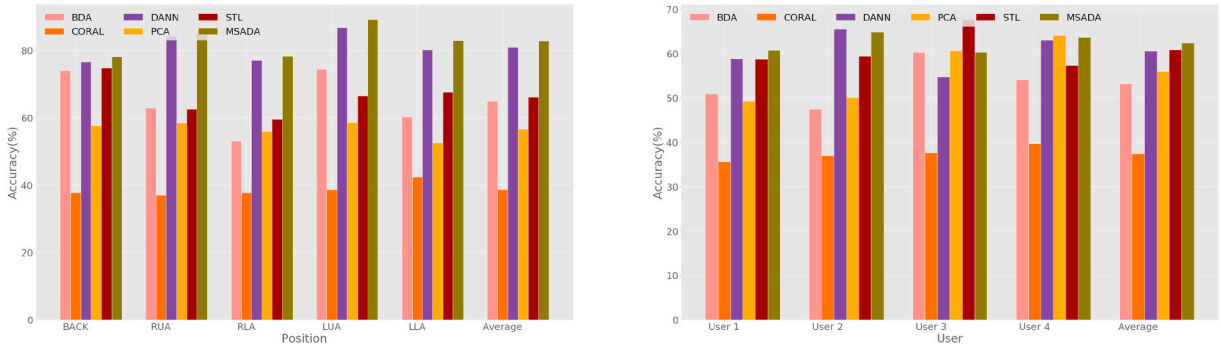

Fig. 3. OPPORTUNITY dataset experimental result comparison of *MSADA* with the baselines (left: cross-person heterogeneity adaptation for a fixed body-position, right: cross-position heterogeneity adaptation for a person).

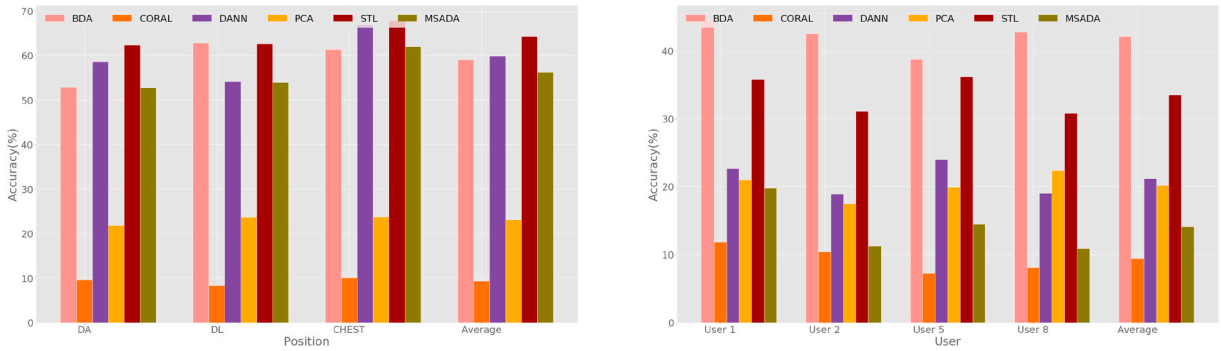

Fig. 4. PAMAP2 dataset experimental result comparison of *MSADA* with the baselines (left: cross-person heterogeneity adaptation for a fixed body-position, right: cross-position heterogeneity adaptation for a person).

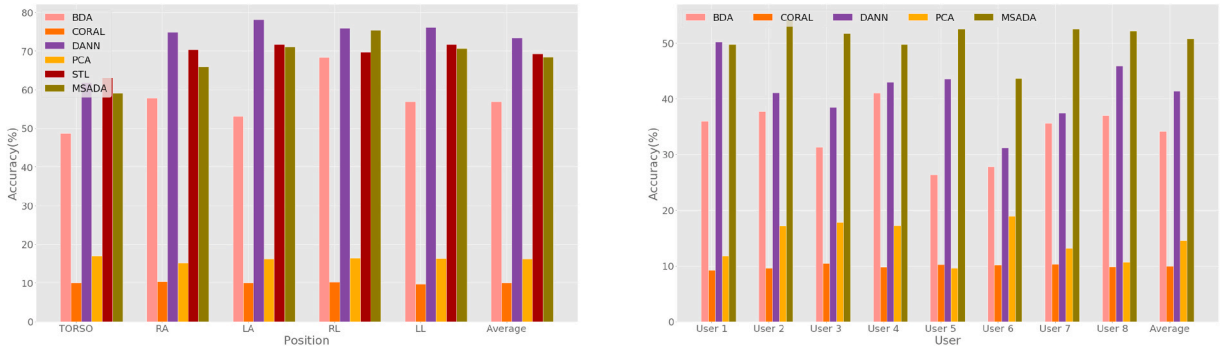

Fig. 5. DSADS dataset experimental result comparison of *MSADA* with the baselines (left: cross-person heterogeneity adaptation for a fixed body-position, right: cross-position heterogeneity adaptation for a person).

were considered the source domains, and User-3 was considered the target domain. After training, the trained perplexity scores for two source domains were 0.502067 and 0.497933, which were reasonable as the body-position for all three users are the same. Once trained, we test the target domain data on the source-1 classifier, source-2 classifier, and *MSADA*, and record the prediction these classifiers made. Figs. 6 and 7 represents the confusion matrix that is generated from the source domain-specific classifiers. It is observable from the confusion matrices that the source-1 classifier is more accurate in identifying the 'Sitting' activity, whereas the source-2 classifier is better in classifying 'Standing' and 'Running' activity. Mis-classification of 'Walking' activity as 'Standing' activity seems to have two possibilities - either 'standing' data may be noisy or 'walking' data may be wrongly annotated.

MSADA utilizes the perplexity scores as a weight to the classifier prediction results. A close observation of the *MSADA* generated confusion matrix in Fig. 8 reveals that the network can take advantage of both source-1 and source-2 classifiers. The confusion matrix

Fig. 6. Source-1 confusion matrix in a cross-person (User-2 as the source-1, User-4 as source-2 and User-3 as the target domain) heterogeneous settings in OPPORTUNITY dataset.

Fig. 7. Source-2 confusion matrix in a cross-person (User-2 as the source-1, User-4 as source-2 and User-3 as the target domain) heterogeneous settings in OPPORTUNITY dataset.

shows that the network preserves the individual classification results and increases prediction accuracy.

Table 3 shows statistics of the correctly predicted target samples. The table reveals the role of the perplexity scores in breaking the ties. Both source-1 and source-2 classifiers correctly predicted 156 samples, whereas there were disagreements on 19 samples. With the perplexity scores, *MSADA* correctly predicts the true labels for these samples and increases the accuracy.

5. Discussion and future work

We studied and investigated the cross-position and cross-person activity recognition using the *MSADA* framework and evaluated its performance extensively on publicly available datasets. In this section, we discuss a few challenges that beg more investigation.

Number of Publicly Available Datasets: We found only countable public datasets that support the multi-source domain adaptation problem formulation. We faced issues with the public datasets regarding data availability from multiple body-positions, the number of data samples and types of activities to fit our designed problem statement. Specifically, for the cross-position heterogeneity settings, the availability of at least three body-positional data is mandatory. For example, we considered Heterogeneity Activity Recognition Data Set (HHAR), but it provides activity data only from two different body positions - Hand and Waist. Moreover, the number of data entries is another essential aspect to take advantage of the deep learning frameworks over traditional machine learning approaches that do not support transfer learning in general. To elaborate, as we considered macro activities, we used a larger windowing size of 128 in the data preprocessing. Even though we have used an overlapping of 90% between two consecutive windows, the number of windowed samples was still low in using deep learning frameworks. Having a larger dataset that provides multiple body-positional data would help the network generalize better and increase performance.

PAMAP2 Dataset Inconsistency: In our experiments, the proposed approach and other baselines did not perform well in the PAMAP2 dataset, which prompted us to inspect PAMAP2 and DSADS dataset visually. We considered three body-positions - TORSO, Right leg, and Right Arm from two participants from both datasets. We applied the windowing technique to extract windowed data and calculated nine time-domain features and nine frequency-domain features for Linear Discriminant Analysis (LDA) presentation. We use the SciPy library as the tool for visualization, and Figs. 9 and 10 plots the LDA visualization of Participant-1 and Participant-8 from PAMAP2 and DSADS dataset, respectively. Fig. 10 presents the opposite alignment of activities from one body-positional data against the two other positions, which are not the case for the DSADS dataset. Such findings could probably result and explain the low performance of *MSADA* in a cross-position heterogeneous environment as such a scenario presents a challenging setting for domain adaptation. We suspect that the PAMAP2 dataset sustains issues with the data annotation for certain activities, although it needs more investigation and experimentation.

Conditional Distribution Alignment: In *MSADA*, we aligned the marginal data distribution of three separate domains for the domain adaptation. In the proposed approach, the classifiers were trained during the pre-training, whereas it is also essential to adapt the classifiers on the target domain data. Such adaptation would be tantamount to conditional data distribution alignment, which could increase the overall performance. As the label for the target domain is unavailable, aligning conditional data distribution is not trivial. The pseudo-labeling concept could be investigated further to acquire labels for a few unlabeled target domain instances. Pseudo-labeling also helps in adapting the class-wise decision boundaries. Adapting marginal distribution and conditional distribution simultaneously in sensor-based domain adaptation is still an open arena for exploration.

Fig. 8. *MSADA* confusion matrix in a cross-person (User-2 as the source-1, User-4 as source-2 and User-3 as the target domain) heterogeneous settings in OPPORTUNITY dataset.

Table 3
Effect of perplexity scores on a cross-person heterogeneous settings in opportunity dataset.

Classifier	Correct Prediction
S1 Classifier	165
S2 Classifier	173
S1 and S2 Classifier Agreement	156
MSADA	175

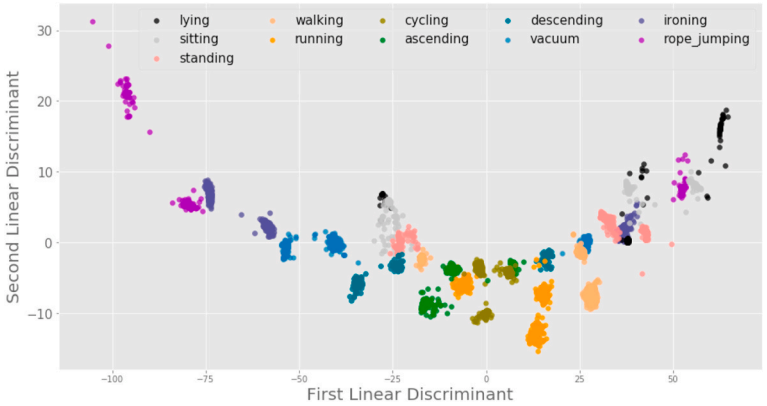

Fig. 9. PAMAP2 dataset LDA presentation of 11 activities from three body-positions of User-5. The presentation reveals the opposite alignment of activities in at least one body-positional samples.

6. Conclusion

The scarcity of annotated data is a widespread phenomenon in a device sparse environment; selecting the relevant domain and annotating these datasets makes the task even more challenging. In this work, we propose a deep-learning-based framework that incorporates novel perplexity score for data annotation in the presence of multiple heterogeneous data distribution. The perplexity scores help to identify the most relevant source domains. We evaluate the proposed framework with multiple publicly available datasets that consist of daily living activities and sports activities. We also analyze the effectiveness of the novel perplexity score that

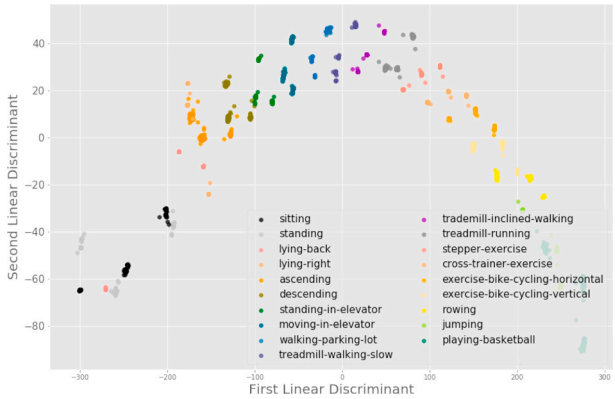

Fig. 10. DSADS dataset LDA presentation of 19 activities from three body-positions of User-8. The presentation reveals the homogeneous alignment order of activities in three body-positional samples.

dynamically helps to identify and select the most relevant source domain from multiple available source domains, which is most often unacknowledged in many approaches that consider heterogeneous data distribution.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Akbari, A., & Jafari, R. (2019). Transferring activity recognition models for new wearable sensors with deep generative domain adaptation. In *Proceedings of the 18th international conference on information processing in sensor networks* (pp. 85–96). ACM.

Barshan, B., & Yüksek, M. C. (2014). Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *The Computer Journal, 57*(11), 1649–1667.

Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and Information Systems, 36*(3), 537–556.

Faridee, A. Z. M., Khan, M. A. A. H., Pathak, N., & Roy, N. (2019). Augtoact: Scaling complex human activity recognition with few labels. In *Proceedings of the 16th EAI international conference on mobile and ubiquitous systems: Computing* (pp. 162–171). Networking and Services.

Faridee, A. Z. M., Ramamurthy, S. R., Hossain, H., & Roy, N. (2018). Happyfeet: Recognizing and assessing dance on the floor. In *Proceedings of the 19th international workshop on mobile computing systems & applications* (pp. 49–54). ACM.

Ganin, Y., & Lempitsky, V. (2014). Unsupervised domain adaptation by backpropagation. In *arXiv preprint arXiv:1409.7495*.

Haeusser, P., Frerix, T., Mordvintsev, A., & Cremers, D. (2017). Associative domain adaptation. In *Proceedings of the IEEE international conference on computer vision* (pp. 2765–2773).

Hossain, H. S., Khan, M. A. A. H., & Roy, N. (2017). Soccermate: A personal soccer attribute profiler using wearables. In *2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)* (pp. 164–169). IEEE.

Jeyakumar, J. V., Lai, L., Suda, N., & Srivastava, M. (2019). Sensehar: A robust virtual activity sensor for smartphones and wearables. In *Proceedings of the 17th conference on embedded networked sensor systems* (pp. 15–28).

Karbalayghareh, A., Qian, X., & Dougherty, E. R. (2018). Optimal bayesian transfer learning. *IEEE Transactions on Signal Processing, 66*(14), 3724–3739.

Khan, M. A. A. H., Roy, N., & Misra, A. (2018). "Scaling human activity recognition via deep learning-based domain adaptation. In *2018 IEEE international conference on pervasive computing and communications (PerCom)* (pp. 1–9). IEEE.

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980.

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. In *Advances in neural information processing systems* (pp. 971–980).

Mansour, Y., Mohri, M., & Rostamizadeh, A. (2009). Multiple source adaptation and the rényi divergence. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 367–374). AUAI Press.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). *Automatic differentiation in pytorch*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res., 12*(Oct), 2825–2830.

Reiss, A., & Stricker, D. (2012). Introducing a new benchmarked dataset for activity monitoring. In *"2012 16th international Symposium on wearable computers*. pp. –, IEEE.

Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., et al. (2010). Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)* (pp. 233–240). IEEE.

Rokni, S. A., & Ghasemzadeh, H. (2017). Synchronous dynamic view learning: A framework for autonomous training of activity recognition models using wearable sensors. In *Proceedings of the 16th ACM/IEEE international conference on information processing in sensor networks* (pp. 79–90). ACM.

Sankaranarayanan, S., Balaji, Y., Castillo, C. D., & Chellappa, R. (2018). Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8503–8512).

Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., Sonne, T., & Jensen, M. M. (2015). Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM conference on embedded networked sensor systems* (pp. 127–140). ACM.

Sun, B., Feng, J., & Saenko, K. (2017). *"Correlation alignment for unsupervised domain adaptation," in domain Adaptation in computer vision applications*. Springer.

Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision* (pp. 443–450). Springer.

Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7167–7176).

Um, T. T., Pfister, F. M. J., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U., & Kulić, D. (2017). "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. ICMI 2017. In *Proceedings of the 19th ACM international conference on multimodal interaction* (pp. 216–220). New York, NY, USA): ACM.

Wang, J., Chen, Y., Hao, S., Feng, W., & Shen, Z. (2017). Balanced distribution adaptation for transfer learning,. In *2017 IEEE international conference on data mining (ICDM)* (pp. 1129–1134). IEEE.

Wang, J., Chen, Y., Hu, L., Peng, X., & Philip, S. Y. (2018a). "Stratified transfer learning for cross-domain activity recognition. In *2018 IEEE international conference on pervasive computing and communications (PerCom)* (pp. 1–10). IEEE.

Wang, J., Zheng, V. W., Chen, Y., & Huang, M. (2018b). Deep transfer learning for cross-domain activity recognition. In *Proceedings of the 3rd international conference on crowd science and engineering* (Vol. 16)ACM.

Wen, J., Indulska, J., & Zhong, M. (2016). "Adaptive activity learning with dynamically available context. In *2016 IEEE international conference on pervasive computing and communications (PerCom)* (pp. 1–11). IEEE.