Deep Transfer Learning for Cross-domain Activity Recognition

Jindong Wang*
Institute of Computing Technology, CAS
Beijing, China
wangjindong@ict.ac.cn

Yiqiang Chen†
Institute of Computing Technology, CAS
Beijing, China
yqchen@ict.ac.cn

Vincent W. Zheng
Advanced Digital Sciences Center
Singapore
vincent.zheng@adsc.com.sg

Meiyu Huang
Qian Xuesen Laboratory of Space Technology, CAST
Beijing, China
huangmeiyu@qxslab.cn

ABSTRACT

Human activity recognition plays an important role in people's daily life. However, it is often expensive and time-consuming to acquire sufficient labeled activity data. To solve this problem, transfer learning leverages the labeled samples from the source domain to annotate the target domain which has few or none labels. Unfortunately, when there are several source domains available, it is difficult to select the *right* source domains for transfer. The right source domain means that it has the most similar properties with the target domain, thus their similarity is higher, which can facilitate transfer learning. Choosing the right source domain helps the algorithm perform well and prevents the negative transfer. In this paper, we propose an effective *Unsupervised Source Selection algorithm for Activity Recognition (USSAR)*. USSAR is able to select the most similar K source domains from a list of available domains. After this, we propose an effective *Transfer Neural Network* to perform knowledge transfer for Activity Recognition (TNNAR). TNNAR could capture both the time and spatial relationship between activities while transferring knowledge. Experiments on three public activity recognition datasets demonstrate that: 1) The USSAR algorithm is effective in selecting the best source domains. 2) The TNNAR method can reach high accuracy when performing activity knowledge transfer.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous computing**; • **Computing methodologies** → **Transfer learning**.

KEYWORDS

Transfer Learning, Activity Recognition, Deep Learning, Domain Adaptation

---

*Jindong Wang and Yiqiang Chen are also with Beijing Key Lab of Mobile Computing and Pervasive Devices and University of Chinese Academy of Sciences.
†Corresponding author.

1 INTRODUCTION

Human activity recognition (HAR) aims to seek the profound high-level information from the low-level sensor inputs [39]. For example, we can predict if a person is walking or running using the on-body sensors such as a smartphone or wristband. HAR has been widely applied to many applications such as indoor localization [43], sleep state detection [46], and smart home sensing [41].

There are different activity patterns of different body parts, so sensors can be put on them to collect activity data and then build machine learning models. The combination of signals from different body parts can be used to reflect meaningful knowledge such as person's detailed health information [18] and working states [33]. Unfortunately, the real situation is that we either do not want all body parts to be equipped with sensors, or the data on certain body part may be easily missing. In these situations, we are unable to train models to recognize the activities of some body parts. Figure 1 illustrates this situation. If the activity information on the arm (the red pentacle, we call it the *target domain*) is missing and we want to predict the activities on this body part, how to utilize the information on other parts (such as torso or leg, we call them the *source domains*) to help build the model? This problem is referred to as the *cross-domain activity recognition (CDAR)*.

This problem is extremely challenging. Firstly, we do not know which body parts are most similar to the target domain since the sensor signals are not independent, but are highly correlated because of the shared body structures and functions. If we use all the body parts, there is likely to be negative transfer [32] because some body parts may be dissimilar. Secondly, we only have the raw activity data on the target domain without the actual activity labels, making it harder to measure the similarities. Thirdly, even we already know the similar body parts to the target domain, it is also difficult to build a good machine learning model using both the source and target domains. The reason is that the signal on different domains are following exactly different distributions, which means there are distribution discrepancies between them. However, traditional machine learning models are built by assuming that all signals are with the same distribution. Moreover, existing neural

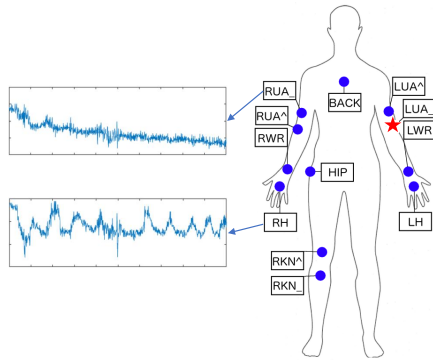Jindong Wang, Vincent W. Zheng, Yiqiang Chen, and Meiyu Huang

Figure 1: An example of cross-domain activity recognition. The activity signals on different body parts are often different. If the labels of a certain part are missing (the red pentacle), how to leverage the well-labeled activity data on other body parts (the blue dots) to acquire its labels?

Figure 2: How to select the most similar body parts and perform activity transfer on multiple persons?

networks are too generic to capture both the time and spatial information in the activity data. Fourthly, when it comes to different persons, there are also similar body parts across persons (Fig. 2). This makes the problem more challenging.

In this paper, we aim to tackle above challenges through multiple source selection and deep neural network. For multiple source selection, we calculate the distance between available sources and the target domains to select the most similar source domains. Our intuition is that the sensor signals may consist of generic and specific relationships about the body parts: the generic relationship means the data distance between two signals such as the Euclidean distance or cosine similarity; the specific relationship refers to the similar moving patterns or body functions between two body parts. By calculating these two significant distances, we can correctly measure the distances between different body parts, and thus select the *right* source domains. Our algorithm does not depend on the availability of the target domain labels. We call this algorithm *Unsupervised Source Selection for Activity Recognition (USSAR)*.

After obtaining the *right* source domains via the USSAR algorithm, we propose a *Transfer Neural Network for Activity Recognition (TNNAR)*. TNNAR is an end-to-end neural network to perform knowledge transfer across different domains. The important thing is that in order to reduce the distance between two domains, we add an *adaptation* layer in the network to calculate the adaptation loss, which can be optimized jointly with the classification loss. The main structure of TNNAR is two convolutional blocks with max-pooling operations, one LSTM layer, and two fully-connected layers. We use the convolutional layers to extract the spatial features from the original activity data. The LSTM layer is mainly for capturing the time features [39]. The fully-connected layers are used for nonlinear transformation. Finally, a softmax function is applied for classification.

To validate the performance of our USSAR and TNNAR algorithms, we conduct extensive experiments on three large-scale public activity recognition datasets: OPPORTUNITY, PAMAP2, and UCI DSADS. All of them contains multiple domains from several body parts. We perform cross-domain activity recognition across
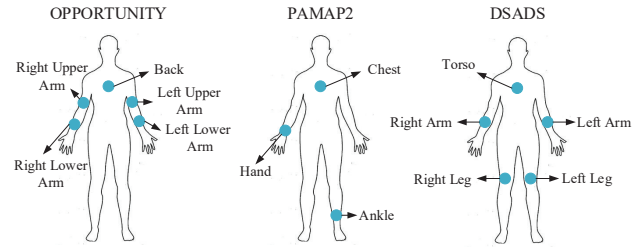
different body parts. The Experimental results demonstrate that US-SAR is effective in choosing the *right* source domains, and TNNAR achieves the best accuracy in activity transfer.

The main contributions of this paper are summarized as follows:

1) We propose the *first* unsupervised source selection algorithm for activity recognition. USSAR measures both the general and specific characteristics of activity information, hence it is capable of capturing the profound relationship between different domains.

2) We propose an end-to-end transfer neural network for cross-domain activity recognition. Different from existing deep transfer learning methods that need to extract features from human knowledge, our TNNAR can simultaneously perform classification and adaptation between two activity domains on the original data.

3) We evaluate the proposed USSAR and TNNAR algorithms in extensive experiments on cross-position activity recognition. Therefore, our algorithms can be applied to many applications to increase the generalization ability of activity models.

2 RELATED WORK

2.1 Activity Recognition

Human Activity recognition has been a popular research topic in pervasive computing [6] for its competence in learning profound high-level knowledge about human activity from raw sensor inputs. Several survey articles have elaborated the recent advance of activity recognition using conventional machine learning [6, 21] and deep learning [39] approaches.

Conventional machine learning approaches have made tremendous progress on HAR by adopting machine learning algorithms such as similarity-based approach [9, 48], active learning [19], crowdsourcing [22], and other semi-supervised methods [20, 29]. Those methods typically treat HAR as a standard time series classification problem. And they tend to solve it by subsequently performing preprocessing procedures, feature extraction, model building, and activity inference. However, they all assume that the training and test data are with the same distribution. As for CDAR where the training (source) and the test (target) data are from different feature distributions, those conventional methods are prune to under-fitting since their generalization ability will be undermined [32].

Deep learning based HAR [33, 39] achieves the state-of-the-art performance than conventional machine learning approaches. The reason is that deep learning is capable of automatically extracting high-level features from the raw sensor readings [23]. Therefore, the features are likely to be more domain-invariant and tend to perform better for cross-domain tasks. A recent work evaluated deep models for cross-domain HAR [28], which provides some

experience for future research on this area. There are still many open problems for deep learning based CDAR.

2.2 Transfer Learning

Transfer learning has been successfully applied in many applications such as Wi-Fi localization [30], natural language processing [4], and visual object recognition [12]. According to the literature survey [32], transfer learning can be categorized into 3 types: instance-based, parameter-based, and feature-based methods.

Instance-based methods perform knowledge transfer mainly through instance re-weighting techniques [7, 36]. Parameter-based methods [44, 47] first train a model using the labeled source domain, then perform clustering on the target domain. Feature-based methods [15, 31, 38] learn a feature transformation between domains when the distance can be minimized. A detailed survey on transfer learning for activity recognition is conducted in [11].

Our work differs from transfer learning in the following two aspects:
1) Source selection

The work [42] first proposed a source-selection free transfer learning approach. They choose the source samples that are close to the target samples using the Laplacian Eigenmap. The work [27] followed this idea in the text classification. However, both of them only focused on the sample selection, while our USSAR focuses on the selection of the whole domain. Collier et al. [10] investigated the transfer performance of different layers of a neural network in a grid search manner. But they did not perform source selection. The work [35] developed a relation network, which can be used to evaluate the distance between different image samples. Yet they still focused on the single sample. Authors in [3] proposed a greedy multi-source selection algorithm. This selection algorithm could iteratively select the best K source domains and then perform transfer learning based on this selection. However, their method is too general to focus on the domain-specific features of activity recognition.

2) Transfer network

In recent years, the deep transfer learning methods have dramatically increased the learning performance of transfer learning tasks. Some popular deep transfer learning methods include Deep Domain Confusion (DDC) [37], Joint Adaptation Network (JAN) [26], and Domain adversarial Neural Network (DANN) [14]. They are typically based on deep neural networks, where they add an adaptation layer to reduce the distribution divergence between domains. Compared to existing deep transfer learning methods, our TNNAR network is tailored according to the characteristics of activity recognition. Hence, we not only consider the generic transfer learning scenario, but also focus on the spatial and time relationship between activities.

3 OUR METHOD

In this section, we describe our Unsupervised Source Selection and Transfer Neural Network for activity recognition. Before that, we will first formally give the problem definition.

3.1 Problem definition

Assume we have an activity domain $\mathcal{D}_t = \{\mathbf{x}_t^j\}_{j=1}^{n_t}$ as the target domain which we want to learn its corresponding activity labels

$\mathbf{y}_t$, i.e. to predict the person's activity state based on the sensor signals. Suppose we have C activity states (labels). There are M labeled source domains available: $\{\mathcal{D}_s^i\}_{i=1}^M$. Each source domain $\mathcal{D}_s^i = \{\mathbf{x}_s^j, y_s^j\}_{j=1}^{n_s^i}$. Note that the data distributions are not the same, i.e. $P(\mathbf{x}_s) \neq P(\mathbf{x}_t)$. We need to design algorithms to: 1) select the best $K(K < M)$ source domains (we denote them as $\mathcal{D}_s(K)$), and 2) perform effective transfer learning from $\mathcal{D}_s(K)$ to $\mathcal{D}_t$ in order to obtain $\mathbf{y}_t$.

3.2 Unsupervised Source Selection

Since the target domain has no labels, it is challenging to measure the distance between $\mathcal{D}_t$ and $\mathcal{D}_s^i$. Moreover, the activity signals are not only normal time series data, there are more mixed information and relationships in different domains, such as the correlated body functions and patterns. Existing distance measurements such as Maximum Mean Discrepancy (MMD) [5] and $\mathcal{A}$ distance [2] are too generic for our problem. They only calculate the data distance without considering the relationship between body parts. Therefore, we should develop a comprehensive measurement to evaluate the distance between different activity domains.

In this paper, we propose the Unsupervised Source Selection for Activity Recognition (USSAR) algorithm to effectively select the *right* source domains for the target activity domain. Our USSAR well considers the *generality* and *specificity* of activity while selecting source domains. Generality means that we have to seek the general relation between activity data, which is a common problem in machine learning. More importantly, specificity means that we should consider the specific information behind different activity domains. Formally, if we denote $D(A, B)$ as the distance between domains A and B, then it can be represented as

$$D(A, B) = D_g(A, B) + \lambda D_s(A, B), \tag{1}$$

where $D_g(A, B)$ is the general distance (*g* for general) and $D_s(A, B)$ is the specific distance (*s* for specific). $\lambda \in [0, 1]$ is the trade-off factor between two terms.

The *general* distance $D_g(A, B)$ can be easily computed by the well-established $\mathcal{A}$-distance [2]:

$$D_g(A, B) = 2(1 - 2\epsilon), \tag{2}$$

where ϵ is the error to classify domains A and B. In order to obtain ϵ, we train a linear binary classifier h on A and B, where A has the label $+1$ and B has the label -1 (or vice versa). Then, we apply prediction on both domains to get the error ϵ.

The *specific* distance $D_s(A, B)$ is composed of two important aspects from activity recognition: the **semantic** and the **kinetic** information. The semantic information refers to the spatial relationship between two domains, i.e. close spatial relations lead to the similar property. The kinetic information refers to the activity signal relationship between domains, i.e. close signal relations lead to the shorter distance. We take the following approaches to calculate the semantic and kinetic distances.

Semantic distance: We basically give each source domain a weight $w \in [0, 1]$ indicating its spatial relationship with the target domain. Therefore, if there are M source domains available, there will be M weights: $\{w_i\}_{i=1}^M$. Currently, the weighting technique

is only based on human experience, i.e. we give relatively small weights if we think two domains are not closely related, while we give relatively large weights if we think two domains are closely related. For instance, if the target domain is the *Right Hand*, we would probably give a larger weight to the source domain *Left Hand* since these two body parts are always correlated; on the other hand, we would probably give the source domain *Torso* a small weight since these two body parts are not exactly correlated. Since w_i is a weight, we bound it by:

$$\sum_{i=1}^{M} w_i = 1. \tag{3}$$

Kinetic distance: There are many approaches available to approximate the signal relationship between two domains. This relationship can be captured by the consistency between two signals. For instance, the Pearson correlation score and the maximum mean discrepancy can be used to calculate this relationship. In this paper, we adopt the well-established *Cosine* similarity function. Specifically, given two domains A and B, their cosine similarity is formulated as:

$$cos(A, B) = \mathbb{E}\left[\sum_{\mathbf{a},\mathbf{b}} \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a} \cdot \mathbf{b}|}\right], \tag{4}$$

where $\mathbf{a}, \mathbf{b}$ are the basic vectors in A and B, respectively. $\mathbb{E}[\cdot]$ is the expectation of samples. Note that when one domain is the target domain, its weight can be set to 1. Once the kinetic distance is calculated, we combine it with the weights generated by the semantic distance, and finally get the specificity distance as:

$$d_s(A, B) = \mathbb{E}\left[\sum_{\mathbf{a},\mathbf{b}} \frac{w_{\mathbf{a}}\mathbf{a} \cdot w_{\mathbf{b}}\mathbf{b}}{|w_{\mathbf{a}}\mathbf{a} \cdot w_{\mathbf{b}}\mathbf{b}|}\right]. \tag{5}$$

Overall distance: we combine the general distance (Eq. 2) and the specific distance (Eq. 5) to get the final distance expression:

$$d(A, B) = 2(1 - 2\epsilon) + \lambda\mathbb{E}\left[\sum_{\mathbf{a},\mathbf{b}} \frac{w_{\mathbf{a}}\mathbf{a} \cdot w_{\mathbf{b}}\mathbf{b}}{|w_{\mathbf{a}}\mathbf{a} \cdot w_{\mathbf{b}}\mathbf{b}|}\right]. \tag{6}$$

Note that this distance measurement does not rely on the labels of the target domain. We call this distance the *Context Activity Distance (CAD)*. Once we have the CAD, we can perform source selection from many available source domains. In this paper, we propose a greedy algorithm to select the available sources. We regard the source selection problem as the process of constructing a finite set S. Initially, we calculate all the CADs between the target domain and every source domain and sort them in an increasing order. At this time, the set $S = \phi$. Then, we add the source domain with the smallest CAD (denoted as $\mathcal{D}_s^{\min}$)to S: $S = \mathcal{D}_s^{\min}$. After that, we add other $K - 1$ source domains to S according to the greedy technique: if $d(S, \mathcal{D}_s^i) < d(\mathcal{D}_t, \mathcal{D}_s^i)$, we add $\mathcal{D}_s^i$ to S. The USSAR algorithm is illustrated in Algorithm 1.

3.3 Transfer Neural Network

After obtaining the selected source domains $\mathcal{D}_s(K)$, we can perform knowledge transfer across the target and the source domains.

Algorithm 1 USSAR: Unsupervised Source Selection for Activity Recognition

---

Input: M available source domains $\{\mathcal{D}_s^i\}_{i=1}^M$, target domain $\mathcal{D}_t$, the number of selected source domains $K(K << M)$
Output: The selected source domain set S.
 1: Calculate the CAD between each source domain and target domain using Eq. (6), and sort them in an increasing order;
 2: Initialize a set $S = \mathcal{D}_s^{\min}$;
   Set $i = 2$
 3: **repeat**
 4:   Calculate the CAD between $\mathcal{D}_s^i$ and S, and denote it as d_{iS};
 5:   If $d_{iS} < d(\mathcal{D}_t, \mathcal{D}_s^i)$, we add $\mathcal{D}_s^i$ to S;
 6:   Else $i = i + 1$;
 7: **until** i = K
 8: **return** S.

In this paper, we propose a neural network to accomplish this transfer. Generally speaking, the goal of the transfer neural network is to learn the classifiers $y = f_s(\mathbf{x})$ and $y = f_t(\mathbf{x})$ and minimize their discrepancies. So the expected target risk is bounded: $R_t(f_t) = \mathbb{E}_{(\mathbf{x},y)}[f_t(\mathbf{x}) \neq y]$. Since there are discrepancies between the source and target domain, the general form of the function f can be expressed as:

$$f(\mathbf{x}) = \ell_c(\mathbf{x}) + \mu\ell_a(\mathcal{D}_s, \mathcal{D}_t), \tag{7}$$

where $\ell_c(\mathbf{x})$ is the classification loss on the labeled data (source domain), and $\ell_a(\cdot, \cdot)$ is the adaptation loss on both of the source and target domains. $\mu \in [0, 1]$ is the trade-off factor.

This is a general form of a transfer neural network. In cross-domain activity recognition setting, the structure of the neural network has to be modified according to the characteristics of activity recognition. In this paper, we propose Transfer Neural Network for Activity Recognition (TNNAR). The structure of our TNNAR is illustrated in Figure 3. The proposed TNNAR consists of two convolutional layers with max-pooling layers, one LSTM layer, and two fully-connected (fc) layers. The convolutional layers are adopted to extract the spatial features for the activities, while the LSTM layer is adopted to capture the time relationship between activities. The fully-connected layer is acting as the classification function.

Beyond this simple structure, we also add an *adaptation* layer to reduce the discrepancy between two domains. The reason we add this layer after the first fully-connected layer is that the features are becoming specific to the higher layer, making adaptation more urgent. The low-level layers are only extracting some common features, thus they do not need to be adapted [45]. The adaptation layer computes the adaptation loss, which can be optimized jointly with the classification loss.

We denote $\mathbf{W}^l$ and $\mathbf{b}^l$ the weights and bias at fc layer l. Each fc layer l will learn a nonlinear mapping $\mathbf{h}^l = f^l(\mathbf{W}^l\mathbf{h}_i^{l-1} + \mathbf{b}^l)$, where $\mathbf{h}^l$ is the lth layer hidden representation and f^l is the activation function. The activation f^l for the last fc layer is computed as $f^l(\mathbf{x}) = e^{\mathbf{x}}/\sum_{j=1}^{|\mathbf{x}|} e^{\mathbf{x}_j}$. The other fc layer takes the ReLU units $f^l = \max(\mathbf{0}, \mathbf{x})$. If we denote $\Theta = \{\mathbf{W}^l, \mathbf{b}^l\}$ the hyperparameters of the neural network, then the empirical risk is
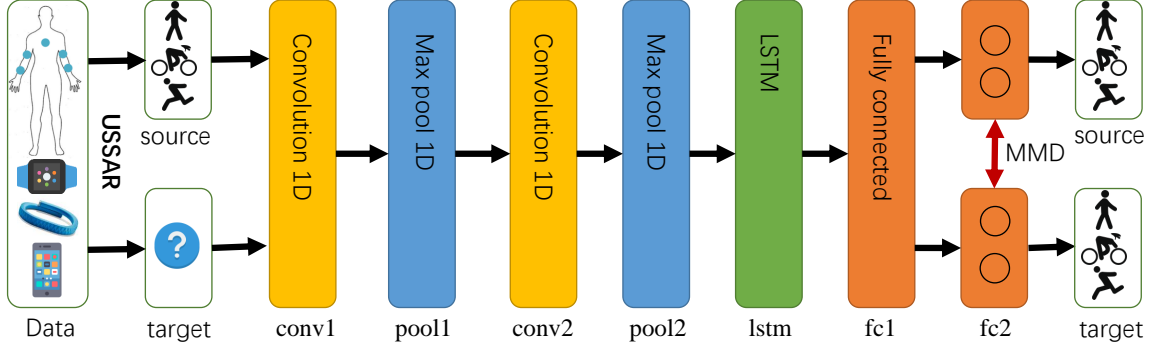
Figure 3: The structure of Transfer Neural Network for Activity Recognition (TNNAR).

$$\ell_c = \min_{\Theta} \frac{1}{n_b} \sum_{i=1}^{b} J(\Theta(\mathbf{x}_i^b), y_i^b), \qquad (8)$$

where J is the cross-entropy function. $(\mathbf{x}_i^b, y_i^b)$ denotes all the labeled samples from the source domain.

As for the adaptation layer, we adopted the well-established Maximum Mean Discrepancy (MMD) [5] as the measurement to reduce the discrepancy between domains. MMD is a popular distance metric, which has been widely used in many existing work [24, 31], and its effectiveness has been verified in [16]. The MMD distance between distributions p and q is defined as $d^2(p, q) = (\mathbb{E}_p[\phi(\mathbf{z}_s)] - \mathbb{E}_q[\phi(\mathbf{z}_t)])^2_{\mathcal{H}_K}$ where $\mathcal{H}_K$ is the reproducing kernel Hilbert space (RKHS) induced by feature map $\phi(\cdot)$. Here, $\mathbb{E}[\cdot]$ denotes the mean of the embedded samples. Therefore, the MMD distance between the source and target domain is

$$MMD(\mathcal{D}_s, \mathcal{D}_t) = \|\mathbb{E}[\mathbf{x}_s] - \mathbb{E}[\mathbf{x}_t]\|^2_{\mathcal{H}_K}. \qquad (9)$$

We train the TNNAR using mini-batch Stochastic Gradient Descent (SGD) strategy. The gradient can be calculated as

$$\nabla_{\Theta^l} = \frac{\partial J(\cdot)}{\partial \Theta^l} + \mu \frac{\partial \ell_a(\cdot)}{\partial \Theta^l}. \qquad (10)$$

4 EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of USSAR and TNNAR.

4.1 Datasets and Setup

We used the same datasets and setup in a recent literature [40] to perform cross-position activity recognition (CPAR). CPAR is an important aspect of cross-domain activity recognition (CDAR). Specifically, it refers to the situation where the activity labels of some body parts are missing, so it is necessary and feasible to leverage the labeled data from other similar body parts to get the labels of those body parts.

There are three public datasets used in [40]: OPPORTUNITY dataset (**OPP**) [8], PAMAP2 dataset (**PAMAP2**) [34], and UCI daily and sports dataset (**DSADS**) [1]. Table 1 provides a brief introduction to these three datasets. In the following, we briefly introduce those datasets, and more information can be found in their original papers. **OPP** is composed of 4 subjects executing different levels of activities with sensors tied to more than 5 body parts. **PAMAP2** is

collected by 9 subjects performing 18 activities with sensors on 3 body parts. **DSADS** consists of 19 activities collected from 8 subjects wearing body-worn sensors on 5 body parts. Accelerometer, gyroscope, and magnetometer are all used in three datasets.

Following the same protocol in [40], we also investigated the same positions in three datasets as in Figure 2. In our experiments, we use the data from all three sensors in each body part since most information can be retained in this way.

The transfer scenarios are obtained according to [40]. There are three scenarios that reflect different similarities between domains: a) *similar body parts of the same person*, b) *different body parts of the same person*, c) *similar body parts of different person*. In the sequel, we use the notation $A \rightarrow B$ to denote labeling the activity of domain B using the labeled domain A. In total, we constructed 22 tasks. Note that there are different activities in three datasets. For scenario a) and b), we simply use all the classes in each dataset; for scenario c) which is cross-dataset, we extract 4 common classes for each dataset (i.e. *Walking, Sitting, Lying,* and *Standing*). In addition, we did not include the scenario '*different body parts of different person*' since 1) all the methods perform poorly in that scenario, and 2) that scenario does not have reasonable feasibility in real applications.

4.2 Evaluation of USSAR

First, we evaluate the performance of the proposed USSAR algorithm for unsupervised source selection. In this experiment, we choose the OPP and DSADS datasets since there are 5 body positions in them. We combine the two datasets together, which means that given a target domain in a dataset, there are 9 source domains available to select. In order to fully evaluate the algorithm, we set $K = 1, 2, 3, 4$.

We compare USSAR with several source selection techniques:

- $\mathcal{A}$-distance [2], which is selected according to the top K smallest $\mathcal{A}$-distances. This distance is acting as the baseline.
- SDM: subspace disagreement measurement [15], which is computing the distance between domains based on the principle angle [17].
- GR: Greedy algorithm [3], which is selecting sources using a greedy technique.
- MMD: Maximum Mean Discrepancy [5], which is a popular metric to measure the distance.

We randomly select two body positions: *Torso (T, or Back in DSADS)* and *Right Arm (RA) in DSADS*. When one of the body parts

Table 1: Brief introduction of three public datasets for activity recognition [40]

Dataset	#Subject	#Activity	#Sample	#Feature	Body parts
OPPORTUNITY	4	4	701,366	459	Back (B), Right Upper Arm (RUA), Right Left Arm (RLA), Left Upper Arm (LLA), Left Lower Arm (LLA)
PAMAP2	9	18	2,844,868	243	Hand (H), Chest(C), Ankle (A)
DSADS	8	19	1,140,000	405	Torso (T), Right Arm (RA), Left Arm (LA), Right Leg (RL), Left Leg (LL)

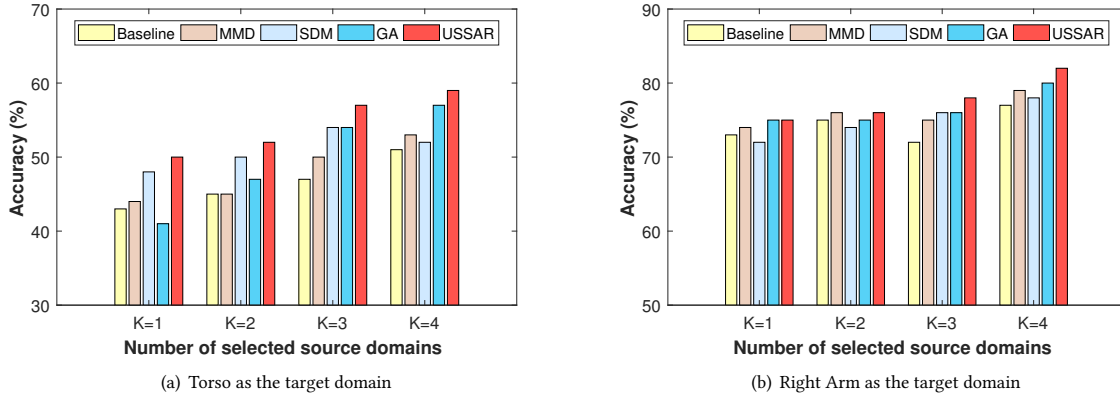(a) Torso as the target domain   (b) Right Arm as the target domain

Figure 4: Classification accuracy of different source selection algorithms.

is selected as the target domain, it means that their labels are missing and we need to predict its labels using the rest 9 domains. The parameters of comparison methods are set according to their original papers. For USSAR algorithm, we assign different weights to each body part according to their relationship to the target domain. In fact, each body part can be a target domain, so these weights are dynamic and relative. For instance, in OPP dataset, if we take the RUA as the target domain, then the weights for the other nine parts can be: $0.1, 0.2, 0.3, 0.1, \cdots, 0.1$, as long as they add up to 1. This weighting technique can be tailored according to human experience. As long as it reflects the functional relationship between body parts, that can be accepted.

Note that there is no ground-truth about the *right* source domains: we can never learn the actual distance between two domains. Therefore, we turn to use the classification accuracy as the evaluation metric. After selecting the source domains, we train the same linear SVM classifier on the selected source domains, and then apply prediction on the target domain to get the classification accuracy. It is intuitive that if we select better source domains, the classification accuracy will also be good. The experimental results are in Figure 4. The results clearly indicated that the proposed USSAR source selection algorithm can choose the right source domains for transfer learning.

4.3 Evaluation of TNNAR

In this part, we evaluate the effectiveness of TNNAR for transfer learning in CPAR. In order to test the performance of TNNAR, we conducted two experiments: transfer learning on single source domain, and transfer learning on multiple source domains. We compare TNNAR with the following methods:

- PCA: Principal component analysis [13].
- KPCA: Kernel principal component analysis [13].
- TCA: Transfer component analysis [31].
- GFK: Geodesic flow kernel [15].
- TKL: Transfer kernel learning [25].
- STL: Stratified Transfer Learning [40].

PCA and KPCA are classic dimensionality reduction methods, while TCA, GFK, TKL, and STL are representative transfer learning approaches. The codes of PCA and KPCA are provided in Matlab. The codes of TCA, GFK, and TKL can be obtained online [1]. The constructed datasets and STL code are available online [2].

The implementations of all comparison methods are following [40]. Different from these work which exploited feature extraction according to human knowledge, we take the original signal as the input. For TNNAR network, we set the learning rate to be 0.005. The batch sizes for source and target domains are selected from [128, 256, 512]. Note that although we selected K source domains, we basically combine them into one large source domain. Since the sensor signal is a multi-channel reading, we treat each channel as a distinct signal and perform 1D convolution on it. Totally, there are 9 channels (i.e. 3 accelerometers, 3 gyroscopes, and 3 magnetometers). The convolution kernel size is 1×1 with the depth 32. Other parameters of the neural network are set accordingly. For the MMD measurement, we take the linear-time MMD as in [16].

Note that in both of the two experiments, all of the comparison methods used the same source and target domains. For the single source domain situation, we follow the settings in [40] and report the results in Table 2. For the experiments on multiple source domains, we extend the results in the last section and set $K = 3$ to select the source domains by USSAR. The results are in Table 3. Note that in order to obtain the steady performance, we perform 10 random permutations of the data and record the average results.

The results clearly indicated that TNNAR dramatically increases the performance of cross-domain activity recognition. Specifically, TNNAR has an average performance improvement of **3.42%** compared to the best method STL. In all levels of similarities, TNNAR outperforms the comparison methods. It indicates that TNNAR is capable of performing transfer learning in all kinds of activity recognition scenarios. The reasons are three folds: 1) Other comparison methods are operated on the extracted features according

---

[1]https://tinyurl.com/y79j6twy
[2]https://tinyurl.com/y7en6owt

Table 2: Classification accuracy (%) of TNNAR and other comparison methods on single source transfer tasks.

Scenario	Dataset	Task	PCA	KPCA	TCA	GFK	TKL	STL	TNNAR
Similar body parts on the same person	DSADS	RA → LA	59.91	62.17	66.15	71.07	54.10	71.04	**75.89**
		RL → LL	69.46	70.92	75.06	79.71	61.63	81.60	**86.76**
	OPP	RUA → LUA	76.12	65.64	76.88	74.62	66.81	83.96	**87.43**
		RLA → LLA	62.17	66.48	60.60	74.62	66.82	83.93	**86.29**
Different body parts on the same person	DSADS	RA → T	38.89	30.20	39.41	44.19	32.72	45.61	**50.22**
	PAMAP2	H → C	34.97	24.44	34.86	36.24	35.67	43.47	**46.32**
	OPP	RLA → T	59.10	46.99	55.43	48.89	47.66	56.88	**59.58**
		RUA → T	67.95	54.52	67.50	66.14	60.49	75.15	**75.75**
Similar body parts on different person	PAMAP2 → OPP	C → B	32.80	43.78	39.02	27.64	35.64	40.10	**45.62**
	DSADS → PAMAP	T → C	23.19	17.95	23.66	19.39	21.65	37.83	**39.21**
	OPP → DSADS	B → T	44.30	49.35	46.91	48.07	52.79	55.45	**57.97**
Average			51.71	48.40	53.23	53.69	48.73	61.37	**64.64**

Table 3: Accuracy (%) of multiple source domains

Target	PCA	TCA	GFK	TKL	STL	TNNAR
RA	66.78	68.43	70.87	70.21	73.22	**78.40**
Torso	42.87	47.21	48.09	43.32	51.22	**55.48**
RL	71.24	73.47	81.23	74.26	83.76	**87.41**
RLA	65.78	67.10	76.38	70.32	84.52	**86.75**
Average	61.67	64.05	69.14	64.53	73.18	**77.01**

(a) λ

(b) μ

Figure 5: Sensitivity analysis of λ and μ.

to human knowledge, which may not be sufficient to capture the resourceful information of the activities. TNNAR is based on the deep neural network to automatically extract features without human knowledge. As previous work has demonstrated the effectiveness of deep neural network on feature extraction [39], it will help the network to extract more high-level features. 2) The structure of the neural network is beneficial for performing transfer learning, since the hyperparameters can be easily shared across domains. 3) The deep neural network has both the convolution and LSTM cells, which enables it to learn the spatial and time information from the activities. Therefore, the network can understand more information about the activity data.

We can also have more insights by combining Table 2 and Table 3. Firstly, for the same target domain, adding more source domains will clearly increase the performance. This is because there is more knowledge contained in multiple domains than a single domain. Secondly, when *Torso* is the target domain, the results are dramatically increased. When the arms and legs are as the target domains, the results did not improve that far. This is probably because that the most similar part to the arms and legs are their opposites (the other arms and legs). Thus adding more source domains can only increase the results a little. However, since the Torso is highly correlated with all the body parts, the performance of this body part will be dramatically increased by adding other source domains.

4.4 Sensitivity Analysis

There are two critical factors in USSAR and TNNAR: the trade-off factors λ and μ. A good choice for these two factors will help the algorithm perform better. In this section, we evaluate the sensitivity of them through empirical experiments. We set $\lambda, \mu \in \{0, 0.1, \cdots, 1.0\}$ for a single task (Torso as the target domain, and $K = 3$) and record the classification accuracy in Figure 5. From the results, we can clearly see that the algorithm is robust with a large width of the

parameter choice. Therefore, these two parameters do not to be cherry-picked. This indicates that our algorithm can be easily applied to real applications.

5 CONCLUSIONS AND FUTURE WORK

If the activity labels of some body parts are missing, it is critical and necessary to exploit the well-labeled information from other body parts to obtain the missing labels. In this paper, we propose the first unsupervised source selection algorithm for activity recognition (USSAR). USSAR could consider both the semantic and kinetic relation between body parts, thus it is able to accurately select the right domains that are closely related to the target domain. We also propose an end-to-end Transfer Neural Network for activity recognition (TNNAR) that can learn transferable representations for activities. Experimental results demonstrate that compared to many source selection and transfer learning algorithms, our proposed USSAR can select the right source domains and TNNAR is able to achieve the best classification accuracy.

In future work, we plan to extend the USSAR and TNNAR algorithms for activity recognition with heterogeneous and distant activity domains.

ACKNOWLEDGMENTS

REFERENCES

[1] Billur Barshan and Murat Cihan Yüksek. 2014. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* 57, 11 (2014), 1649–1667.

[2] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. 2007. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*. 137–144.

[3] Himanshu S Bhatt, Arun Rajkumar, and Shourya Roy. 2016. Multi-Source Iterative Adaptation for Cross-Domain Classification.. In *IJCAI*. 3691–3697.

[4] John Blitzer, Ryan McDonald, and Fernando Pereira. 2006. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 120–128.

[5] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola. 2006. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics* 22, 14 (2006), e49–e57.

[6] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 33.

[7] Rita Chattopadhyay, Qian Sun, Wei Fan, Ian Davidson, Sethuraman Panchanathan, and Jieping Ye. 2012. Multisource domain adaptation and its application to early detection of fatigue. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6, 4 (2012), 18.

[8] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. 2013. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters* 34, 15 (2013), 2033–2042.

[9] Yiqiang Chen, Yang Gu, Xinlong Jiang, and Jindong Wang. 2016. OCEAN: A New Opportunistic Computing Model for Wearable Activity Recognition. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 33–36.

[10] Edward Collier, Robert DiBiano, and Supratik Mukhopadhyay. 2018. Cactus-Nets: Layer Applicability as a Metric for Transfer Learning. *arXiv preprint arXiv:1804.07846* (2018).

[11] Diane Cook, Kyle D Feuz, and Narayanan C Krishnan. 2013. Transfer learning for activity recognition: A survey. *Knowledge and information systems* 36, 3 (2013), 537–556.

[12] Lixin Duan, Ivor W Tsang, and Dong Xu. 2012. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 3 (2012), 465–479.

[13] Imola K Fodor. 2002. A survey of dimension reduction techniques. *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory* 9 (2002), 1–18.

[14] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35.

[15] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. 2012. Geodesic flow kernel for unsupervised domain adaptation. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2066–2073.

[16] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. 2012. A kernel two-sample test. *Journal of Machine Learning Research* 13, Mar (2012), 723–773.

[17] Jihun Hamm and Daniel D Lee. 2008. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning*. ACM, 376–383.

[18] Nils Yannick Hammerla, James Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plötz. 2015. PD Disease State Assessment in Naturalistic Environments Using Deep Learning.. In *AAAI*. 1742–1748.

[19] HM Sajjad Hossain, Nirmalya Roy, and Md Abdullah Al Hafiz Khan. 2016. Active learning enabled activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–9.

[20] Lisha Hu, Yiqiang Chen, Shuangquan Wang, Jindong Wang, Jianfei Shen, Xinlong Jiang, and Zhiqi Shen. 2016. Less Annotation on Personalized Activity Recognition Using Context Data. In *Proceedings of the 2016 International IEEE Conference on Ubiquitous Intelligence Computing (UIC)*. 327–332.

[21] Oscar D Lara and Miguel A Labrador. 2013. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys & Tutorials* 15, 3 (2013), 1192–1209.

[22] Walter S Lasecki, Young Chol Song, Henry Kautz, and Jeffrey P Bigham. 2013. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1203–1212.

[23] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.

[24] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. 2013. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2200–2207.

[25] Mingsheng Long, Jianmin Wang, Jiaguang Sun, and S Yu Philip. 2015. Domain invariant transfer kernel learning. *IEEE Transactions on Knowledge and Data Engineering* 27, 6 (2015), 1519–1532.

[26] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. 2017. Deep Transfer Learning with Joint Adaptation Networks. In *International Conference on Machine Learning*. 2208–2217.

[27] Zhongqi Lu, Yin Zhu, Sinno Jialin Pan, Evan Wei Xiang, Yujing Wang, and Qiang Yang. 2014. Source Free Transfer Learning for Text Classification.. In *AAAI*. 122–128.

[28] Francisco Javier Ordóñez Morales and Daniel Roggen. 2016. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. ACM, 92–99.

[29] Le T Nguyen, Ming Zeng, Patrick Tague, and Joy Zhang. 2015. I did not smoke 100 cigarettes today!: avoiding false positives in real-world activity recognition. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1053–1063.

[30] Sinno Jialin Pan, James T Kwok, and Qiang Yang. 2008. Transfer Learning via Dimensionality Reduction. In *Proceedings of the 23rd AAAI conference on Artificial intelligence*, Vol. 8. 677–682.

[31] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. 2011. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks* 22, 2 (2011), 199–210.

[32] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on* 22, 10 (2010), 1345–1359.

[33] Thomas Plötz, Nils Y Hammerla, and Patrick Olivier. 2011. Feature learning for activity recognition in ubiquitous computing. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22. 1729.

[34] Attila Reiss and Didier Stricker. 2012. Introducing a new benchmarked dataset for activity monitoring. In *Wearable Computers (ISWC), 2012 16th International Symposium on*. IEEE, 108–109.

[35] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. 2017. Learning to Compare: Relation Network for Few-Shot Learning. *arXiv preprint arXiv:1711.06025* (2017).

[36] Ben Tan, Yu Zhang, Sinno Jialin Pan, and Qiang Yang. 2017. Distant Domain Transfer Learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[37] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014).

[38] Jindong Wang, Yiqiang Chen, Shuji Hao, Wenjie Feng, and Zhiqi Shen. 2017. Balanced Distribution Adaptation for Transfer Learning. In *The IEEE International conference on data mining (ICDM)*. 1129–1134.

[39] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. 2018. Deep Learning for Sensor-based Activity Recognition: A Survey. *Pattern Recognition Letters* (2018).

[40] Jindong Wang, Yiqiang Chen, Lisha Hu, Xiaohui Peng, and Philip S Yu. 2018. Stratified Transfer Learning for Cross-domain Activity Recognition. In *IEEE international conference on pervasive computing and communications (PerCom)*.

[41] Jiahui Wen, Jadwiga Indulska, and Mingyang Zhong. 2016. Adaptive activity learning with dynamically available context. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–11.

[42] Evan Wei Xiang, Sinno Jialin Pan, Weike Pan, Jian Su, and Qiang Yang. 2011. Source-selection-free transfer learning. In *IJCAI proceedings-international joint conference on artificial intelligence*, Vol. 22. 2355.

[43] Han Xu, Zheng Yang, Zimu Zhou, Longfei Shangguan, Ke Yi, and Yunhao Liu. 2016. Indoor localization via multi-modal sensing on smartphones. In *UbiComp*. ACM, 208–219.

[44] Yi Yao and Gianfranco Doretto. 2010. Boosting for transfer learning with multiple sources. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*. IEEE, 1855–1862.

[45] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.

[46] Ming Zhao, Shichao Yue, Dina Katabi, and Tommi Jaakkola. 2017. Learning sleep stages from radio signals: A deep adversarial architecture. In *ICML*.

[47] Zhongtang Zhao, Yiqiang Chen, Junfa Liu, Zhiqi Shen, and Mingjie Liu. 2011. Cross-people mobile-phone based activity recognition. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence (IJCAI)*, Vol. 11. Citeseer, 2545–2550.

[48] Vincent W Zheng and Qiang Yang. 2011. User-dependent aspect model for collaborative activity recognition. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence (IJCAI)*, Vol. 22. 2085–2090.