



Cross-User Activity Recognition via Temporal Relation Optimal Transport

Xiaozhou Ye^(✉)  and Kevin I-Kai Wang 

Department of Electrical, Computer, and Software Engineering,
The University of Auckland, Auckland, New Zealand
xye685@aucklanduni.ac.nz, kevin.wang@auckland.ac.nz

Abstract. Current research on human activity recognition (HAR) mainly assumes that training and testing data are drawn from the same distribution to achieve a generalised model, which means all the data are considered to be independent and identically distributed (*i.i.d.*). In many real-world applications, this assumption does not hold, and collected training and target testing datasets have non-uniform distribution, such as in the case of cross-user HAR. Domain adaptation is a promising approach for cross-user HAR tasks. Existing domain adaptation works based on the assumption that samples in each domain are *i.i.d.* and do not consider the knowledge of temporal relation hidden in time series data for aligning data distribution. This strong assumption of *i.i.d.* may not be suitable for time series-related domain adaptation methods because the samples formed by time series segmentation and feature extraction techniques are only coarse approximations to *i.i.d.* assumption in each domain. In this paper, we propose the temporal relation optimal transport (TROT) method to utilise temporal relation and relax the *i.i.d.* assumption for the samples in each domain for accurate and efficient knowledge transfer. We obtain the temporal relation representation and implement temporal relation alignment of activities via the Hidden Markov model (HMM) and optimal transport (OT) techniques. Besides, a new regularisation term that preserves temporal relation order information for an improved optimal transport mapping is proposed to enhance the domain adaptation performance. Comprehensive experiments are conducted on three public activity recognition datasets (i.e. OPPT, PAMAP2 and DSADS), demonstrating that TROT outperforms other state-of-the-art methods.

Keywords: Human activity recognition · Out-of-distribution · Domain adaptation · Transfer learning · Time series classification

1 Introduction

Human activity recognition (HAR) is a crucial research area in Human-Computer Interaction, ubiquitous computing [3] and the Internet of Things [1], which aims to correctly classify the current activity of a human or a group of

humans performs based on physical sensor observations and context information [18]. HAR is applied to a variety of applications, including medical treatment, ambient assisted living, fitness and sports, rehabilitation, security surveillance, health monitoring, automatic security and home automation. Existing state-of-the-art HAR methods for processing time series sensor data are under the assumption that training and testing data are drawn from the same distribution, which means all the data are independent and identically distributed *i.i.d.* [38]. This assumption implies that a generalised model is achieved across the source training data and the target testing data. In other words, the training and testing data are taken from the same domain where the feature space and data distribution characteristics are the same [23]. However, in many real-world applications, this assumption does not hold, and the collected training and target testing datasets have non-uniform distribution due to the issue of data heterogeneity or sometimes referred to as the out-of-distribution (*o.o.d.*) problem. The performance of a model that is trained on the source domain will likely decline when tested on the target domain [24]. Therefore, it is necessary to consider a more practical scenario where the data distributions of the training and testing set can be different.

Transfer learning (or more specifically domain adaptation) [23] is a promising method for handling data heterogeneity. In general, the principle and key idea of transfer learning is to find some common knowledge and reduce the data distribution differences between the source domain and target domain. Most of the current domain adaptation research focus on static data area such as computer vision with the assumption that each sample(image) is *i.i.d.* in its corresponding domain [24]. The same assumption is also applied to time series transfer learning research, such as sensor-based HAR, which means each sliding window is *i.i.d.* in its corresponding domain [8, 11, 30, 40]. However, time series data typically have inherent temporal relation, which means each sliding window is not independent of each other [13]. For an intuitive example of activity recognition, walking is made up of three sub-activities: raising the leg, thrusting forward and feet to the ground. These three sub-activities have temporal dependency relations. So, existing domain adaptation approaches do not fully utilize the characteristics of time series data effectively because temporal relation knowledge is overlooked in the process of aligning data distribution.

In this paper, we propose a novel temporal relation optimal transport (TROT) method targeting time series domain adaptation for cross-user activity recognition, which considers temporal relations across time. This method captures the data distributions of the common sub-activities while preserving the same temporal relation between the sub-activities across source and target users. Temporal relation preservation acts like a regularization for reducing the data distribution differences of the sub-activities across users. In this case, the data distribution mapping process can achieve better domain adaptation performance. Specifically, it is a combination of optimal transport and the Hidden Markov Model (HMM) method that performs cross-user HAR considering temporal relation knowledge. Moreover, we propose a new regularizer for the optimal transport domain adaptation problem that preserves temporal relation

order information to ensure correct data distribution alignment. Comprehensive experiments on three public activity recognition datasets demonstrate the better performance of TROT in time series domain adaptation applications (e.g. cross-user HAR). The rest of the paper is organized as follows: Sect. 2 presents the related work, delving into cross-user HAR and the concepts of transfer learning and domain adaptation. Section 3 describes the proposed method, including problem formulation, the motivation behind it, and the specifics of the TROT method-focusing on temporal relation representation and OT-based data distribution alignment. In Sect. 4, we detail the experimental setup and provide a comparison between our method and existing methods. Finally, Sect. 5 offers conclusions drawn from the study and suggests potential avenues for future research.

2 Related Work

2.1 Cross-User Human Activity Recognition

HAR is one crucial part of ubiquitous computing because of its important role in supporting daily human life. It aims at recognizing and analysing human behaviours from learned high-level knowledge from multi-modal sensor observations and context information. Various types of sensors are utilized to implement activity recognition in different scenarios. From the sensor modality view, HAR can be classified into five types: Smartphones/wearable sensors-based HAR, ambient sensors-based HAR, device-free sensors-based HAR, vision-based sensors HAR and other modality sensors-based HAR [7, 19, 35]. In this paper, we focus on wearable sensors-based HAR.

Sensor-based HAR is considered a problem of time series classification in machine learning research [3]. A great number of classification models, such as ensemble learning [29], SVM [5], and HMM [2] are proposed to solve the HAR problem. With the recent research development, deep learning has achieved many state-of-the-art results for a wide variety of tasks. Deep learning-based HAR approaches [12] are able to learn high-level features and automatically extract features from massive amounts of data [33]. However, these approaches mostly depend on the assumption that training and testing data are drawn from the same distribution to ensure the generalization ability of the model, which means all the data are *i.i.d.* [38]. In many real-world applications, this assumption does not hold, and the collected training and testing datasets are *o.o.d.*. In this paper, we focus on the sensor-based HAR *o.o.d.* problem.

There are various categories of sensor-based HAR *o.o.d.*: First, data is generated from different sensors. Different sensor types, platforms, manufacturers and modalities may cause different data formats and distributions [39]. Second, the data pattern may change over time, which is also called concept drift [21]. For example, the walking pattern of the same person may be impacted by health status. Third, the behaviour differentiation between different people may also be significant [28]. For instance, different people may walk at different paces. Forth, physical sensors may be installed in different body positions [27] or environmental layouts in smart homes [31], which may lead to different data distribution.

Our current research focuses on the sensor-based HAR *o.o.d.* of behaviour differentiation between different people.

2.2 Transfer Learning and Domain Adaptation

Transfer learning is a process that enables a model to be trained in one or more source domains and applied to one or more related target domains that have no label or few labels. During the process, one of the key tasks is to solve the *o.o.d.* problem by reducing the distribution difference between source and target domains. Domain adaptation [24,38] is a special type of transfer learning. It still focuses on solving *o.o.d.* problem between source and target domains but adds an additional constraint of assuming the same task of source and target domains. This paper focuses on the unsupervised domain adaptation problem, which means the availability of labelled data from the source domain and unlabelled data from the target domain [38].

Domain adaptation has been developed for many years, and most of the approaches are feature-based transfer learning [23]. The Large Margin Nearest Neighbour (LMNN) approach, as presented in [37], utilizes Mahalanobis distance metric learning. The goal of this method is to learn a linear transformation, such that data from different classes are distinctly separated from each other within the feature space. Subspace Alignment(SA) [14] applied PCA to generate subspace for the source domain and target domain. Then, SA is achieved by minimizing Bregman matrix divergence to learn a linear transformation matrix. CORrelation Alignment(CORAL) [32] learns a linear transformation by minimizing the distance between the second-order statistics of the source domain and target domain features. Optimal transport(OT) theory is a promising approach and is also applied to solve the domain adaptation problem [9] via Sinkhorn distances. There are several works extending the research of optimal transport domain adaptation. Optimal Transport for Domain Adaptation(OTDA) [15] is proposed to learn the coupling between two probability density functions. The source domain is then transformed into the target domain via barycentric mapping. Substructural optimal transport(SOT) [22] explores the substructure of domains to complete substructure-level mapping to achieve a balance between coarse-grained mapping and fine-grained mapping.

These feature-based domain adaptation methods focus on static data such as images, whereas time series data is often treated directly by the same domain adaptation framework [20,22]. However, these existing domain adaptation approaches may not work well with time series data because the temporal relation knowledge from the time series data is overlooked in the process of aligning data distribution. Temporal relation is an essential characteristic of time series data. Existing approaches that only make use of temporal invariant knowledge lead to models with unreliable performance. This is especially the case in real-world cross-user HAR applications, where a model is trained by one (or more) user(s) and later applied to other user(s). Furthermore, temporal

relation knowledge may be useful for finding the common knowledge between users for better sensor-based HAR domain adaptation. The focus of this paper is the exploration of temporal relation knowledge.

3 Method

3.1 Problem Formulation

In a cross-user HAR problem, a labelled source user $S^{train} = \{(x_i^{train}, y_i^{train})\}_{i=1}^{n^{train}}$ drawn from a joint probability distribution $P^{Source}(X, y)$ and a target user $S^{test} = \{(x_i^{test}, y_i^{test})\}_{i=1}^{n^{test}}$ drawn from a joint probability distribution $P^{Target}(X, y)$, where n^{train} and n^{test} are the number of source and target samples respectively. S^{train} and S^{test} have the same feature space and label space. In real-world applications, a source user and a target user are highly likely to have different data distributions, i.e., $P^{Source}(X, y) \neq P^{Target}(X, y)$. Given source user data $\{(x_i^{train}, y_i^{train})\}_{i=1}^{n^{train}}$ and target user data $\{(x_i^{test})\}_{i=1}^{n^{test}}$, the goal is to achieve correct activity classification for unlabeled target user data.

3.2 Motivation

In general, the principle and key idea of domain adaptation are to reduce the difference between the source domain(s) and the target domain(s) and bridge the gap in the data distributions between the domains. Existing cross-user HAR domain adaptation approaches are mainly based on sample level, user level, activity level and sub-activity level as shown in Fig. 1. However, these methods ignore the temporal relation across chronological time series samples and treat samples as *i.i.d.* in different domains. In fact, temporal relation knowledge is an important property hidden in time series data that can be used as a bridge for cross-user domain adaptation. Different users' time series activity data follow the same temporal relation in various activities. For example, the walking activity includes three sub-activities, namely raising the leg, thrusting forward and feet to the ground, regardless of different users. These temporal relations are valuable knowledge that can better support data distribution alignment across different users. Therefore, we propose a new method called TROT that introduces temporal relation knowledge for solving domain adaptation.

3.3 Temporal Relation Optimal Transport

In this section, we propose an OT implementation that utilizes temporal relation knowledge called TROT. In TROT, HMM is utilized to capture temporal relation knowledge, and OT is used to align the data distributions between the source user and the target user. In addition, we propose to add a new regularizer which preserves temporal order information to enhance OT optimization for better domain adaptation performance. TROT method mainly includes four

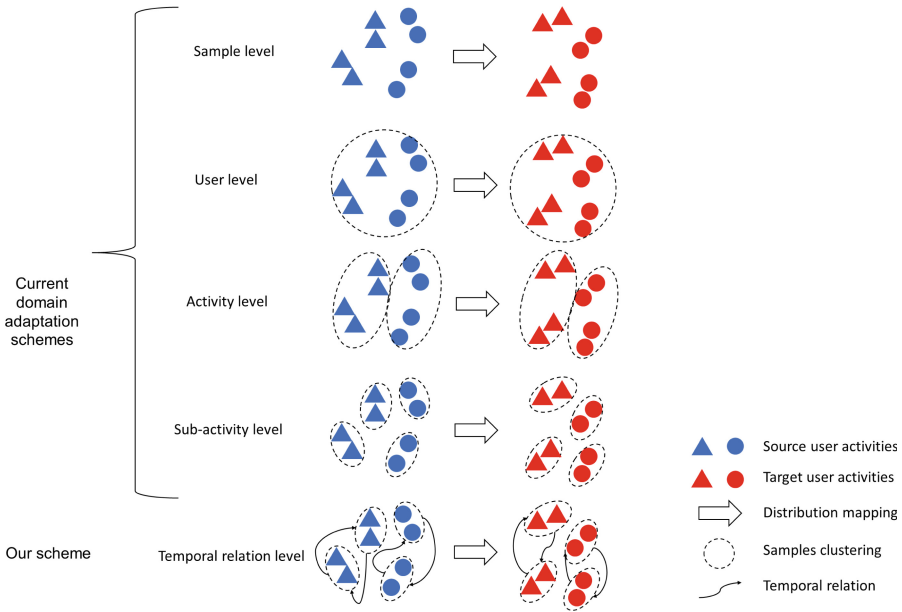


Fig. 1. Comparison of different domain adaptation schemes in cross-user HAR.

steps as listed below, and detailed explanations and discussions are provided in the following subsections. Figure 2 shows the overall TROT learning process.

1. **Data pre-processing:** Data segmentation and feature extraction for both source and target users is implemented as in Fig. 2 from figure (a) to figure (b). In Fig. 2, the red colour is associated with the source user, while the blue colour is associated with the target user. The different shapes, such as circle and triangle, represents different activity classes. Sensor-based HAR data consists of observations taken sequentially over time. When working with time series data, the sliding window technique is often employed for data segmentation. Then, for each window of data, the time and frequency domain features are extracted. Time domain features capture raw temporal patterns in the data, while frequency domain features provide insight into the rhythmic or periodic components of the data. These features are crucial in cross-user HAR because they can capture general patterns of human movement that are less dependent on individual idiosyncrasies, thereby improving the model’s ability to generalize across different users. In this step, the raw data are transformed into samples with features.
2. **Temporal relation extraction:** Temporal state information is extracted via HMM for source and target users as in Fig. 2 from figure (b) to figure (c). The number in each shape indicates the corresponding temporal state information hidden in the time series data. For example, the numbers one, two and three in triangles may correspond to raising the leg, thrusting forward

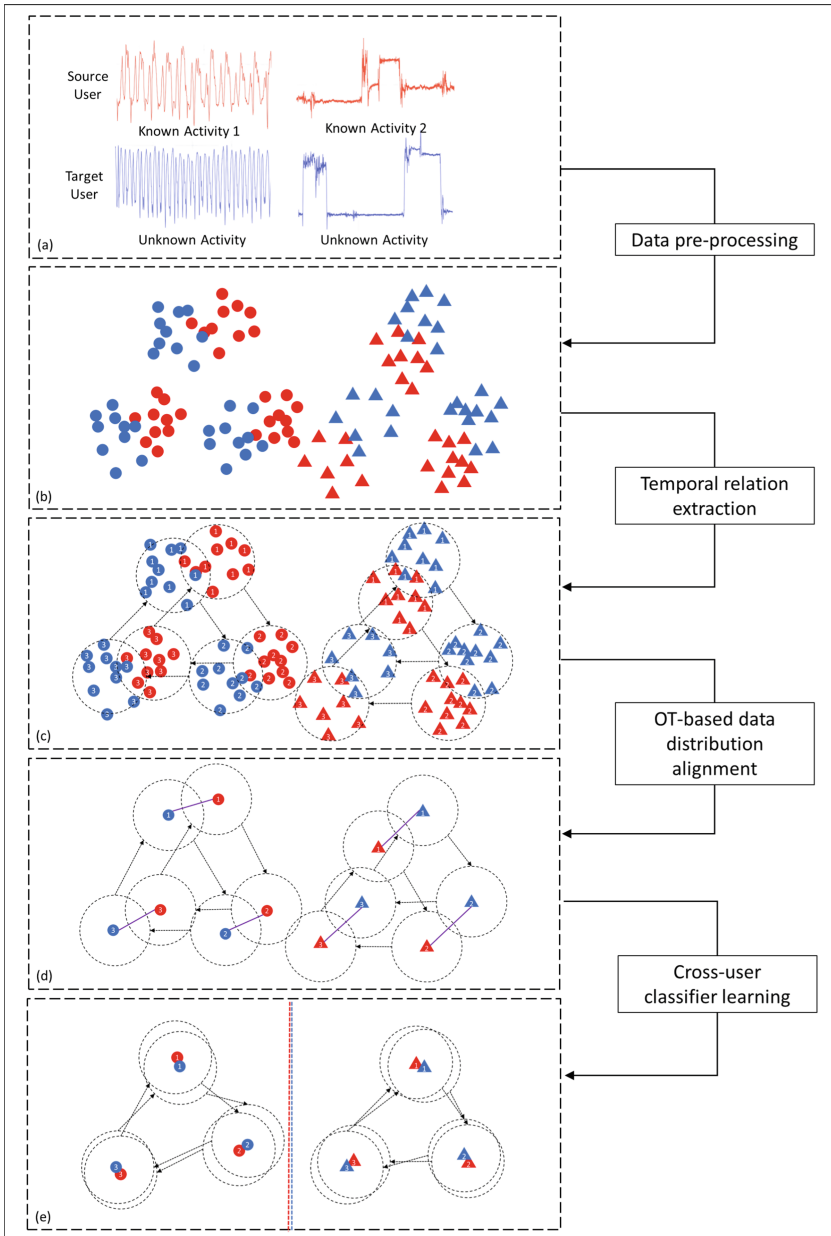


Fig. 2. TROT Learning Process Figure.

and feet to the ground of walking activity. The black dash arrow between the numbers means the temporal relation. The learned temporal relation knowledge is the common knowledge across users that is useful to bridge the gap between source and target users for better domain adaptation. In this step, the samples are clustered and form several clusters based on their temporal relation information.

3. **OT-based data distribution alignment:** The data distribution alignment task is achieved by transforming the samples from source user to target user as in Fig. 2 from figure (c) to figure (d). The cluster centres are used to represent the corresponding cluster, and then the centres of the source user are projected to the centres of the target user considering the temporal order information for learning the mapping plan. The purple solid line in Fig. 2 means the data distribution transformation from the source user to the target user. In this step, the samples from the source user are transformed based on the learned mapping plan.
4. **Cross-user classifier learning:** After the above data distribution alignment step between source and target users, the decision boundary of activity classes of the source and target users should be relatively consistent. Therefore, the transformed source user samples with labels are utilized for training a classifier for recognizing target user samples as in Fig. 2 from figure (d) to figure (e). The red dashed line is the decision boundary of activity classes of the source user, while the blue dashed line is the decision boundary of the target user. In this step, the learned classifier from the transformed source user is also suitable for the target user.

The highlight of the TROT method is the temporal relation extraction step and the OT-based data distribution alignment step. Discussion is focused on these two steps.

3.3.1 Temporal Relation Extraction

As in Fig. 2, after the data pre-processing step, there is implicit temporal relation information that needs to be captured. Here, we consider HMM [25] learning for temporal relation representation. HMM is the directed temporal probability graph model that describes the dependency of two related sequences. These two related sequences are called state sequence $S = (s_1, s_2, s_3, \dots, s_T)$ and observation sequence $X = (x_1, x_2, x_3, \dots, x_T)$. $Q = (q_1, q_2, q_3, \dots, q_N)$ is the state set of all the possible states. The value of the state sequence at t time point s_t is only related to the value of the state sequence of $t - 1$ time point s_{t-1} , and the value of the observation sequence at t time point x_t is only related to the value of the state sequence at t time point s_t . Here, there are three types of probability that HMM needs to learn: 1) $p(S_1)$ initial state probability, 2) $p(S_t|S_{t-1})$ state transition probability and 3) $p(X_t|S_t)$ emission probability.

For the cross-user HAR scenario, each sub-activity can be seen as a state. The projection of the observed samples and the corresponding sub-activity can be learned via HMM emission probability. The temporal relation between the

sub-activities can be learned via HMM state transition probability. In this way, the hidden temporal information can be captured and represented.

In detail, initial state probability is the possibility of the first sub-activity and is assumed to be the same across users in our study. Therefore, the first captured state is set as the initial set for the source user and the target user. Emission probability is the probability of a sample occurring, given a specified sub-activities (i.e. state) at time point t . The observation sequence, X_t , can be expressed as a function of S_t based on the learning of emission probability, which means $X_t = f(S_t)$. For the emission probability, we assume it conforms to Gaussian distribution. The emission probability in each class can be expressed as $B = [b_k]_{N \times 1}$, $b_k = \mathcal{N}(\mu_k, \sigma_k)$, $X_k \sim b_k$, $k = 1, 2, \dots, N$, N is the number of sub-activities in each class. Here, μ_k means the centre value of k^{th} state's Gaussian distribution, σ_k means the covariance value of k^{th} state's Gaussian distribution, and X_k means the samples belong to the k^{th} state. Assuming there are C activity classes, then there are a total of $N \times C$ Gaussian distributions need to be learned for each user, i.e., $B^{Source} = [\mathbf{b}^{Source}]_{N \times C}$ and $B^{Target} = [\mathbf{b}^{Target}]_{N \times C}$. In this way, we abstract the observation samples to a higher level of sub-activities, and then the temporal relation extraction happens at the sub-activities level.

The state transition probability is the probability of a sub-activities at time point $t-1$ change to another sub-activity at time point t . For the state transition probability, we reduce the difference between the source user and the target user via common temporal relation knowledge of the human activity. The common knowledge is that the state transition follows the sub-activities order because of the continuity of human activities, which means the current state can only change to the following state in a future time point. For example, walking activity should first raise the leg and then thrust forward in the next state, but not jump to the last state of feet to the ground step. Therefore, $p(S_t|S_{t-1})$ state transition probability can be expressed as a $N \times N$ matrix $A = [a_{i,j}]_{N \times N}$, $a_{i,j} = P(S_t = q_j | S_{t-1} = q_i)$, $i, j = 0, 1, 2, \dots, N, i \neq j$. $a_{i,j}$ means the transition probability from state q_i in $t-1$ time point to state q_j in t time point. There are total N sub-activities. q_0 is the initial state and q_N is the ending state. If an activity such as running or going upstairs has periodic cycle characteristics, then $q_0 = q_N$. The form of the state transition matrix for HMM is given below. Here, we set $\forall i \ a_{i,i+1} = 1$ to make the state transition probability a common temporal relation knowledge in both the source and target users.

$$\begin{bmatrix} a_{0,1} & 0 & \dots & \dots & 0 & 0 \\ 0 & a_{1,2} & 0 & \dots & \dots & 0 \\ 0 & 0 & a_{2,3} & 0 & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 0 & 0 & a_{N-1,N} \end{bmatrix}$$

With the observation sequence X , our goal is to calculate the parameters of emission probability in each activity for each user. The process of parameter learning can be achieved by Expectation Maximum (EM) [10] algorithm. After

the HMM parameter learning, the source user data distribution can be expressed as B^{Source} , and the target user can be expressed as B^{Target} . As Fig. 2 (c) shows, now each class is represented by multiple temporally-related Gaussian distributions. Our next task is the alignment of sub-activities in each class between source and target users.

3.3.2 Optimal Transport Based Data Distribution Alignment

Once the temporal-related distribution is extracted by HMM, the data distribution alignment is achieved via an OT-based method that maps the Gaussian distributions between the source and target users. OT has drawn attentions in recent years for domain adaptation, and it aims at moving source domain data distribution on top of target domain data distribution with an optimization transportation plan [17].

Based on the OT-based method, the source user and target user can be expressed as data distributions respectively:

$$P^{Source} = \sum_{i=1}^{i=k^{Source}} w_i^{Source} \delta_{\mu_i^{Source}},$$

$$P^{Target} = \sum_{j=1}^{j=k^{Target}} w_j^{Target} \delta_{\mu_j^{Target}}.$$

Here, $\sum_{i=1}^{i=k^{Source}} w_i^{Source} = 1$, and $\sum_{j=1}^{j=k^{Target}} w_j^{Target} = 1$. μ is the center value of Gaussian distributions in B learned from HMM, and δ_{μ} is the Dirac function corresponding to the location of μ . The probabilities w are the probability masses associated with μ . In our case, $k^{Source} = k^{Target} = C \times N$, where C is the number of activity classes, and N is the number of states. We set $\forall i, j, w_i^{Source} = w_j^{Target} = \frac{1}{C \times N}$.

Similar to [15], we also apply regularization approach to OT to prevent overfitting and keep structural risk minimization. Entropy regularization [9] $H(\gamma)$ is introduced with the idea of lowering the sparsity of coupling matrix γ via increasing its entropy to find a smoother optimal transport plan. Here, the coupling matrix represents the transport plan between two probability distributions as source user and target user. The smoother optimal transport plan often translates to being more generalizable, which is beneficial for domain adaptation, instead of having very specific, point-to-point transport, which might overfit the source user. Group-sparse regularization [15] $\Omega(\gamma)$ considers the use of label distribution knowledge that makes each sample in the target domain receive masses only from the samples in the source domain with the same label. By preserving the inherent label structure of the data, group-sparse regularization promotes the correct transfer of class structures between domains and makes models more generalizable to the target user.

Besides, we propose to add a new regularization term to the regularized optimal transport that preserves temporal order information. The main intuition underlying this temporal order preservation regularization is that even two

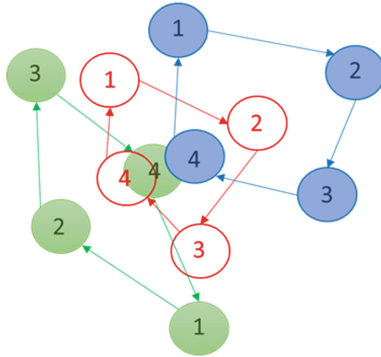


Fig. 3. OT-based Temporal Order Preservation Regularization Figure.

Gaussian distribution centres may be closer between the source user and target user; however, different temporal relation order may lead to wrong data distribution alignment. As Fig. 3 shows, the red is an activity from the source user, and the blue and green are two different activities from the target user. In this example, both green and blue are very close to red in terms of the overall distance between the centres of sub-activities distributions. However, blue is the only suitable one after considering the temporal order of the state Gaussian distribution.

Table 1. Temporal Order Mapping Between Source and Target Users Table.

	1	2	3	4	5	6	7	8
	1	2	3	4	1	2	3	4
1	1	✓			✓			
2	2		✓				✓	
3	3			✓				✓
4	4				✓			✓
5	1	✓				✓		
6	2		✓				✓	
7	3			✓				✓
8	4				✓			✓

This temporal order preservation regularizer is defined as:

$$T(\gamma) = \sum_i \|\gamma(i, I_i)\|_2$$

The coupling matrix γ has the indices of the column, each of which corresponds to a Gaussian distribution from the target user, while the indices of the row correspond to a Gaussian distribution from the source user. For example, in Table 1, the indices of the coupling matrix are from 1 to 8 for both the column and the row. I_i is a set of indices that tell which columns in γ correspond to the

same temporal order as the i^{th} Gaussian distribution of the source user. In other words, I_i helps to find the matching temporal order in the target user for each Gaussian distribution in the source user. As in Table 1, the red and yellow are two activities from the source user, while the blue and green are two activities from the target user. All the activities have four temporal states (i.e. Gaussian distributions). The number from 1 to 4 in each activity represents the temporal order. For example, the temporal order 1 in red activity of the source user matches the temporal order 1 in blue and green activities of the target user, as the checkmark symbol shows. $\gamma(i, I_i)$ is a vector that is achieved by picking out the coefficients from the i^{th} row of γ that correspond to these matched temporal orders from the column. These coefficients are the transport plan from the i^{th} Gaussian distribution of the source user to the matched Gaussian distributions in the target user.

In summary, this regularizer encourages the Gaussian distributions of the source user to match the Gaussian distributions with the same temporal order of the target user. For example, in Fig. 3, order 1 of red is encouraged to map to order 1 of blue instead of mapping to order 3 of green.

With this temporal order preservation regularization, the masses corresponding to each Gaussian distribution of specified temporal order value match samples of the source and target users with the same temporal order value. So, the overall goal of TROT is:

$$\begin{aligned} \gamma^* &= \arg \min_{\gamma} \langle \gamma, \mathcal{C} \rangle_F + \lambda H(\gamma) + \eta \Omega(\gamma) + \tau T(\gamma) \\ &\text{s.t.} \\ \gamma \mathbf{1}_{k^{\text{Target}}} &= P^{\text{Source}} \\ \gamma^T \mathbf{1}_{k^{\text{Source}}} &= P^{\text{Target}} \end{aligned}$$

$\gamma \in \mathbb{R}^{k^{\text{Source}} \times k^{\text{Target}}}$ is the coupling matrix between the probability distributions of the source domain and target domain, $\mathbf{1}_k$ is k -dimensional vectors of ones, $\langle \cdot, \cdot \rangle_F$ is the Frobenius dot product. \mathcal{C} is the cost matrix of the squared Euclidean distance between two locations, $\mathcal{C}(i, j) = \left\| \mu_i^{\text{Source}} - \mu_j^{\text{Target}} \right\|_2^2$, denotes the cost of pushing forward a probability mass from μ_i^{Source} to μ_j^{Target} . λ, η, τ are coefficients.

The generalized conditional gradient (GCG) approach is used to solve the regularized optimization problem [15]. GCG algorithm addresses the general case of constrained minimization of composite functions defined as

$$\min_{\gamma} f(\gamma) + g(\gamma)$$

$f(\gamma)$ is a differentiable and possibly non-convex function, $g(\gamma)$ is a convex, possibly non-differentiable function. Here, we set

$$\begin{aligned}
f(\gamma) &= \langle \gamma, \mathcal{C} \rangle_F + \lambda H(\gamma) \\
g(\gamma) &= \eta \Omega(\gamma) + \tau T(\gamma) \\
\text{s.t. } & H'(\gamma)
\end{aligned}$$

Now, the TROT task is

$$\gamma^* = \arg \min_{\gamma} \langle \gamma, \mathcal{C} + \lambda \nabla H(\gamma^{iter}) \rangle_F + (\eta \Omega(\gamma) + \tau T(\gamma))$$

$iter$ is the number of iterations. After getting the optimal coupling matrix γ^* , barycentric mapping of the source domain is implemented for mapping the source domain to the target domain as

$$\widehat{P^{Target}} = \text{diag}(\gamma^* \mathbf{1}_{k^{Target}})^{-1} \gamma^* P^{Target}$$

In the final step, transformed source user data is utilized to train a classifier for classifying target user samples. Here, any classifier can be applied to learn a model with $\widehat{P^{Target}}$ and the labels of the source domain y^{Source} . We use the 1-Nearest Neighbour classification method to get labels $\{y_j^{Target}\}_{j=1}^{n^{Target}}$. 1-Nearest Neighbour is a simple instance of the k-Nearest Neighbours algorithm [16], which is a type of instance-based learning. Given a sample from the target user, the algorithm searches the entire samples from the source user to find the single sample that is closest to the sample from the target user. Here, the distance metric is the Euclidean metric. The label of the nearest sample from the source user is then assigned to the sample from the target user.

4 Experiments

4.1 Datasets and Pre-processing

We used three common sensor-based HAR public datasets as Table 2 shows for validating our cross-user transfer learning method. Here, we only use the sensor values of the accelerometer and gyroscope in the position of the right lower arm to explore the practical scenario of using the smartwatch. In the following, the information of each dataset is introduced briefly.

OPPORTUNITY (OPPT) [6] dataset contains recordings of subjects in a daily living scenario performing morning activities. During the recordings, subjects perform the activities without any restriction by following a loose description of the overall actions to perform. The sampling frequency is 30 Hz. Physical activity monitoring (PAMAP2) dataset [26] is collected following a protocol of specified activities for each subject. Over 10 h of data were collected altogether. The sampling frequency is 100 Hz. Daily and Sports Activities Data Set (DSADS) [4] is collected by asking the subjects to perform the activities in their own fashion. Without any given instructions, this is likely to lead to greater inter-subject variations, and would be more similar to real-life situations. Each activity

Table 2. Three sensor-based HAR datasets information Table.

Dataset & Ref.	Subjects	#Activities	Common Activities
OPPT	S1, S2, S3	4	lying, sitting, standing, walking
PAMAP2	1, 5, 6	11	lying, sitting, standing, walking, running, cycling, Nordic walking, ascending stairs, descending stairs, vacuum cleaning, ironing
DSADS	2, 4, 7	19	sitting, standing, lying on back, lying on right, ascending stairs, descending stairs, standing in an elevator still, moving around in an elevator, walking in a parking lot, walking on a treadmill in flat, walking on a treadmill inclined positions, running on a treadmill in flat, exercising on a stepper, exercising on a cross trainer, cycling on an exercise bike in horizontal positions, cycling on an exercise bike in vertical positions, rowing, jumping, playing basketball

is performed by each subject for 5 min. Sensor units are calibrated to acquire data at 25 Hz sampling frequency. The three datasets are selected in order of increasing domain adaptation task difficulty (DSADS 19 activities > PAMAP2 11 activities > OPPT 4 activities). In this case, DSADS dataset contains more activity classes and more similar activities (3 different walking activities), which make it harder to distinguish.

The data pre-processing step for all the methods follows the same setup as the TROT data segmentation and feature extraction step. Time series data is continuous sequence data. The sliding window technique is a common data segmentation method. For the purpose of temporal relation extraction, each window is set as a fixed time interval. Each window is set as 3 s, and overlapping is 50% as the common setting in sensor-based HAR task [34]. Following a similar feature extraction setup as [36], the x, y, z axis data in each sensor of accelerometer and gyroscope are combined as $\sqrt{x^2 + y^2 + z^2}$, 19 features from both time and frequency domains are extracted for a single sensor as in Table 3. Therefore, in each window, there are total of 38 features because of two sensors of accelerometer and gyroscope. Besides, standardization of maximum absolute value scaler is applied to scales and translates each feature individually in the range of $[-1, 1]$ to prevent the side effect of the features with different numerical order.

4.2 Experiments and Methods Comparison

Here are the categories of methods for comparison, including no domain adaptation methods and domain adaptation methods.

Table 3. Features Extraction Per Sensor

ID	Feature	Description
1	Mean	Average value of samples in window
2	Var	Variance
3	STD	Standard deviation
4	Mode	The value with the largest frequency
5	Maximum	Maximum
6	Minimum	Minimum
7	Mean crossing rate	Rate of times signal crossing mean value
8	Range	Maximum minus minimum
9	DC	Direct component
10–14	Five amplitude features	Mean, variance, STD, skewness, kurtosis
15–19	FFT variance	Mean, variance, STD, skewness, kurtosis

1. No domain adaptation:
 - (a) NA: No Adaptation
 - (b) TD: Target domain 1NN
2. Domain adaptation:
 - (a) The usage of temporal relation: TROT
 - (b) No usage of temporal relation:
 - i. LMNN: Large Margin Nearest Neighbour [37]
 - ii. SA: Subspace Alignment [14]
 - iii. OT: Optimal transport [9]
 - iv. CORAL: CORrelation Alignment [32]
 - v. OTDA: Optimal transport for domain adaptation [15]
 - vi. SOT: Substructural optimal transport [22]

The tables provide a comprehensive comparison of different methods applied to the OPPT, PAMAP2, and DSADS datasets. The performance of each method is evaluated based on the transition from one user to another user. Three users are randomly selected from all the users in each dataset. Then, a one-to-one cross-user HAR task is implemented, encompassing all possible pairings between users in each dataset.

For the NA method, the 1-Nearest Neighbour model is trained by source user samples and is applied directly to target user samples without any domain adaptation mapping. For the TD method, the 1-Nearest Neighbour model is trained and predicted by target user samples, which is the optimal baseline that we aim to approach. For the domain adaptation methods, hyper-parameters tuning is performed, and we follow a similar process as suggested in [15] to prevent overfitting on the testing set. The target user is partitioned into two equal datasets the validation set and the test set. For the TROT method that captures the temporal relation, it is particularly crucial to assess its performance on unseen future time-series data that exhibit a distinction from the data utilized for aligning the

distribution. Therefore, the target domain is partitioned in equal time duration of two parts that keep temporal order. The validation set is used to obtain the combination of hyper-parameters that achieve the highest accuracy from the hyper-parameter space. Then, the performance of the testing set is evaluated with the selected combination of hyper-parameters. Classification accuracy on the target domain is the evaluation metric that we used.

Table 4. OPPT Dataset Methods Comparison Table

Method	S1 \rightarrow S2	S1 \rightarrow S3	S2 \rightarrow S1	S2 \rightarrow S3	S3 \rightarrow S1	S3 \rightarrow S2
NA	44.18	54.11	50.36	53.22	47.94	46.53
TD	100	100	100	100	100	100
LMNN	77.77	73.10	75.04	69.94	65.96	71.06
SA	79.50	84.11	83.39	83.22	78.18	76.53
OT	80.37	81.33	81.69	79.05	81.69	78.89
CORAL	74.18	84.74	80.36	84.74	78.67	75.92
OTDA	83.96	80.44	86.42	78.04	86.66	80.62
SOT	75.60	79.94	82.76	73.62	74.09	69.53
TROT	100	100	100	100	100	100

In general, all the domain adaptation methods can achieve increased performance than the baseline method of NA in the OPPT dataset as shown in Table 4. TROT method achieved the highest accuracy across all the six tasks, with a perfect score of 100%. This suggests that the method is extremely efficient and reliable for the OPPT dataset. OTDA also showed promising results, especially in the tasks from S2 to S1 and from S3 to S1. SA and OT methods were relatively consistent in their performance across different tasks. The SA method, in particular, performed quite well, especially in the task from S1 to S3 and S2 to S1.

When applied to the PAMAP2 dataset in Table 5, all the domain adaptation methods except for LMNN achieved better performance than NA in most of the cross-user tasks. Negative transfer learning (i.e., worse performance) happened on LMNN for all the tasks of the PAMAP2 dataset. TROT maintained the best performance compared to other methods. Moreover, SOT and OTDA showed notable performance, with SOT showing particularly high efficiency in the tasks from 1 to 5 and 5 to 1. OTDA showed a good performance in the tasks from 1 to 6 and from 6 to 1.

On the DSADS dataset in Table 6, CORAL, SOT, OTDA and TROT achieved higher performance than NA in more than half of the tasks in the DSADS dataset. TROT maintained its superior performance. Among other methods, TROT performed best, especially in the task from 2 to 4. SOT showed reasonable performance, especially in the tasks from 2 to 7 and from 4 to 2.

Table 5. PAMAP2 Dataset Methods Comparison Table

Method	$1 \rightarrow 5$	$1 \rightarrow 6$	$5 \rightarrow 1$	$5 \rightarrow 6$	$6 \rightarrow 1$	$6 \rightarrow 5$
NA	28.72	42.02	47.91	55.46	60.43	35.75
TD	100	100	100	100	100	100
LMNN	22.25	20.17	26.04	26.41	36.03	21.34
SA	31.33	43.58	49.81	58.82	60.18	40.07
OT	53.69	61.70	57.77	61.22	57.14	57.32
CORAL	32.69	42.62	50.95	57.14	59.04	38.25
OTDA	64.47	65.23	63.08	65.19	66.12	64.47
SOT	69.82	57.80	64.54	61.95	62.48	56.14
TROT	84.80	75.67	78.28	83.05	82.46	73.76

Table 6. DSADS Dataset Methods Comparison Table

Method	$2 \rightarrow 4$	$2 \rightarrow 7$	$4 \rightarrow 2$	$4 \rightarrow 7$	$7 \rightarrow 2$	$7 \rightarrow 4$
NA	48.57	41.12	39.76	38.61	44.24	37.88
TD	100	100	100	100	100	100
LMNN	18.45	16.47	13.13	15.63	11.78	17.67
SA	48.83	40.91	39.86	38.30	43.04	38.82
OT	47.26	35.64	46.59	38.09	36.89	36.63
CORAL	49.35	41.27	40.70	39.24	43.30	39.76
OTDA	48.10	38.25	45.23	45.28	35.28	43.30
SOT	43.23	46.72	48.47	45.64	43.00	41.11
TROT	53.95	52.47	49.27	47.21	53.68	46.37

OTDA also performed relatively well but had a decrease in performance compared to the PAMAP2 dataset. LMNN had the lowest performance among all the methods, much like in the other two datasets. This suggests that LMNN might not be the best choice for cross-user HAR task.

TROT method gets the highest accuracy among all the three datasets of 18 cross-user HAR tasks, the performance of other methods varies based on the dataset and the specific tasks. The results showed that the methods without considering temporal relation knowledge have difficulties (i.e. unstable performance) when performing time series transfer learning tasks, because the design rationale assumes the samples follow *i.i.d.* in each domain. In contrast, TROT

relaxes the constraint condition of domain adaptation without the *i.i.d.* assumption of samples. Besides, TROT can capture temporal relation knowledge in each domain and reduce the difference by taking into account the temporal relation between the source and target users, which showed superior performance in the time series domain adaptation problem.

5 Conclusion

Current works in domain adaptation are mainly based on the assumption that samples in each domain are *i.i.d.* drawn from a joint probability distribution. However, existing research works overlook the temporal relation among the samples in the process of aligning data distribution and considering the temporal relation knowledge embedded in time series data. In this paper, we propose the TROT method to utilize temporal relation and relax the *i.i.d.* assumption for the source and target users and an OT and HMM based implementation for cross-user HAR. A new regularization term that preserves temporal order information for the regularized optimal transport mapping is proposed to extend the OT capability. Compared to existing domain adaptation, TROT obtains the temporal relation representation and completes the feature-based data distribution alignment to solve domain adaptation problem. Comprehensive experiments applied to three HAR public datasets demonstrate the superiority of the TROT method over other state-of-the-art methods.

In the future, we plan to remove the Markov assumption and extend HAR to capture and align more complex temporal relations in sensor-based HAR cross-user domain adaptation. Moreover, in the pursuit of refining and enhancing the applicability of the TROT, it is imperative to delineate the specific scenarios under which this approach may exhibit suboptimal performance. A critical aspect warranting investigation is the conceptualization and quantification of the “distance” between users or their corresponding data. This “distance” metric could potentially encapsulate the dissimilarity in types, positions, and other attributes of sensors utilized by different users. Establishing a threshold for this “distance” could provide a criterion to ascertain whether the application of TROT would yield satisfactory results. Such an exploration could illuminate the constraints of TROT, and guide the optimization of its algorithm to cater to a broader spectrum of user scenarios. Through a comprehensive analysis of these facets, the endeavor is to elevate the robustness and versatility of TROT, thus propelling it closer to being a universally applicable solution.

References

1. Abdallah, Z.S., Gaber, M.M., Srinivasan, B., Krishnaswamy, S.: Activity recognition with evolving data streams: a review. *ACM Comput. Surv. (CSUR)* **51**(4), 1–36 (2018)
2. Amft, O., Junker, H., Troster, G.: Detection of eating and drinking arm gestures using inertial body-worn sensors. In: *Ninth IEEE International Symposium on Wearable Computers (ISWC 2005)*, pp. 160–163. IEEE (2005)

3. Andreas, B., Blanke, U., Schiele, B.: A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv. (CSUR)* **46**(3), 33 (2014)
4. Barshan, B., Yükses, M.C.: Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* **57**(11), 1649–1667 (2014)
5. Bulling, A., Ward, J.A., Gellersen, H.: Multimodal recognition of reading activity in transit using body-worn sensors. *ACM Trans. Appl. Percept. (TAP)* **9**(1), 1–21 (2012)
6. Chavarriaga, R., et al.: The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. *Pattern Recogn. Lett.* **34**(15), 2033–2042 (2013)
7. Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., Liu, Y.: Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. *ACM Comput. Surv. (CSUR)* **54**(4), 1–40 (2021)
8. Chen, Y., Wang, J., Huang, M., Yu, H.: Cross-position activity recognition with stratified transfer learning. *Pervasive Mob. Comput.* **57**, 1–13 (2019)
9. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: *Advances in Neural Information Processing Systems*, vol. **26** (2013)
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **39**(1), 1–22 (1977)
11. Deng, W.Y., Zheng, Q.H., Wang, Z.M.: Cross-person activity recognition using reduced kernel extreme learning machine. *Neural Netw.* **53**, 1–7 (2014)
12. Ding, R., et al.: Empirical study and improvement on deep transfer learning for human activity recognition. *Sensors* **19**(1), 57 (2018)
13. Esling, P., Agon, C.: Time-series data mining. *ACM Comput. Surv. (CSUR)* **45**(1), 1–34 (2012)
14. Fernando, B., Habrard, A., Sebban, M., Tuytelaars, T.: Unsupervised visual domain adaptation using subspace alignment. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2960–2967 (2013)
15. Flamary, R., Courty, N., Tuia, D., Rakotomamonjy, A.: Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 1–40 (2016)
16. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K.: KNN model-based approach in classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds.) *OTM 2003. LNCS*, vol. 2888, pp. 986–996. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39964-3_62
17. Kerdoncuff, T., Emonet, R., Sebban, M.: Metric learning in optimal transport for domain adaptation. In: *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 2162–2168 (2021)
18. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutorials* **15**(3), 1192–1209 (2012)
19. Lentzas, A., Vrakas, D.: Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review. *Artif. Intell. Rev.* **53**(3), 1975–2021 (2020)
20. Li, S., Xie, B., Wu, J., Zhao, Y., Liu, C.H., Ding, Z.: Simultaneous semantic alignment network for heterogeneous domain adaptation. In: *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 3866–3874 (2020)
21. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G.: Learning under concept drift: a review. *IEEE Trans. Knowl. Data Eng.* **31**(12), 2346–2363 (2018)
22. Lu, W., Chen, Y., Wang, J., Qin, X.: Cross-domain activity recognition via sub-structural optimal transport. *Neurocomputing* **454**, 65–75 (2021)

23. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2009)
24. Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: Visual domain adaptation: a survey of recent advances. *IEEE Signal Process. Mag.* **32**(3), 53–69 (2015)
25. Rabiner, L., Juang, B.: An introduction to hidden markov models. *IEEE ASSP Mag.* **3**(1), 4–16 (1986)
26. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th International Symposium on Wearable Computers, pp. 108–109. IEEE (2012)
27. Rokni, S.A., Ghasemzadeh, H.: Autonomous training of activity recognition algorithms in mobile sensors: a transfer learning approach in context-invariant views. *IEEE Trans. Mob. Comput.* **17**(8), 1764–1777 (2018)
28. Saeedi, R., Sasani, K., Norgaard, S., Gebremedhin, A.H.: Personalized human activity recognition using wearables: a manifold learning-based knowledge transfer. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1193–1196. IEEE (2018)
29. Sekiguchi, R., Abe, K., Yokoyama, T., Kumano, M., Kawakatsu, M.: Ensemble learning for human activity recognition. In: Adjunct proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pp. 335–339 (2020)
30. Soleimani, E., Nazerfard, E.: Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing* **426**, 26–34 (2021)
31. Sukhija, S., Krishnan, N.C.: Supervised heterogeneous feature transfer via random forests. *Artif. Intell.* **268**, 30–53 (2019)
32. Sun, B., Feng, J., Saenko, K.: Return of frustratingly easy domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30 (2016)
33. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds.) ICANN 2018, Part III. LNCS, vol. 11141, pp. 270–279. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01424-7_27
34. Wang, G., Li, Q., Wang, L., Wang, W., Wu, M., Liu, T.: Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors. *Sensors* **18**(6), 1965 (2018)
35. Wang, J., Gao, Q., Pan, M., Fang, Y.: Device-free wireless sensing: challenges, opportunities, and applications. *IEEE Network* **32**(2), 132–137 (2018)
36. Wang, J., Chen, Y., Hu, L., Peng, X., Philip, S.Y.: Stratified transfer learning for cross-domain activity recognition. In: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 1–10. IEEE (2018)
37. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* **10**(2), 207–244 (2009)
38. Wilson, G., Cook, D.J.: A survey of unsupervised deep domain adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)* **11**(5), 1–46 (2020)
39. Xing, T., Sandha, S.S., Balaji, B., Chakraborty, S., Srivastava, M.: Enabling edge devices that learn from each other: cross modal training for activity recognition. In: Proceedings of the 1st International Workshop on Edge Systems, Analytics and Networking, pp. 37–42 (2018)
40. Zhao, Z., Chen, Y., Liu, J., Shen, Z., Liu, M.: Cross-people mobile-phone based activity recognition. In: Twenty-Second International Joint Conference on Artificial Intelligence. Citeseer (2011)