ORIGINAL ARTICLE

A holistic multi-source transfer learning approach using wearable sensors for personalized daily activity recognition

Qi Jia[2,3] · Jing Guo[1,2] · Po Yang[1,2,4] · Yun Yang[1,2]

Abstract

Human activity recognition (HAR) aims to collect time series through wearable devices to precisely identify specific actions. However, the traditional HAR method ignores the activity variances among individuals, which will cause low generalization when applied to a new individual and indirectly enhance the difficulties of personalized HAR service. In this paper, we fully consider activity divergence among individuals to develop an end-to-end model, the multi-source unsupervised co-transfer network (MUCT), to provide personalized activity recognition for new individuals. We denote the collected data of different individuals as multiple domains and implement deep domain adaptation to align each pair of source and target domains. In addition, we propose a consistent filter that utilizes two heterogeneous classifiers to automatically select high-confidence instances from the target domain to jointly enhance the performance on the target task. The effectiveness and performance of our model are evaluated through comprehensive experiments on two activity recognition benchmarks and a private activity recognition data set (collected by our signal sensors), where our model outperforms traditional transfer learning methods at HAR.

Keywords Human activity recognition · Personalized service · Transfer learning · Domain adaptation

Introduction

Human activity recognition (HAR) is a fundamental technology in medical services and healthcare that can extract information from time series data collected by wearable sig-

Qi Jia and Jing Guo have contributed equally to this work.

✉ Yun Yang
   yangyan19@hotmail.com

   Qi Jia
   jiaqi9669@hotmail.com

   Jing Guo
   guojing@mail.ynu.edu.cn

   Po Yang
   Poyangcn@gmail.com

1   School of Software, Yunnan University, Kunming 650500, Yunnan, China

2   Yunnan Key Laboratory of Software Engineering, Yunnan University, Kunming 650500, Yunnan, China

3   School of Information Science and Engineering, Yunnan University, Kunming 650500, Yunnan, China

4   Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK
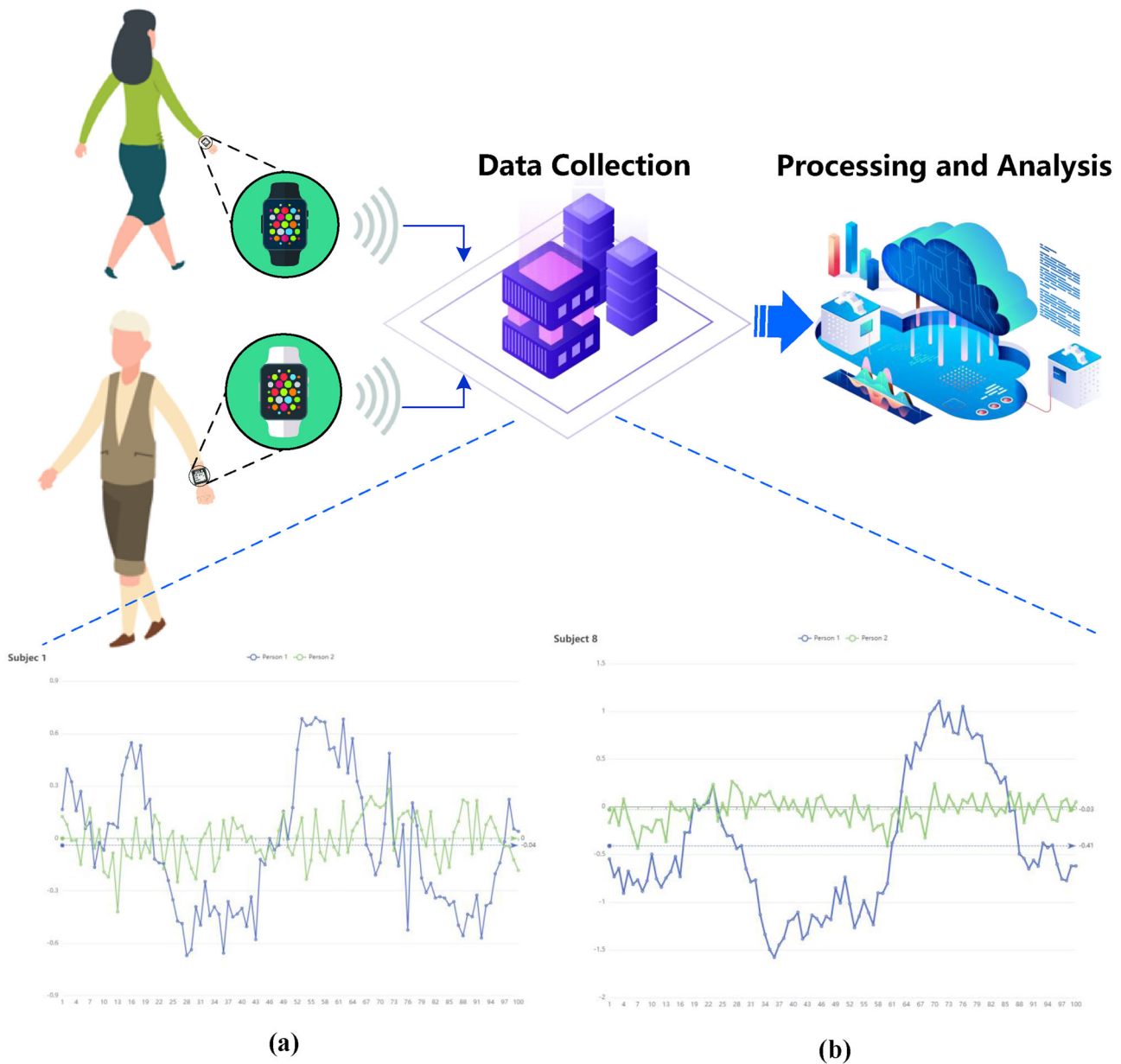
nal sensors to predict the activities of human beings [1]. Due to its real-time characteristic and portability, HAR is widely used in rehabilitation monitoring [2], geriatric monitoring [3], and other fields [4]. With the development of technology and the maturity of intelligent algorithms, traditional HAR has been far from satisfying human needs, but more personalized services, which can accurately recognize the activities for new individuals, have become the focus of current research.

In recent years, deep learning based HAR algorithm has got evolutionary progress [5, 6]. Specifically, it can overcome the dependency on manual feature extraction by powerful representation capability to obtain the correlation between time series data and automatically extracting features for better classification. Traditional deep learning relies on large amount of high-quality labeled training data to obtain a robust model [7, 8]. However, the deviation of activities among individuals will lead to data distribution discrepancy, makes the deep learning models hard to perform well on new individuals, as shown in Fig. 1. Therefore, how to effectively eliminate the discrepancy among different individuals becomes the key to enhance the generalization ability of personalized HAR services.

Fig. 1 Overview of HAR processing. Data and mean of two individuals collected by *z*-axis acceleration sensor on private data set: **a** left foot, performing subject 1; **b** right foot, performing subject 8

Transfer learning breaks the assumption of machine learning that the distribution of training and testing data must be the same [9]. It can improve the generalization ability of the model by reducing the distribution discrepancies cross domains [10, 11], which has been widely applied in computer vision [7, 12], medical decision-making [13–16], natural language processing (NLP) [17–19], and other fields [20–23]. By utilizing massive, similar and high-quality activity time series data captured by other individuals, transfer learning can effectively improve the generalization abilities of the model on new individuals, thus realizing personalized HAR service. According to the terminology of transfer learning,

we annotate the labeled activity data as the source domain, and the unlabeled activity data collected from new individual as the target domain [24]. In addition, considering the labeled data often comes from multiple independent individuals in practical application, we conduct multi-source transfer learning [25, 26] to separately eliminate the differences of activity data among individuals to realize more precisely HAR on target task.

On the other hand, due to the complexity of data distribution in real scenario, the effectiveness of the transfer learning model is constrained. Ensemble learning can improve the classification performance by constructing and combining

multiple weak classifiers, which has been proved to be an efficient way for multi-source transfer learning [27, 28]. In general, we extend the advantages of ensemble learning for two main purpose: (1) adopt a co-training strategy to select high-confident target samples for training and (2) combine the transfer results from different source domains to reduce sensitivity to distribution variances and achieve a better average performance. Our previous studies have systematically illustrated the advantages of ensemble learning on transfer learning [29, 30].

In this paper, we fully consider the variance in the distribution of different individuals and propose an end-to-end model, multi-source unsupervised co-transfer network (MUCT), to establish a personalized and precise classification model for HAR. We use feature extractors to automatically extract the time series data feature and use domain adaptation to align the distribution variance cross domains. Furthermore, a consistency filter is developed, consisting of two heterogeneous classifiers to assign the pseudo-label for the target domain with a final agreement. Based on the co-training strategy with a consistency filter, high-confidence target samples are selected to reintegrate the training set and iteratively boost the effectiveness of the classifiers. We aggregate the classification results from the multi-source domain to acquire the final result. The main contributions of this model can be summarized as follows:

- We propose an unsupervised multi-source transfer learning network that provides a feasible end-to-end way to realize personalized HAR service;
- An adaptive feature extractor is presented to automatically extract the feature of time series data and align the distribution variance among domains to improve the transferability of the model;
- MUCT can iteratively enhance the robustness of the model by training with high-confidence target samples collected by a consistency filter;
- Experimental results on benchmark data sets and a real-world data set (collected by our signal sensors) demonstrate the superior performance of our model over traditional unsupervised domain adaptation methods.

The remainder of this paper is structured as follows. "Related work" section presents related work. "Methods" section describes the workflow of MUCT. "Experiment" section reports simulation results on benchmark data sets and a real-world data set. In "Discussion" section, we analyze the results and discuss future work. In the last section, we describe the conclusions of our model.

Related work

We discuss work on human activity recognition via traditional machine learning methods and multi-source transfer learning.

Human activity recognition

In recent years, the spotlight has been on HAR [31], which can learn advanced information about human activities from raw sensor input [32]. Some recent surveys introduce traditional machine learning models on HAR [33–37]. Bayat et al. [38] proposed a digital low-pass filter that could extract the component of gravity acceleration from body acceleration to improve classifier performance. Hossain et al. [39] combined K-means with active learning to reduce the reliance on training data labels in HAR. Some recent methods have bridged deep learning and HAR, utilizing deep learning methods to extract features from raw time series data. Zeng et al. [40] proposed deep learning based on convolutional neural networks (CNNs) and a partial weight-sharing technique, which could capture the local dependency and scale invariance of a signal. A partial weight-sharing technique could apply to sensor signals to further improve data. Lee et al. [41] proposed a one-dimensional CNN-based method using smartphone tri-axial accelerometer data to transform x, y, and z acceleration data to vector magnitudes.

Many traditional HAR methods require manual design for the features, which is time and labor consuming. Although deep learning methods such as CNNs have gradually become mainstream methods in HAR, they fail to account for individual variance, which will cause the poor generalization of classifiers for different individuals.

Multi-source domain adaptation

Domain adaptation is a significant branch of transfer learning that aligns a labeled source and unlabeled target domains in a specific feature space. In real-world applications, multi-source data can be drawn. Hence multi-source domain adaptation is an appropriate method to gain knowledge from multiple perspectives to enhance the classification ability of the target domain [42]. Zhu et al. [43] used Maximum Mean Discrepancy (MMD) loss to align domain-specific distributions, and Disagreement Classifier (DISC) loss to align domain-specific classifiers, enabling classifiers to achieve satisfactory results in the target domain. Fang et al. [27] presented a method that mapped target domain samples into

multi-label samples, which could comprehensively consider the correlation between tags. The model creates a shared label subspace in multi-source domains, applying the obtained knowledge to the target domain classifier.

Recently, domain adaptation methods have been applied in HAR. Wang et al. [44] measured the distance of activity data distributions among multiple humans to find the best domain to accomplish transfer tasks, utilizing CNN and long short-term memory (LSTM) layers to extract time series and spatial features, and MMD loss to align them as similarly as possible. Zhao et al. [45] proposed a transfer learning-embedded decision tree that integrates decision trees with the K-means algorithm in personalized HAR. Methods based on domain adaptation cannot deal with multi-source domains, and their generalization ability is poor in practical applications without a suitable ensemble paradigm, especially on unbalanced data sets.

We propose a MUCT model that leverages knowledge gained from multiple individuals with annotated information (source domain) to precisely recognize activities for a specific person (target domain). By adopting a boost strategy, the generalization ability of the algorithm is improved on both benchmark and real-world data sets.

Methods

We show how our model addresses individual invariance and enhances generalization when applying HAR. Our model consists of multi-source feature alignment and a consistency filter, as shown in Fig. 2. Multi-source feature alignment extracts discriminative and domain-invariant features for source and target domains. Heterogeneous classifiers are included in the consistency filter to provide predictive labels for samples from different views. The consistency filter can select high-confidence samples with pseudo-labels to enrich the diversity of training samples and promote the learning process of target information. The pseudocode is shown in Algorithm 1.

Problem statement

We propose a personalized HAR service solution. We denote K items of source domain data with labels as $\mathcal{D}_{s_k} = \left\{\left(x_i^{s_k}, y_i^{s_k}\right)\right\}_{i=1}^{n_{s_k}}$, where n_{s_k} is the number of instances of the kth source domain. $\mathcal{D}_t = \left\{\left(x_i^t\right)\right\}_{i=1}^{n_t}$ is the target domain, with n_t unlabeled instances. $\mathcal{D}_{s_k}$ and $\mathcal{D}_t$ are sampled from data distributions p and q, respectively, $p \neq q$. MUCT trains a network $\mathbf{y} = H(\mathbf{x})$ that can effectively reduce the distribution variance between the source and target domains, and the target risk $R_t(H) = \mathbb{E}_{(\mathbf{x},\mathbf{y})q}[H(\mathbf{x}) \neq \mathbf{y}]$ can be bounded by leveraging the source domain.

Algorithm 1 MUCT

Input: Data: K source domains $\mathcal{D}_{s_k} = \left\{(x_i^{s_k}, y_i^{s_k})\right\}_{i=1}^{n_{s_k}}$ and target domain $\mathcal{D}_t = \left\{x_j^t\right\}_{j=1}^{n_t}$. Networks: MUCT model H is consists of feature extractor $\{f_k(\cdot)\}_{k=1}^K$ and heterogeneous classifiers $\left\{C_{k_1}, C_{k_2}\right\}_{k=1}^K$ for each pair of source and target domain. Parameters: hyperparameter β.

Output: The optimal MUCT H.

1: $\mathcal{D}_t^* = \phi$.
2: **for** n to N **do**
3: 　　$f_{s_k} \leftarrow f_k(x_i^{s_k})$; $f_{t_k} \leftarrow f_k(x_j^t)$.
4: 　　Calculate $\mathcal{L}_{mmd}$ in Eq. (3) with each pair of f_{s_k} and f_{t_k}.
5: 　　**if** $n \bmod \beta == 0$ **then**
6: 　　　　Update $\mathcal{D}_t^*$ in Eq. (4).
7: 　　**end if**
8: 　　Calculate $\mathcal{L}_{cls}$ in Eq.(5) with $\mathcal{D}_t^* \cup \mathcal{D}_{s_k}$
9: 　　Compute the $\mathcal{L}_{muct}$ in Eq.(6)
10: 　　Update parameters of H by minimizing $\mathcal{L}_{muct}$
11: **end for**
12: **return** H

Multi-source feature alignment

In personalized HAR, the distributional variance between domains reduces the generalizability of the model over the target domain. Specifically, the feature extractor cannot extract domain-invariant features due to domain distribution variance. The classifier is unable to classify the target domain data effectively. Multi-source feature alignment aligns the domain distribution variance and motivates the feature extractor to extract domain-invariant features.

To extract domain-invariant features, we introduce MMD to measure the variance between domains. MMD measures the sample feature distributions in the Reproducing Kernel Hilbert Space (RKHS) and calculates the mean embedding distances from different domains [46]. The MMD distance can be estimated as

$$d_{mmd}(p, q) \triangleq \left\| \mathbf{E}_{x^s \sim p}\left[\phi\left(x^s\right)\right] - \mathbf{E}_{x^t \sim q}\left[\phi\left(x^t\right)\right] \right\|_{\mathcal{H}}^2, \quad (1)$$

where $\mathcal{H}$ is the RKHS endowed with a characteristic kernel, $kernel\left(X_s, X_t\right) = <\phi\left(X_s\right), \phi\left(X_t\right)>$, where $<\cdot,\cdot>$ is the inner product of vectors, $\phi(\cdot)$ denotes the mapping of a feature distribution to RKHS, and p and q are the distributions of x^s and x^t, respectively. In general, the empirical estimate of MMD $d_{mmd}(p, q)$ can be further factorized as

$$\begin{aligned}
\hat{d}_{mmd}(p, q) &= \left\| \frac{1}{n_s}\sum_{x_i \in \mathcal{D}_s}\phi\left(x_i\right) - \frac{1}{n_t}\sum_{x_j \in \mathcal{D}_t}\phi\left(x_j\right) \right\|_{\mathcal{H}}^2 \\
&= \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{j=1}^{n_s}k\left(x_i^s, x_j^s\right) + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}k\left(x_i^t, x_j^t\right) \\
&\quad - \frac{2}{n_s n_t}\sum_{i=1}^{n_s}\sum_{j=1}^{n_t}k\left(x_i^s, x_j^t\right).
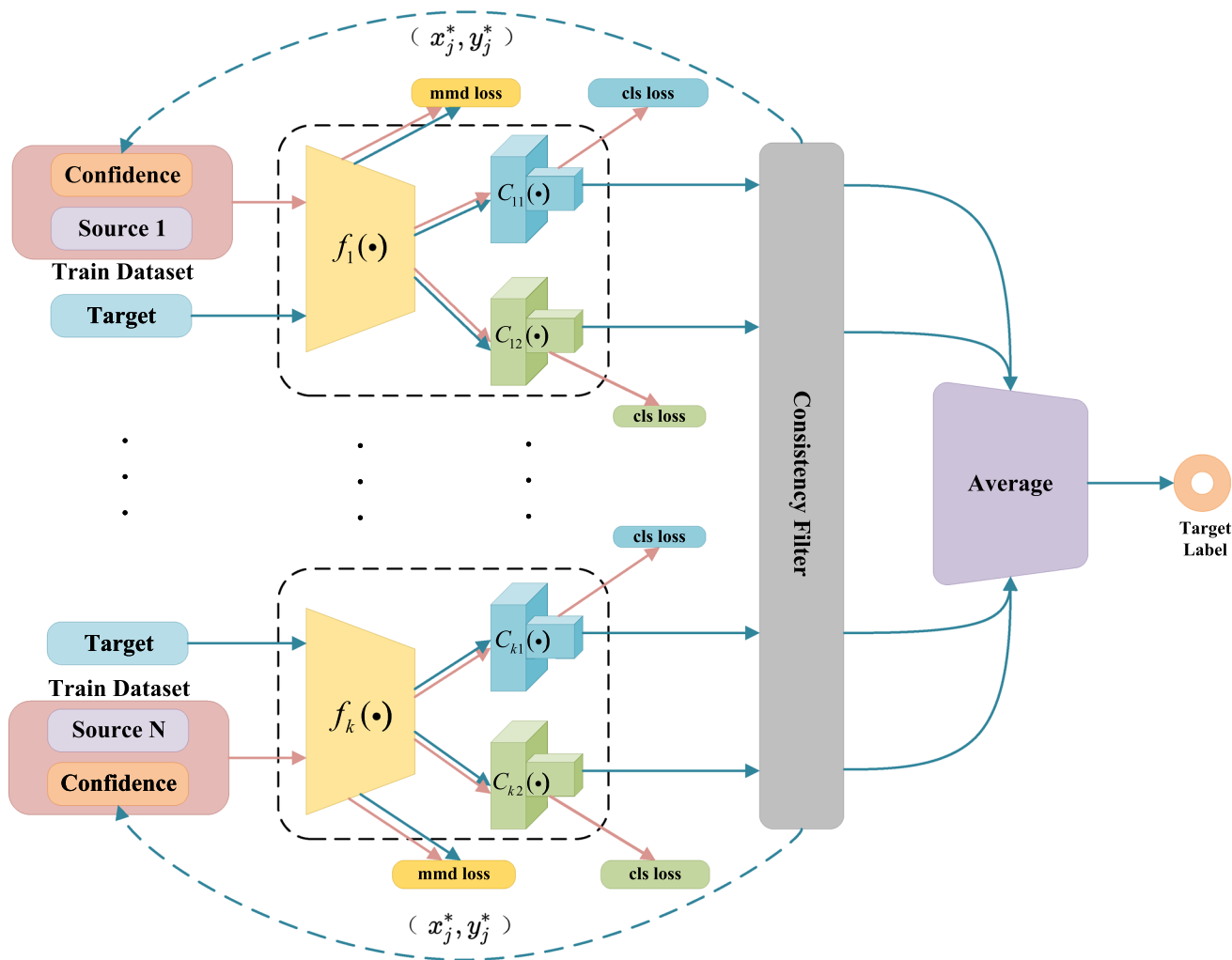\end{aligned} \quad (2)$$

Fig. 2 Overview of MUCT model. Feature extractor obtains a specific common feature space in each pair of source and target domains and can extract their domain-invariant feature representations to alleviate the problem of variance distribution among different domains.

High-confidence data in the source domain are filtered by classification results of multiple classifiers and added with their pseudo-labels to the training set. Average probability is calculated for results of each classifier to obtain final classification result

In multi-source transfer learning, it is challenging to construct a suitable common latent feature space for all domains because the data distributions in each domain differ. When the number of domains is too large, the common latent feature space will reduce the discriminability of features to adapt all common domain features [43]. To enhance the representation capability of the feature extractor, we propose K subnetworks $\{f_k(\cdot)\}_{k=1}^{K}$ as feature extractors to construct a common latent feature space for each pair of source and target domains. The source domain feature extraction process can be expressed as $f_{s_k} = f_k\left(x_i^{s_k}\right)$. Similarly, we use each feature extractor to perform feature extraction on the target domain data and obtain the features of the target domain. The features of the target domain extracted using the kth feature extractor can be expressed as $f_{t_k} = f_k\left(x_j^t\right)$. We use MMD to measure the

variance in the distribution of the source and target domains. The MMD loss is

$$\mathcal{L}_{mmd} = \sum_{k=1}^{K} \hat{d}_{mmd}(p, q)(f_k(x^{s_k}), f_k(x^t)). \qquad (3)$$

Consistency filter

We propose a consistency filter to enrich the diversity of training samples and promote the learning process of the target information by assigning pseudo-labels to target domain samples and collecting the prediction results of all classifiers for the target domain samples. When all classifiers predict the same result for the target domain sample, the consistency filter will make the prediction a pseudo-label.

This sample is eventually joined to the training set of each classifier. With increasing iterations, the number of target domain samples in the training set grows. Therefore, the feature extractor will focus on extracting discriminable features in the target domain, which are unique in it and are not domain-invariant. In other words, the consistency filter helps in the extraction of potential features of the target domain, ultimately improving the classifier's performance. We use multi-classifiers in the final recognition task. For each source domain, two heterogeneous classifiers, $C_{k1}(\cdot)$ and $C_{k2}(\cdot)$, perform classification tasks from different perspectives, receiving the domain-invariant feature after feature extractor $f_k(\cdot)$ for the kth source human domain. We compute the instance's filter label, S_i^i, which when 0 indicates that an instance is highly trusted data, which helps improve the classifier's performance; otherwise, it is untrusted data. S_i calculation formula as follows:

$$S_i = \sum_{k=1}^{K} EQ(|C_{k_1}(f_k(x_j^t))|, |C_{k_2}(f_k(x_j^t))|). \tag{4}$$

$EQ(\cdot, \cdot)$ is used to judge whether the predictions of the two classifiers are consistent, outputting 0 when they are, and 1 otherwise. After selection, we add the confidence data and pseudo-label $(x_j^t, C_{1_1}(f_1(x_j^t)))$ to the confidence data set $\mathcal{D}_t^* = \{x_l^*, y_l^*\}_{l=1}^{n_*}$, and add this to the classifier's training set, where n_* is the number of samples in the confidence data set. To prevent the classifier from non convergence, we run the consistency filter after each iteration of hyperparameter β.

In HAR, a single classifier cannot accurately classify the data from multiple views because of the spatial and time series characteristics of the data, such as frequency and sequence. These heterogeneous classifiers not only are trained from different domain data but can combine the advantages of different classification algorithms. Therefore, these filtered target domain data have a high confidence level. Therefore, we use two heterogeneous classifiers to classify the samples from different perspectives to obtain predictive labels with high execution for each source domain. We use cross-entropy to optimize the heterogeneous classifiers. Cross-entropy loss measures the degree of difference between two probability distributions in the same random variable. We estimate the difference between the probability distribution and its prediction. The smaller the cross-entropy, the better the classifier predicts. We take the cross-entropy loss as the classification loss of each classifier:

$$\mathcal{L}_{cls} = \sum_{k=1}^{K} \sum_{z=1}^{2} \mathcal{L}_{ce}(C_{k_z}\left(f_k\left(x^{s_k}\right)\right), y^{s_k})$$
$$+ \sum_{k=1}^{K} \sum_{z=1}^{2} \mathcal{L}_{ce}(C_{k_z}\left(f_k\left(x^*\right)\right), y^*), \tag{5}$$

where $\mathcal{L}_{ce}$ is cross-entropy loss, K is the number of source domains, and C_{k_z} is the zth classifier in the kth domain.

Multi-source unsupervised co-transfer learning network

Design features are difficult for human activity recognition. In addition, the connections between multiple features between activities are challenging to capture. To this end, we propose MUCT. Our model comprises K feature extractors, $2K$ classifiers, and a consistency filter. The total loss of our model generally consists of domain and classification loss, where domain loss is measured by MMD, and classification loss is calculated by cross-entropy. The feature extractor extracts domain-invariant features and measures their differences to represent the domain loss. The classification loss improves the classifier's performance, and therefore network performance. We use a consistency filter to obtain the confidence data of the target domain and add pseudo-labels to the training set. Because the target domain data are added to the training set, the classifier performs better on the target domain. The total loss of our model is

$$\mathcal{L}_{muct} = \mathcal{L}_{cls} + \mathcal{L}_{mmd}. \tag{6}$$

In MUCT, the feature is extracted by the feature extractor, and the feature distribution of each pair of source and target domains is as similar as possible to MMD. After this stage, we can approximately consider the source and target domain data distributions to be similar. This provides the basis for our second stage. MUCT improves the classifier's performance on the target through a consistency filter. We determine its final label by averaging the weights. We use each classifier to predict the probability of the class to which the target human sample belongs by summing the possibilities and averaging them. The category with the highest probability is its label. This process can be expressed as follows:

$$H(x_i^t) = \frac{1}{2K} \sum_{k=1}^{K} \sum_{z=1}^{Z} C_{kz}(f_k(x_i^t)). \tag{7}$$

Experiment

We conducted a number of evaluations and experiments to compare the effectiveness and performance of our methods with other state-of-the-art (SOTA) domain adaptation methods on two benchmark HAR data sets and a real-world data set.

Data sets and setup

We compared the effectiveness and performance of MUCT with ResNet [47], JDA [48], TCA [49], DAN [50], DAAN [51], and MFSAN [43], which include traditional deep learning, single-source domain adaptation, and multi-source domain adaptation methods.

We performed three standard comparisons: (1) Source combine unites different source domains into one, despite the distribution variance; (2) Single best reports the optimal results in multiple single-source domain adaptations among multi-source domain, to compare the differences between the upper bound of single-source and multi-source domain adaptation; (3) Multi-source compares the classification results of multi-source domain adaptation methods. To further validate the effectiveness of every component in our model, we carried out ablation experiments and evaluated several MUCT variants: (1) $MUCT_{SV}$, with a single classifier; (2) $MUCT_{NC}$, without a consistency filter; (3) $MUCT$, the whole pipeline, combining multiple classifiers and a consistency filter. Considering calculation and time constraints, we followed the experimental setup in FedHealth [52], randomly selecting four subjects in each data set to evaluate the effectiveness and performance of our model.

Experimental data sets included two HAR benchmark data sets, the UCI daily and sports data set (DSADS) [53] and the WISDM Lab human activity recognition data set (WISDM) [54], and a real-world data set collected by our signal sensors (real data set). Table 1 describes these data sets.

The DSADS data set was collected from eight subjects who wore sensors in five positions while completing 19 movements. Each subject performed all movements for five minutes. Four males and four females were subjects, and the sensors collected data at 25 Hz. The five minutes of data collected from each subject were divided into 60 segments. Sensors were worn on the subject's torso (T), right arm (RA), left arm (LA), right leg (RL), and left leg (LL). We randomly selected four individuals from the DSADS data set and noted them as P_1–P_4. We built four transfer tasks using them as target domains: $P_1, P_2, P_3 \rightarrow P_4$; $P_1, P_2, P_4 \rightarrow P_3$; $P_1, P_3, P_4 \rightarrow P_2$, and $P_2, P_3, P_4 \rightarrow P_1$.

WISDM included six activities intelligently collected from 36 subjects. We randomly selected four individuals from this data set and labeled them as P_1–P_4. For WISDM, cell phone accelerometers collected six types of human activity data: Walking, Jogging, Upstairs, Downstairs, Sitting, and Standing. We built four transfer tasks, P_1–P_4, as target domains: $P_1, P_2, P_3 \rightarrow P_4$; $P_1, P_2, P_4 \rightarrow P_3$; $P_1, P_3, P_4 \rightarrow P_2$ and $P_2, P_3, P_4 \rightarrow P_1$.

The real data set included 13 activities intelligently collected from four subjects. We collected 19 types of activity information from four volunteers using wearable sensors in Kunming, Yunnan Province, China. We labeled the volunteers as P_1–P_4. Sensor units were on the waist (W), right arm (RA), left arm (LA), right leg (RL), and left leg (LL). These units collected activity data 20 times per second, and this was grouped in 5-second increments after completing the collection. We used ResNet-18 to extract features.

We compared our MUCT model with the five SOTA algorithms for human activity recognition problems. Our model was based on the PyTorch framework. We set the initial learning rate at 0.2, and it decreased with iterations. We used the SGD optimizer and set its momentum to 0.9. We used ResNet-18 to extract features. Softmax and DNN were used as classifiers. Among ResNet, DAN, DAAN, and MFSAN, we used ResNet18 as the backbone network. We used a learning rate of 0.2 and SGD in these models as the optimizer.

Effectiveness of MUCT

Performance on DSADS. We compared our model with others on DSADS and recorded the best results, obtaining an average over five trials. Table 2 shows that the algorithm's experimental effect on Source Combine was better than on Single Best. This situation does not square with our previous experience. The reasons for the experimental results are that the activity movements between humans are small in DSADS, which leads to slight variance in the generated data distribution, so the model's performance is not degraded because of the difference in data distribution in Source Combine. On the contrary, it can increase the generalization performance of the model. It can be seen from the experiment that in this case, better classification results can be obtained even without the use of transfer learning. For these reasons, the performance of the multi-source model MFSAN did not exceed that of Source Combine.

Although there is little difference in data distribution, there is still the issue of complex data characteristics in time and space. However, our algorithm performs better because multiple heterogeneous classifiers and consistent filters can address these aspects. Due to the small difference in data distributions, our consistency filter can screen out more high-confidence samples in the target domain to participate in the classifier's training. Our multiple heterogeneous classifiers can capture more features of the data for classification. In addition, the consistency filter adds the target domain data to the training set so that the classifiers can adapt to the target distribution. The accuracy of our ablation experiment on
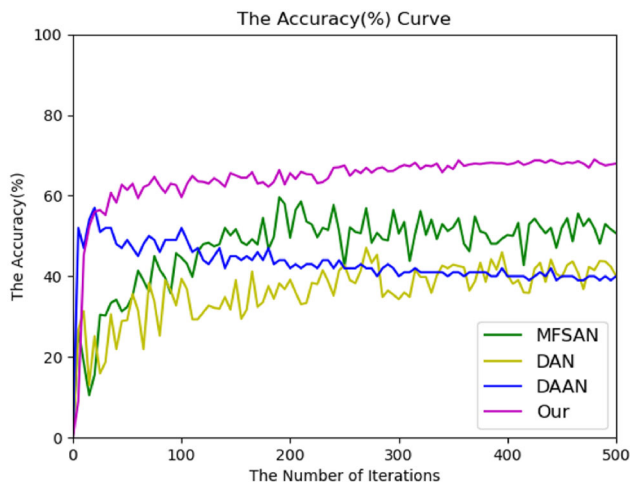
Table 1 Benchmark and real HAR data sets

Data set	Subject	Activity	Sample	Body part
DSADS	8	19	1,140,000	5
WISDM	36	6	1,098,207	1
Real data set	4	13	330,800	5

Fig. 3 ACC curve on real data set

Fig. 4 Box plots on real data set

$MUCT_{NC}$ and $MUCT_{SV}$ also decreased after removing the consistency filter, which confirmed our conjecture. We can see that our model achieved optimal results in four groups of experiments on DSADS.

Performance on WISDM. We compared the results of our model to others on the WISDM data set. WISDM was collected on smartphones with a single sensor, making little information available. Table 3 compares classification accuracy on WISDM. We can clearly see that the performance on Single Best was better than on source combine. DAAN had a better accuracy rate in classification tasks due to the few category labels in the WISDM data set. DAAN could be more accurate based on label subdomains, thus improving the ability to extract common features with data labels. The multi-source model MFSAN had better results than the source combine models because MFSAN can capture information from multi-source domains.

As shown in Table 3, our model can improve the performance of the HAR task more effectively with multiple classifiers and a consistency filter. In WISDM, MUCT achieves the best results in terms of average accuracy.

Performance on real data set. We evaluated our model's performance on a real application data set, with results as shown in Table 4, from which we can see that the model performance with single best was better than with source combine. This is consistent with our previous experience. Multi-source model MFSAN was better than single best and source combine because multi-source models can obtain more source domain information and improve classifier performance by adapting to the data distribution of the target domain. Our model achieved the best performance. To further verify our model's performance, we implemented task $P_2, P_3, P_4 \rightarrow P_1$ on the real data set and plotted the ACC curve, as shown in Fig. 3, which shows the ACC curves of all
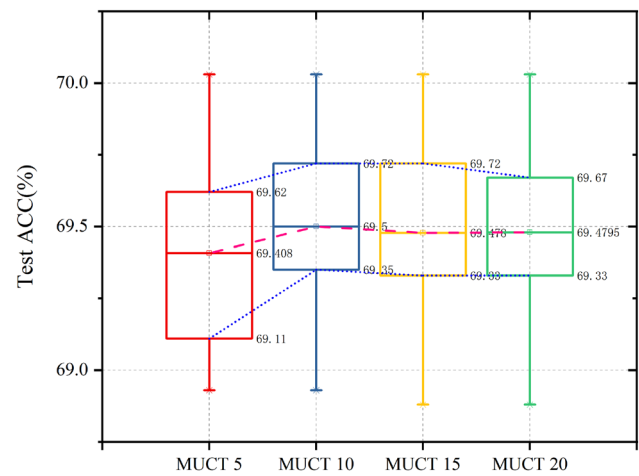
four models, where the horizontal axis represents the number of training iterations, and the vertical axis represents the accuracy rate of the model. From the ACC curve, we can observe the model's rate of convergence and stability. Although the curve of DAAN is sharper, the accuracy of our model is higher, and the algorithm more stable, after convergence. To verify the reliability of our model, we repeated the experiment 5, 10, 15, and 20 times and drew box plots, as shown in Fig. 4. We can find that the results of the four groups of experiments are similar of the mean, maximum, and minimum values. This shows that our model has reliable performance in HAR service.

Evaluation of MUCT on unbalanced data set

We tested the performance of our model on unbalanced data sets through experiments with the MUCT model on the real data set, whose data and label distribution are shown in Table 5. We can see that the subject 10 data accounted for 20.73%. However, the data collected from subject 8 only accounted for 4.69%. This data imbalance exists not only between subjects but between domains. For example, in subject 8, there are only 15 samples of P_1, while there are 33, 61, and 46 samples of P_2, P_3, and P_4, respectively. This imbalance is a challenge for the HAR task.

To confirm that our model handles unbalanced data, we calculated the confusion matrix on the real data set with source domains P_2, P_3, P_4, and target domain P_1. Figure 5 shows the confusion matrices of DAN, DAAN, MFSAN, and our model, where each row represents the true label of a category of samples, and each column represents the predicted label of the model for predicting a category. The diagonal of the confusion matrix is the result of correct classification. We can observe the performance of the models on the unbalanced data set by the confusion matrix. The MSFAN approach is

Table 2 Comparison of classification accuracy (%) on DSADS

Standard	Method	$P_1, P_2, P_3 \rightarrow P_4$	$P_1, P_2, P_4 \rightarrow P_3$	$P_1, P_3, P_4 \rightarrow P_2$	$P_2, P_3, P_4 \rightarrow P_1$	Avg
Single best	ResNet	66.32	63.60	63.95	58.95	63.21
	JDA	44.39	46.49	43.42	49.12	45.86
	TCA	46.97	46.70	40.83	52.48	46.75
	DAN	54.74	61.23	63.77	65.96	61.43
	DAAN	70.79	69.91	73.42	68.25	70.59
Source combine	ResNet	75.18	76.40	80.44	81.67	78.42
	JDA	71.93	62.19	46.14	50.04	57.58
	TCA	52.84	58.07	43.12	56.14	52.54
	DAN	61.23	70.35	69.56	62.72	65.97
	DAAN	79.65	80.09	78.77	58.25	74.19
Multi-source	MFSAN	69.30	71.05	68.42	65.88	68.66
	$MUCT_{SV}$	83.95	88.86	80.26	**86.84**	84.98
	$MUCT_{NC}$	80.53	86.49	79.74	82.98	82.44
	$MUCT$	**85.00**	**93.25**	**82.46**	84.39	**86.28**

Bold indicates the best result

Table 3 Comparison of classification accuracy (%) on WISDM

Standard	Method	$P_1, P_2, P_3 \rightarrow P_4$	$P_1, P_2, P_4 \rightarrow P_3$	$P_1, P_3, P_4 \rightarrow P_2$	$P_2, P_3, P_4 \rightarrow P_1$	Avg
Single best	ResNet	78.21	89.92	53.63	81.10	75.72
	JDA	74.51	78.29	63.10	**88.58**	76.12
	TCA	66.16	89.87	52.02	87.55	73.90
	DAN	61.67	**93.02**	71.77	85.63	78.02
	DAAN	77.04	**93.02**	72.18	81.10	80.84
Source combine	ResNet	85.21	71.32	59.48	76.57	73.15
	JDA	68.09	78.29	44.76	81.30	68.11
	TCA	67.03	81.01	43.27	66.81	64.53
	DAN	71.40	83.72	76.21	84.84	79.04
	DAAN	85.02	89.92	65.12	82.48	80.64
Multi-source	MFSAN	60.31	73.64	76.61	77.95	72.13
	$MUCT_{SV}$	88.33	88.37	73.19	85.43	83.83
	$MUCT_{NC}$	86.38	88.37	65.73	84.84	81.33
	$MUCT$	**88.91**	91.47	**76.81**	86.02	**85.80**

Bold indicates the best result

superior to DAN and DAAN with unbalanced data sets. DAN is more inclined to classify data as label 1 in the classification task, and DAAN is more prone to labels 5 and 12. Our model weights the data, which enables it to perform better on unbalanced data sets. Our model has fewer misclassifications and higher accuracy on unbalanced data sets.

To further analyze the performance of our model on unbalanced data sets, we experimented on the real data set with source domains P_2, P_3, P_4 and target domain P_1, and drew the ROC curve. We compared our model with MFSAN, DAN, and DAAN. Figure 6 shows the Receiver Operating

Characteristic (ROC) curve, which can accurately reflect the relationship between the true example rate and a certain learner's false-positive rate. It can be found from the figure that the Area Under Curve (AUC) region of MFSAN is significantly larger than that of DAN and DAAN. Due to richer samples, the multi-source domain models perform better on unbalanced data sets. We can find that the AUC region of our model is more significant than those of other models, which confirms its stronger robustness on unbalanced data sets.
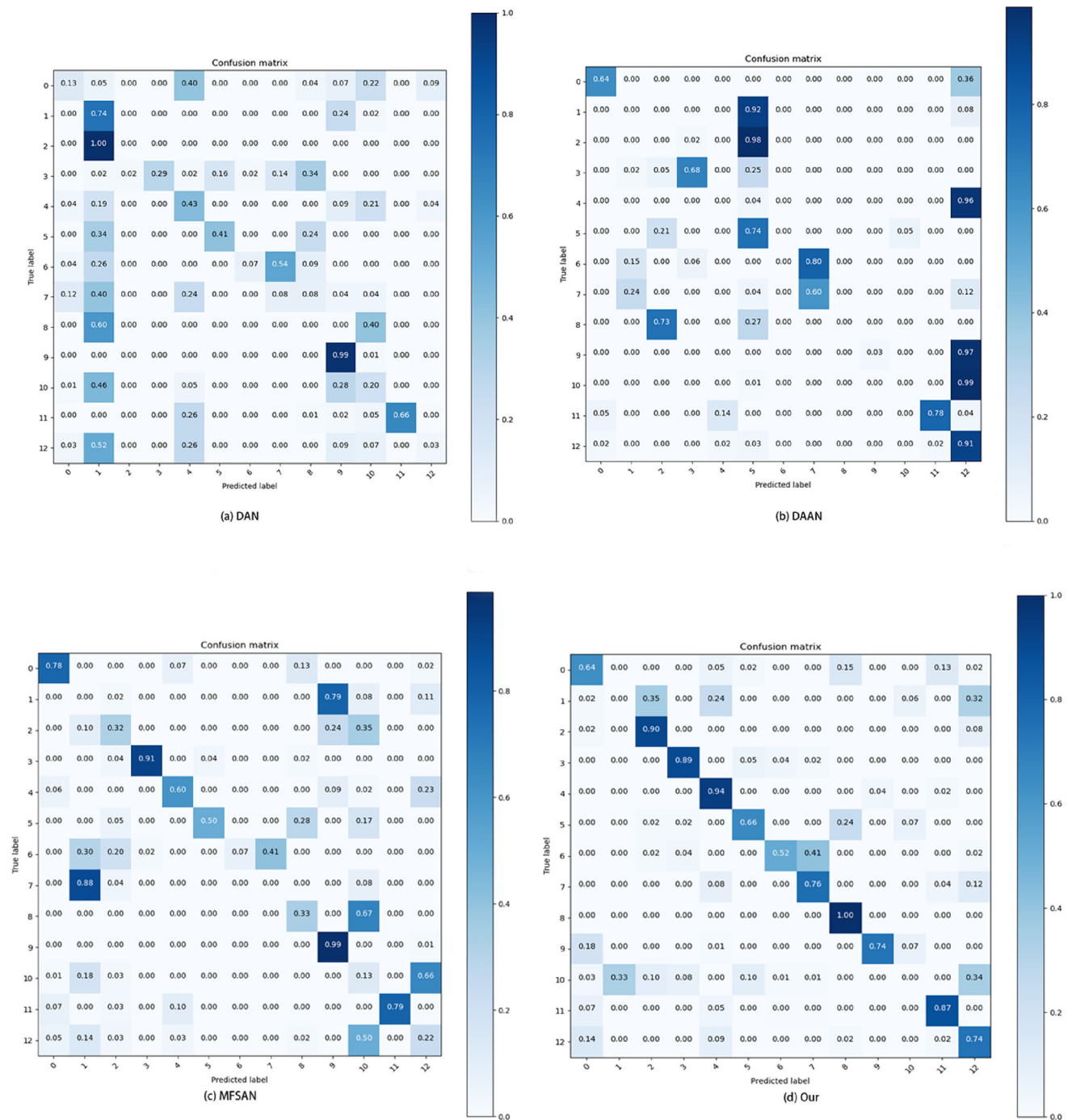
Fig. 5 Confusion matrix on real data set

Discussion

Our experiments illustrated that our model can achieve impressive classification results for personalized HAR without target data labels. Our model acquired high-confidence data through a Consistency Filter to assist in classifier training, and it could still maintain a high AUC on unbalanced data sets. Hence, it provides a promising, easy-to-use tech-

nique to address personalized HAR problems. We summarize the method's innovations and drawbacks and propose future work.

Our model can work in a personalized human environment without new human data labels in human activity recognition. This multi-source unsupervised transfer learning method requires less target data than the traditional deep learning method. The information of multiple source

Table 4 Comparison of classification accuracy (%) on real data set

Standards	Method	$P_1, P_2, P_3 \rightarrow P_4$	$P_1, P_2, P_4 \rightarrow P_3$	$P_1, P_3, P_4 \rightarrow P_2$	$P_2, P_3, P_4 \rightarrow P_1$	Avg
Single best	ResNet	53.27	69.38	65.47	69.97	64.52
	JDA	64.45	82.85	73.32	76.55	74.29
	TCA	58.61	**83.31**	71.46	72.85	71.56
	DAN	53.39	71.54	72.19	76.42	68.39
	DAAN	58.23	72.31	74.46	73.83	69.71
Source combine	ResNet	60.84	73.32	78.22	71.76	71.04
	JDA	65.80	75.22	69.18	75.26	71.37
	TCA	58.01	80.19	68.63	75.07	70.48
	DAN	49.32	70.01	71.96	73.83	66.28
	DAAN	57.22	76.37	81.92	73.96	72.37
Multi-source	MFSAN	59.48	72.17	83.43	80.96	74.01
	$MUCT_{SV}$	62.64	73.06	81.46	80.96	74.53
	$MUCT_{NC}$	64.45	73.82	80.64	80.83	74.94
	$MUCT$	**69.41**	79.42	**84.94**	**82.51**	**79.07**

Bold indicates the best result

Table 5 Data distribution of real data set

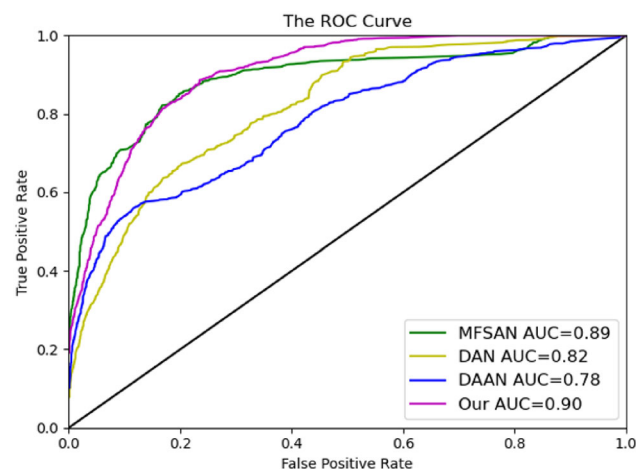| Subject	Total	P_1	P_2	P_3	P_4
0	214	55	59	47	53
1	233	62	40	64	67
2	223	63	42	63	55
3	220	56	45	56	63
4	213	53	43	59	58
5	202	58	38	60	46
6	220	54	56	66	44
7	184	25	51	62	46
8	155	15	33	61	46
9	273	72	63	68	70
10	258	79	53	70	56
11	664	236	202	121	105
12	249	58	62	66	63

Fig. 6 ROC curve on real data set

domains improves the classification performance of target domains. Since our model does not need the target domain data label, it can quickly process the data of new users without manual data processing. Because of its multi-source domain nature, it can take full advantage of previous data from multiple humans rather than just using a single human.

We use the idea of co-training and using classifiers and high-confidence data screening to improve the performance and stability of our model. Tables 2, 3 and 4 compare the performance of our model and other methods. After introducing high-confidence data, our classifier can better adapt to the data distribution in the target domain. With the addition of a multi-perspective classifier, it can classify data from multiple perspectives. It can make better use of some hidden features of the data and improve the stability of the model. Figure 3 shows our model's better stability and performance. Our model better adapts to the data imbalance problem using a weighted sample, classifiers, and high-confidence data. Figures 5 and 6 show the excellent performance of our model on unbalanced data sets.

Our model avoids the manual design of activity recognition features and is an end-to-end neural network. It does not need to design features by hand, as with traditional methods. It significantly reduces manual work compared with traditional and supervised models. In the feature extraction module, we use a CNN as the backbone network. For long-term and repetitive behaviors, CNN has a huge advantage. An RNN is more suitable for short-time and natural sorting behaviors [55]. Our model can replace the backbone network with more realistic scenarios.

While experiments showed that our model has good results on public and real-world data sets for personalized HAR,

some issues still must be considered. The quality of the pseudo-label of high-confidence data is related to the performance of our model. Therefore, our model requires the classifier to perform better when filtering high-confidence data. We must rely on experience to set suitable parameters for filtering high-confidence data, and to find the best parameters is a challenge. Moreover, as the data in the target domain are more important to the classifier, the classifier should pay more attention to it.

To address the shortcomings of our model, we propose some future work. After obtaining high-confidence data, we can use a method similar to TrAdaboost [56] to weigh the source domain and high-confidence data. In addition, since samples in the source domain contribute differently to classification results, we can weight them according to the similarity between the source and target domains when integrating multiple classification results.

Conclusion

We proposed the MUCT model, which can offers a viable solution for personalized HAR. Our model communicates differences in the activities of individuals and can quickly recognize new human activity data using previous such data. In the absence of target domain data labels, our model utilizes co-training to make full use of target domain data. Multiple-view classifiers are added to improve the model's performance. The model is an end-to-end network that can automatically extract features with less labor than traditional methods. Experimental results on public and real-world data sets show that our model yields superior personalized classification results in HAR. The model also shows its robustness in practical applications.

References

1. Arzani MM, Fathy M, Azirani AA, Adeli E (2021) Switching structured prediction for simple and complex human activity recognition. IEEE Trans Cybern 51(12):5859–5870

2. Panwar M, Biswas D, Bajaj H, Jöbges M, Turk R, Maharatna K, Acharyya A (2019) Rehab-net: deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation. IEEE Trans Biomed Eng 66(11):3026–3037

3. Chen Y, Yu L, Ota K, Dong M (2018) Robust activity recognition for aging society. IEEE J Biomed Health Inform 22(6):1754–1764

4. Zhou X, Liang W, Wang KI-K, Wang H, Yang LT, Jin Q (2020) Deep-learning-enhanced human activity recognition for internet of healthcare things. IEEE Internet Things J 7(7):6429–6438

5. Hassan MM, Uddin MZ, Mohamed A, Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. Future Gener Comput Syst 81:307–313

6. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2021) Deep learning for sensor-based human activity recognition: overview, challenges, and opportunities. ACM Comput Surv (CSUR) 54(4):1–40

7. Shao L, Zhu F, Li X (2014) Transfer learning for visual categorization: a survey. IEEE Trans Neural Netw Learn Syst 26(5):1019–1034

8. Du F, Yang P, Jia Q, Nan F, Chen X, Yang Y (2023) Global and local mixture consistency cumulative learning for long-tailed visual recognitions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15814–15823

9. Torrey L, Shavlik J (2010) Transfer learning. Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, Hershey, pp 242–264

10. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big data 3(1):1–40

11. Bengio Y (2012) Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML workshop on unsupervised and transfer learning. JMLR Workshop and Conference Proceedings, pp. 17–36

12. Wang P, Yang Y, Xia Y, Wang K, Zhang X, Wang S (2022) Information maximizing adaptation network with label distribution priors for unsupervised domain adaptation. IEEE Trans Multimedia. https://doi.org/10.1109/TMM.2022.3203574

13. Zhu Y, Chen Y, Lu Z, Pan SJ, Xue G-R, Yu Y, Yang Q (2011) Heterogeneous transfer learning for image classification. In: Twenty-fifth AAAI conference on artificial intelligence

14. Quattoni A, Collins M, Darrell T (2008) Transfer learning for image classification with sparse prototype representations. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8

15. Gao Y, Mosalam KM (2018) Deep transfer learning for image-based structural damage recognition. Comput Aided Civ Infrastruct Eng 33(9):748–768

16. Jia Q, Guo J, Du F, Yang P, Yang Y (2022) A fast texture-to-stain adversarial stain normalization network for histopathological images. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp. 2294–2301

17. Houlsby N, Giurgiu A, Jastrzebski S, Morrone B, De Laroussilhe Q, Gesmundo A, Attariyan M, Gelly S (2019) Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning. PMLR, pp. 2790–2799

18. Ruder S, Peters ME, Swayamdipta S, Wolf T (2019) Transfer learning in natural language processing. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials, pp. 15–18

19. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ (2019) Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683

20. Pan SJ, Zheng VW, Yang Q, Hu DH (2008) Transfer learning for wifi-based indoor localization. In: Association for the advancement of artificial intelligence (AAAI) workshop, vol. 6. The Association for the Advancement of Artificial Intelligence Palo Alto

21. Diba A, Fayyaz M, Sharma V, Karami AH, Arzani MM, Yousefzadeh R, Van Gool L (2017) Temporal 3d convnets: New architecture and transfer learning for video classification. arXiv preprint arXiv:1711.08200

22. Tian X, Tao D, Rui Y (2012) Sparse transfer learning for interactive video search reranking. ACM Trans Multimedia Comput Commun Appl (TOMM) 8(3):1–19

23. Yang Y, Guo J, Wang P, Wang Y, Yu M, Wang X, Yang P, Sun L (2021) Reservoir hosts prediction for covid-19 by hybrid transfer learning model. J Biomed Inform 117:103736

24. Rozantsev A, Salzmann M, Fua P (2018) Beyond sharing weights for deep domain adaptation. IEEE Trans Pattern Anal Mach Intell 41(4):801–814

25. Yao Y, Doretto G (2010) Boosting for transfer learning with multiple sources. In: IEEE computer society conference on computer vision and pattern recognition. IEEE 2010, pp. 1855–1862

26. Peng X, Bai Q, Xia X, Huang Z, Saenko K, Wang B (2019) Moment matching for multi-source domain adaptation. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1406–1415

27. Fang M, Guo Y, Zhang X, Li X (2015) Multi-source transfer learning based on label shared subspace. Pattern Recognit Lett 51:101–106

28. Li X, Jiang H, Xie M, Wang T, Wang R, Wu Z (2022) A reinforcement ensemble deep transfer learning network for rolling bearing fault diagnosis with multi-source domains. Adv Eng Inform 51:101480

29. Shen S, Sadoughi M, Li M, Wang Z, Hu C (2020) Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. Appl Energy 260:114296

30. Singh J, Hanson J, Paliwal K, Zhou Y (2019) Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nat Commun 10(1):1–13

31. Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput Surv (CSUR) 46(3):1–33

32. Wang J, Chen Y, Hu L, Peng X, Philip SY (2018) Stratified transfer learning for cross-domain activity recognition. In: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, pp. 1–10

33. Capela N, Lemaire E, Baddour N, Rudolf M, Goljar N, Burger H (2016) Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. J Neuroeng Rehabilit 13(1):1–10

34. Anguita D, Ghio A, Oneto L, Parra Perez X, Reyes Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In: Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning, pp. 437–442

35. Lara OD, Labrador MA (2012) A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor 15(3):1192–1209

36. Kim E, Helal S, Cook D (2009) Human activity recognition and pattern discovery. IEEE Pervasive Comput 9(1):48–53

37. Vrigkas M, Nikou C, Kakadiaris IA (2015) A review of human activity recognition methods. Front Robot AI 2:28

38. Bayat A, Pomplun M, Tran DA (2014) A study on human activity recognition using accelerometer data from smartphones. Procedia Comput Sci 34:450–457

39. Hossain HS, Khan MAAH, Roy N (2017) Active learning enabled activity recognition. Pervasive Mob Comput 38:312–330

40. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J (2014) Convolutional neural networks for human activity recognition using mobile sensors. In: 6th international conference on mobile computing, applications and services. IEEE, pp. 197–205

41. Lee S-M, Yoon SM, Cho H (2017) Human activity recognition from accelerometer data using convolutional neural network. In: IEEE international conference on big data and smart computing (bigcomp). IEEE 2017, pp. 131–134

42. Sun S, Shi H, Wu Y (2015) A survey of multi-source domain adaptation. Inf Fusion 24:84–92

43. Zhu Y, Zhuang F, Wang D (2019) Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In: Proceedings of the AAAI Conference on Artificial Intelligence, vo. 33(01), pp. 5989–5996

44. Wang J, Zheng VW, Chen Y, Huang M (2018) Deep transfer learning for cross-domain activity recognition. In: Proceedings of the 3rd International Conference on Crowd Science and Engineering, pp. 1–8

45. Zhao Z, Chen Y, Liu J, Shen Z, Liu M (2011) Cross-people mobile-phone based activity recognition. In: Twenty-second international joint conference on artificial intelligence

46. Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A (2012) A kernel two-sample test. J Mach Learn Res 13(1):723–773

47. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778

48. Long M, Wang J, Ding G, Sun J, Yu PS (2013) Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE international conference on computer vision, pp. 2200–2207

49. Pan SJ, Tsang IW, Kwok JT, Yang Q (2010) Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 22(2):199–210

50. Long M, Cao Y, Wang J, Jordan M (2015) Learning transferable features with deep adaptation networks. In: International conference on machine learning. PMLR, pp. 97–105

51. Yu C, Wang J, Chen Y, Huang M (2019) Transfer learning with dynamic adversarial adaptation network. In: 2019 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 778–786

52. Chen Y, Qin X, Wang J, Yu C, Gao W (2020) Fedhealth: a federated transfer learning framework for wearable healthcare. IEEE Intell Syst 35(4):83–93

53. Barshan B, Yüksek MC (2014) Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. Comput J 57(11):1649–1667

54. Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SigKDD Explor Newslett 12(2):74–82

55. Hammerla N, Halloran S, Ploetz T (2016) Deep, convolutional, and recurrent models for human activity recognition using wearables. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. Newcastle University

56. Dai W, Yang Q, Xue G-R, Yu Y (2007) Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning, ser. ICML '07. New York, NY, USA: Association for Computing Machinery, pp. 193–200. https://doi.org/10.1145/1273496.1273521