

“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Heterogeneous Unsupervised Domain Adaptation Based on Fuzzy Feature Fusion

Feng Liu, Guangquan Zhang, Jie Lu

Decision Systems and e-Service Intelligence (DeSI) Lab,
Centre for Artificial Intelligence, Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia

Feng.Liu-2@student.uts.edu.au; {Guangquan.Zhang, Jie.Lu}@uts.edu.au

Abstract— Domain adaptation is a transfer learning approach that has been widely studied in the last decade. However, existing works still have two limitations: 1) the feature spaces of the domains are homogeneous, and 2) the target domain has at least a few labeled instances. Both limitations significantly restrict the domain adaptation approach when knowledge is transferred across domains, especially in the current era of big data. To address both issues, this paper proposes a novel fuzzy-based heterogeneous unsupervised domain adaptation approach. This approach maps the feature spaces of the source and target domains onto the same latent space constructed by fuzzy features. In the new feature space, the label spaces of two domains are maintained to reduce the probability of negative transfer occurring. The proposed approach delivers superior performance over current benchmarks, and the heterogeneous unsupervised domain adaptation (HeUDA) method provides a promising means of giving a learning system the associative ability to judge unknown things using related knowledge.

Keywords—domain adaptation, transfer learning, fuzzy features, heterogeneous feature space

I. INTRODUCTION

The aim of domain adaptation is to use the knowledge learned from one or more source domains in a specific target domain, with the intention of improving learning accuracy in an almost unknown but similar field [1,2]. For example, knowledge from English documents could be used to classify a French document; credit records collected from Germany could be applied to improve the assessment accuracy of Australia’s credit records which suffer from data shortage; labeled databases could be used to classify similar but unlabeled datasets; or the forecasting accuracy of a new electricity grid could be improved by using the existing knowledge of an old grid. These examples clearly indicate that good domain adaptation methods can reduce the cost of manually labeling a huge dataset while improving classification or regression accuracy in a new, almost unknown field. The amount of data in the current era is increasing with incredible speed, which means that labeling all existing data comes at a high cost or is an impossible mission. Using the knowledge (labeled data) from source domains to explore unknown areas (unlabeled data) in an infinite number of target domains has thus become extremely important.

Many approaches have been proposed to deal with five types of domain adaptation: 1) homogeneous supervised domain adaptation (HoSDA) [3]; 2) homogeneous semi-supervised domain adaptation (HoSSDA) [4]; 3) homogeneous unsupervised domain adaptation (HoUDA) [5]; 4) heterogeneous supervised domain adaptation (HeSDA) [6]; and 5) heterogeneous semi-supervised domain adaptation (HeSSDA) [7]. In each of these, “homogeneous” represents source domains and target domains have the same number of related features, and “heterogeneous” represents source domains and target domains have different number of related features.

In domain adaptation types 1), 2), 4) and 5), however, existing approaches all need labeled instances in the target domain, which is a huge obstacle to applying these approaches to an infinite number of target domains. To deal with type 3), approaches are limited to source domains that have the same feature space as the target domain, which does not apply to some situations in the current environment.

It is therefore necessary to consider another type of domain adaptation: heterogeneous unsupervised domain adaptation (HeUDA), which is a widespread type in this big data era. Kernel canonical correlation analysis (KCCA) [8] has been proposed in existing works for heterogeneous unsupervised domain adaptation when the source domain and target domain have paired instances. However, when source domain and target domain do not have paired instances, KCCA is not appropriate for addressing the issue in which a target domain is unlabeled and has a different feature from with the source domain.

To overcome the drawbacks of current research in which a target domain is unlabeled and has a different feature space from the source domain, this paper proposes a novel HeUDA approach based on fuzzy feature fusion (FFF), which can map heterogeneous domains onto the same latent space using fuzzy features. In the mapped latent space, domains can be adapted by geodesic flow kernel (GFK) [5]. Compared to KCCA (the only existing model for HeUDA), the proposed model, referred to as FFF-GFK, has the following advantages:

- 1) FFF-GFK does not require the source domain and the target domain to have the same number of instances;
- 2) FFF-GFK does not require the source domain and the target domain to have paired instances;
- 3) FFF-GFK has a much lower probability of returning negative transfer results than KCCA.

These advantages result in the proposed model having better performance than KCCA and other benchmarks. The main contributions of this paper are as follows:

- 1) An effective approach for domain adaptation on the problem type HeUDA is proposed;
- 2) A new criterion is used to evaluate negative transfer for classification tasks;
- 3) The proposed model outperforms the current HeUDA approaches.

This paper is organized as follows. Section II reviews related works. Section III introduces the proposed novel fuzzy-based HeUDA approach. Section IV demonstrates the classification results of the proposed HeUDA approach and benchmarks. Section V concludes the paper and discusses future works.

II. RELATED WORK

In this section, we will briefly review related domain adaptation approaches developed in recent years, which fall into three categories: supervised domain adaptation (SDA), semi-supervised domain adaptation (SSDA) and unsupervised domain adaptation (UDA).

A. Supervised domain adaptation

Many approaches have been proposed for the traditional research topic of SDA. SDA approaches only use labeled instances in source and target domains to train prediction models, and unlabeled instances are unable to improve the training accuracy of prediction models. There are two categories of SDA, based on the relationship between the feature spaces of the source domain and the target domain: homogeneous SDA (HoSDA) and Heterogeneous SDA (HeSDA). The representative HoSDA approaches are transfer component analysis (TCA) [3], geodesic flow kernel (GFK), information-theoretical learning (ITL) [9] and stationary subspace analysis (SSA) [10]. The representative HeSDA approaches are asymmetric regularized cross-domain transformation (ARC-t) [11], and heterogeneous feature augmentation (HFA) [6].

B. Semi-supervised domain adaptation

SSDA approaches aim to apply the information of unlabeled data in the source and target domains to help improve classification accuracy. Similar to SDA approaches, there are two major categories of SSDA: homogeneous SSDA (HoSSDA) and Heterogeneous SDA (HeSSDA). The representative HoSSDA approaches are covariate shift [12], self-labeling adaptation [4], semi-supervised kernel matching for domain adaptation (SSKMDA) [7] and fuzzy sets [13–15]. The representative HeSSDA approaches are heterogeneous spectral mapping (HeMap) [16], manifold alignment-based approaches (MA) [17], SSKMDA [7], and the DASH-N method [18].

C. Unsupervised domain adaptation

When there are no labeled instances in the target domain, the knowledge transfer process is regarded as UDA. There is one basic assumption: two domains are related and have similar prediction tasks. Since there is no reliable method to transfer knowledge from heterogeneous source domains to a target domain, the principal UDA technologies deal with the issue as if the source domain and target domain are

homogeneous. This kind of technology is known as homogeneous UDA (HoUDA). The representative HoUDA approaches are similar to the HoSDA approaches and include TCA, GFK, ITL and SSA. As mentioned in Section I, kernel canonical correlation analysis (KCCA) is the only approach for HeUDA.

III. FUZZY-BASED HETEROGENEOUS UNSUPERVISED DOMAIN ADAPTATION

This section will introduce the proposed approach for the HeUDA issue based on fuzzy feature fusion. To clearly demonstrate our approach, the first subsection introduces notations used in this section and the second subsection explains the motivation of our paper. The fuzzy feature fusion method will then be proposed. The final subsection presents a novel fuzzy-based HeUDA approach to address the situation in which the target domain is unlabeled and has a feature space that is heterogeneous with the source domain.

A. Notations and problem statement

In this section, we assume for simplicity that there is one source domain D_S and one target domain D_T with different feature spaces, in which $D_S = \{X_{S_i}, y_{S_i}\}$ and $D_T = \{X_{T_i}\}$, where $X_{S_i} \in \mathbb{R}^n$, $X_{T_i} \in \mathbb{R}^m$, $y_{S_i} \in \{0, 1\}$, $m \neq n$ and $i = 1, 2, \dots, L$ (L is the number of instances in each domain). Our aim is to label each instance in D_T using the knowledge from D_S . It is important to clarify that there are no labeled instances in the target domain.

B. Motivation

When a source domain and a target domain have different feature spaces, the model trained by the source domain cannot be applied to label instances of the target domain. A common idea is thus to map two domains onto the same feature space so that instances of the target domain can be predicted by the model trained by the source domain. There are three common ways: 1) dimension reduction, 2) feature selection, and 3) feature fusion [19].

The dimension reduction approach aims to find the most representative information of a feature space and construct this important information as a new feature space. It has the following representative methods: principal component analysis (PCA), linear discriminant analysis (LDA) and kernel PCA (KPCA). In traditional machine learning tasks, training sets and testing sets have the same feature space, so the dimension reduction method can reduce the dimension of two sets simultaneously, which will not change the relationship between the label spaces of the training set and testing set. However, in the field of heterogeneous domain adaptation, these methods can only be used to separately reduce the dimension of the source and target domains, which does not ensure that the label spaces of the two domains remain unchanged in the dimension reduction process. This will cause the negative transfer (shown in the experiments). The negative transfer will let original label space change a lot. For example, the original label “1” will be changed to “2” with higher probability after the negative transfer happens, which will significantly influence the prediction accuracy.

Feature selection and feature fusion methods aim to find a matrix U to map an original feature space to a new feature space, which can be expressed by the following function:

$$X_S^{new} = X_S U^T \quad (1)$$

$X_S \in \mathbb{R}^{L \times n}$ is the original feature space and U is an $r \times n$ matrix, while r is the number of new features. For feature selection methodologies, $U \in \{0, 1\}^{r \times n}$ because these methodologies only use existing features, rather than generating new features. For feature fusion methodologies, $U \in [0, 1]^{r \times n}$, which means that feature fusion methods are able to use more information than feature selection.

C. Fuzzy feature fusion

In this section, the proposed model will be introduced in two parts: 1) model establishment, and 2) model learning process.

1) Model establishment

We propose a new feature fusion approach, fuzzy feature fusion (FFF), to overcome the drawbacks of dimension reduction methodologies. This new approach has a higher probability of maintaining unchanged relationships between the label spaces of the source domain and the target domain than traditional dimension reduction methodologies, as shown in the experiment section of this paper. As shown in Eq. (1), we need to calculate U_S for $X_S \in \mathbb{R}^{L \times n}$ and U_T for $X_T \in \mathbb{R}^{L \times m}$. We first provide a definition of fuzzy features.

Definition 1 (fuzzy features). Consider that there are n original features denoted by f_1, \dots, f_n , the i^{th} fuzzy feature is constructed using the following formula ($i = 1, 2, \dots, r$).

$$f_i^{fuzzy} = \frac{\sum_{j=1}^n \mu_{ij} f_j}{\sum_{j=1}^n \mu_{ij}}, \sum_{i=1}^r \mu_{ij} = 1 \quad (2)$$

Where μ_{ij} is the membership grade of the j^{th} original feature to the i^{th} fuzzy feature.

From Definition 1, it is clear that fuzzy features are combined by original features with corresponding membership grades. Calculating μ_{ij} is thus the key to fuzzy features. For a clear demonstration of μ_{ij} , the definition of a fuzzy features matrix follows.

Definition 2 (fuzzy features matrix). Given an instance matrix $X \in \mathbb{R}^{L \times n}$ where L is the number of instances and n is the number of features, an instance matrix based on fuzzy features can be expressed by the following term:

$$X^{fuzzy} = \begin{pmatrix} \frac{1}{\sum_{j=1}^n \mu_{1j}} & \dots & \frac{1}{\sum_{j=1}^n \mu_{1r}} \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \\ \frac{1}{\sum_{j=1}^n \mu_{1j}} & \dots & \frac{1}{\sum_{j=1}^n \mu_{1r}} \end{pmatrix} \odot XU^T, \sum_{i=1}^r \mu_{ij} = 1$$

$U \in [0, 1]^{r \times n}$ is a membership matrix, and μ_{ij} is the element in the i^{th} row and j^{th} column of U , and \odot represents the element-wise product of vectors.

The definition of the cost function follows.

Definition 3 (cost function). Given an instance matrix $X \in \mathbb{R}^{L \times n}$ and a corresponding membership matrix $U \in [0, 1]^{r \times n}$ where L is the number of instances and n is the number of features, the cost function of the proposed model is defined as follows:

$$J(U, X) = \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \left\| X_{*j} - \frac{1}{\sum_{j=1}^n \mu_{ij}} \odot XU_{i*}^T \right\|^2 \quad (3)$$

Where X_{*j} is the j^{th} column of X , U_{i*} is the i^{th} row of U , and μ_{ij} is the element in the i^{th} row and j^{th} column of U .

In Definition 3, XU_{i*}^T represents the i^{th} fuzzy feature and $\|X_{*j} - XU_{i*}^T\|^2$ is the Euclidean distance between the j^{th} original feature and the i^{th} fuzzy feature. We can thus establish our model by minimizing the cost function $J(U, X)$. In detail, the proposed model can be expressed as:

Model (fuzzy feature fusion). Given an instance matrix $X \in \mathbb{R}^{L \times n}$ where L is the number of instances and n is the number of features, the proposed model, denoted by fuzzy feature fusion, aims to obtain the fuzzy feature matrix by finding the best membership matrix U^* . U^* satisfies the following equation:

$$J(U^*, X) = \min_U \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \left\| X_{*j} - \frac{1}{\sum_{j=1}^n \mu_{ij}} \odot XU_{i*}^T \right\|^2 \quad (4)$$

2) Model Learning

This section will introduce how to learn the best membership matrix U^* of the proposed model. For simplicity, we define the following terms:

$$\frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} := \frac{1}{\sum_{j=1}^n \mu_{ij}} \odot XU_{i*}^T \quad (5)$$

$$(U_{i*}^T)^m := \underbrace{U_{i*}^T \odot \dots \odot U_{i*}^T}_m \quad (6)$$

We then consider that there are r latent centers C_i , which are defined using Eqs. (5) and (6).

$$C_i := \frac{X(U_{i*}^T)^2}{\sum_{j=1}^n \mu_{ij}^2} \quad (7)$$

Eq. (7) is taken from the clustering centers updating step of the fuzzy c-means algorithm and, based on Eq. (7), we have the following inequality.

$$\left\| X_{*j} - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \leq \|X_{*j} - C_i\|^2 + \left\| C_i - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \quad (8)$$

Based on (8), Theorem 1 is provided to minimize the cost function $J(U, X)$ using another cost function $J_1(U, X)$.

Theorem 1. Given an instance matrix $X \in \mathbb{R}^{L \times n}$ where L is the number of instances and n is the number of features, the cost function $J(U, X) \rightarrow 0$ if the cost function $J_1(U, X) \rightarrow 0$. $J_1(U, X)$ is defined as

$$J_1(U, X) = \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \|X_{*j} - C_i\|^2$$

Proof. Based on Eqs. (3) and (8), we arrive at the following inequalities.

$$J(U, X) \leq \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \left(\|X_{*j} - C_i\|^2 + \left\| C_i - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \right)$$

$$\begin{aligned}
&= \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \|X_{*j} - C_i\|^2 + \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \left\| C_i - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \\
&= \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \|X_{*j} - C_i\|^2 + \sum_{i=1}^r \left\| C_i - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \sum_{j=1}^n \mu_{ij}
\end{aligned}$$

Since

$$\left\| \frac{X(U_{i*}^T)^2}{\sum_{j=1}^n \mu_{ij}^2} - \frac{XU_{i*}^T}{\sum_{j=1}^n \mu_{ij}} \right\|^2 \sum_{j=1}^n \mu_{ij} < \sum_{j=1}^n \mu_{ij} \left\| X_{*j} - \frac{X(U_{i*}^T)^2}{\sum_{j=1}^n \mu_{ij}^2} \right\|^2$$

we have the following inequality

$$\mathbf{J}(U, X) \leq 2 \sum_{i=1}^r \sum_{j=1}^n \mu_{ij} \|X_{*j} - C_i\|^2$$

It is therefore evident that $\mathbf{J}(U, X) < 2\mathbf{J}_1(U, X)$. This means that the cost function $\mathbf{J}(U, X)$ will approach 0 if the cost function $\mathbf{J}_1(U, X)$ approaches 0. \square

Theorem 1 shows that we can use U^* , which minimizes $\mathbf{J}_1(U, X)$ as the best U^* in Eq. (4). Then, applying the method of Lagrange multipliers, the updating equation of membership matrix U is as follows:

$$\mu_{ij} = \left(\sum_{k=1}^r \left(\frac{\left\| X_{*j} - \frac{X(U_{k*}^T)^2}{\sum_{j=1}^n \mu_{kj}^2} \right\|^2}{\left\| X_{*j} - \frac{X(U_{i*}^T)^2}{\sum_{j=1}^n \mu_{ij}^2} \right\|^2} \right)^2 \right)^{-1} \quad (9)$$

Based on Eq. (9), Algorithm 1 describes the learning steps of the proposed model, which can obtain the fuzzy features matrix (FFM) generated by X and U^* .

Algorithm 1. Fuzzy feature fusion (\mathbf{X} , r , ϵ)

Input:

\mathbf{X} : a matrix representing the set of instances
 r : the number of fuzzy features
 ϵ : minimum posteriori error

Output:

FFM (fuzzy features matrix) generated by X

```

1   $[l, n] = \text{size}(\mathbf{X})$ ; % to obtain the size of  $\mathbf{X}$ 
2   $U = \text{rand}(r, n)$ ; % to generate initial  $U$ 
3   $\epsilon_0 = 1$ ; % to initialize posteriori error
4  WHILE  $\epsilon_0 > \epsilon$ 
5       $E = \mathbf{J}_1(U, \mathbf{X})$ ;
6      Update  $\mu_{ij}$  using Eq. (9);
7       $\epsilon_0 = |E - \mathbf{J}_1(U, \mathbf{X})|$ ; % to calculate posteriori error
8  END
9   $\text{FFM} = \mathbf{X}U^T$ ;
10 Return FFM

```

D. A novel fuzzy-based HeUDA approach: FFF-GFK

This section will introduce the detailed procedures of the novel fuzzy-based HeUDA approach proposed in this paper. There are four major procedures: 1) data pre-processing, 2) fuzzy feature fusion, 3) domain adaptation, and 4) prediction. Figure 1 shows the main steps of the proposed approach.

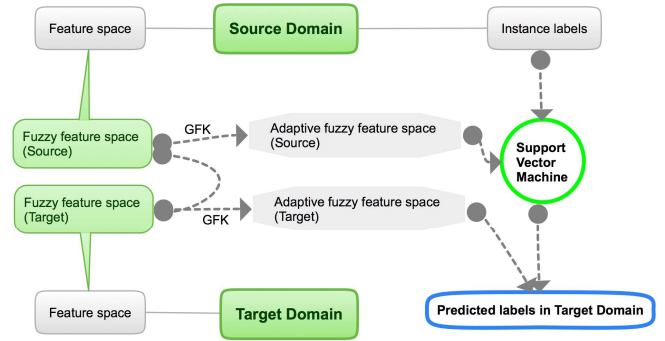


Fig. 1. Main steps of FFF-GFK. For the source domain (normalized), it is constructed by two parts: the feature space and instance labels. For the target domain, it is constructed by the feature space. Both feature spaces are used to construct fuzzy feature spaces, next, fuzzy feature spaces are mapped by GFK. On the mapped space, knowledge can be transferred to the target domain.

1) Data pre-processing

For two feature spaces X_S and X_T , we adopt the same method of normalizing them, which can be expressed using the following functions:

$$f_{pp}(X_{*j}) = \frac{X_{*j} - \min(X_{*j})}{\max(X_{*j}) - \min(X_{*j})}, j = 1, \dots, n$$

$$F_{pp}(X) = (f_{pp}(X_{*1}) f_{pp}(X_{*2}) \dots f_{pp}(X_{*n}))$$

X_{*j} is the j^{th} column of X and n is the number of features of X . $f_{pp}(X_{*j})$ is an L by 1 vector and $F_{pp}(X)$ is an L by n matrix.

We can then obtain two pre-processed feature spaces using $X_S = F_{pp}(X_S)$ and $X_T = F_{pp}(X_T)$.

2) Fuzzy feature fusion

For pre-processed X_S and X_T , we can use Algorithm 1 to obtain the corresponding fuzzy feature matrixes, denoted by FFM_S and FFM_T .

3) Domain adaptation

After mapping two features spaces onto the same feature space (X_S is mapped onto FFM_S and X_T is mapped onto FFM_T), FFM_S and FFM_T still have different distributions. We therefore need to map FFM_S and FFM_T onto another feature space where both have the same distribution. This process can be regarded as HoUDA.

We apply GFK to map FFM_S and FFM_T onto another feature space where they have the same distribution. There are five steps to GFK [5]:

- 1) standardize (z-score) FFM_S and FFM_T ;
- 2) calculate subspaces of FFM_S and FFM_T ;
- 3) use geodesic flow defined on Grassmann manifold to calculate kernel parameter G ;
- 4) calculate square root of G such that $R^T R = G$;
- 5) map FFM_S and FFM_T using R : $\text{FFM}_{SG} = R \times \text{FFM}_S$ and $\text{FFM}_{TG} = R \times \text{FFM}_T$ (in step 1 and step 2, FFM_S and FFM_T are replaced with corresponding calculation results).

4) Label Prediction

After adapting FFM_S and FFM_T to FFM_{SG} and FFM_{TG} , we can use FFM_{SG} to train a support vector machine (SVM) and

predict labels of instances in the target domain using trained SVM and FFM_{TG}. In this paper, we will use LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>) with default parameters.

For the source and target domains, they will first be normalized by step 1, and both feature spaces are separately used to construct fuzzy feature spaces. Based on two fuzzy feature spaces, GFK is applied to map two fuzzy feature spaces onto a latent space. Finally, the model trained by the mapped fuzzy feature space can be used to predict labels of instances in the target domain.

IV. EXPERIMENTS

In this section, we will apply our FFF-GFK to real datasets to test its prediction performance. We first introduce the relevant datasets. We will then describe the related benchmarks. Lastly, the prediction result will be analyzed.

A. Datasets for HeUDA

To vividly show the significance of HeUDA, we select two datasets from the UCI Machine Learning Repository (UMLR, <http://archive.ics.uci.edu/ml/index.html>) to test the proposed model and benchmarks. Both datasets are related to personal credit assessment, namely German Credit Data and Australian Credit Approval. They have totally different feature spaces. Our aim is to address two HeUDA tasks based on both datasets:

Task 1: G2A: Assume that the German data is labeled and the Australian data is unlabeled. Label “1” means “good credit” and label “2” means “bad credit”. This task is equivalent to the question: “Can we use knowledge from German credit records to label unlabeled Australian data?”

Task 2: A2G: Assume that the Australian data is labeled and the German data is unlabeled. Label “1” means “good credit” and label “2” means “bad credit”. This task is equivalent to the question: “Can we use knowledge from Australian credit records to label unlabeled German data?”

Table I shows the details of two HeUDA tasks. It should be clarified that the similarity of the two domains (both are datasets for evaluating personal credit) is the reason that the domain adaptation technique works on them (G2A or A2G).

TABLE I. DETAILS OF TWO HEUDA TASKS

Source Domain	Target Domain	Labels	Task Name
German Credit Data	Australian Credit Approval	1: Good	G2A
Australian Credit Approval	German Credit Data	1: Good	A2G

B. Benchmarks and parameter setting

As mentioned in previous sections, dimensional reduction technology can be applied to force the two domains to have the same number of features. We denote this method as DR_GFK (DG) where the dimensions of the mapped features are set as r . KCCA is regarded as another transfer method.

We also consider two non-transfer methods: all_1 (A1) and the clustering method (CM). A1 labels all instances as “1” and CM clusters the instances and randomly labels the

categories using the k -means algorithm. Because there are no labeled instances in the target domain, CM naturally has extremely unstable prediction results.

The parameters for the four benchmarks and the proposed approach are set as shown in Table II (r is the dimension of the mapped feature space and ϵ is the posteriori error of FFF-GFK and CM). For the DG approach, we only set the dimension of the mapped feature space, and other parameters are set as default values.

TABLE II. PARAMETER SETTINGS

Parameters	A1	CM	DG	FFF-GFK
r	-	-	10	10
ϵ	-	0.00001	-	0.00001

C. Prediction results and analysis

Because of the basic assumption that the source domain and target domain are similar and related, a highly unbiased situation should be avoided. Of the selected datasets, German Credit Data has an unbiased situation, so we randomly selected unbiased samples from German Credit Data in every experiment. We ran the experiment 50 times for each model and each task. All prediction results are reported in Table III, according to three criteria: mean accuracy (with the value of standard derivation), minimum accuracy, and maximum accuracy.

TABLE III. PREDICTION RESULTS OF FFF-GFK AND BENCHMARKS

Approach	Criteria	A2G	G2A
A1	Average Accuracy	50%	50%
	Min Accuracy	50%	50%
	Max Accuracy	50%	50%
	Accuracy		
CM	Average Accuracy	50.97%±5.21%	44.89%±0.40%
	Min Accuracy	43.67%	43.77%
	Max Accuracy	57.17%	56.23%
	Accuracy		
DG	Average Accuracy	50.92%±1.06%	45.19%±1.96%
	Min Accuracy	48.50%	41.16%
	Max Accuracy	53.00%	52.61%
	Accuracy		
KCCA	Average Accuracy	50.52%±4.64%	51.05%±9.72%
	Min Accuracy	41.00%	31.50%
	Max Accuracy	59.67%	71.00%
	Accuracy		
FFF-GFK	Average Accuracy	59.52%±2.92%	69.66%±6.91%
	Min Accuracy	50.33%	47.83%
	Max Accuracy	65.00%	77.50%
	Accuracy		

From Table III, it is clear that FFF-GFK outperforms other approaches. KCCA is better than DG, CM and A1 when the task is G2A. CM is better than A1, DG and KCCA when the task is A2G, which indicates that KCCA is invalid in this

task. FFF-GFK is actually better than other approaches, due mainly to the lower probability of causing negative transfer. For example, KCCA has high Max accuracy when the task is G2A but it always has the situation of negative transfer, resulting in its average accuracy being much lower and having higher standard derivation. The performance of G2A is better than that of A2G mainly because that German data has more features to describe the “credit” than Australian data has, which leads that more knowledge is transferred from German data to Australian data.

To show the negative transfer situation, we propose a new criterion to evaluate which approaches can avoid negative transfer with high probability. The new criterion (for the binary classification issue) is expressed as follows:

$$P_{ng} = \frac{1}{T} \text{Card}(\{i | a_i \leq 0.5, i = 1, \dots, T\})$$

T is the number of times the experiments are run ($T = 50$ in this paper) and a_i is the accuracy of the i^{th} experiment. Lower P_{ng} values represent higher probability of avoiding negative transfer. Table IV shows the values of P_{ng} in each approach.

TABLE IV. P_{ng} VALUES OF EACH APPROACH

Approach	A2G	G2A
A1	100%	100%
CM	22.00%	94.00%
DG	16.00%	98.00%
KCCA	42.00%	46.00%
FFF-GFK	0.00%	2.00%

Table III and Table IV clearly show that the proposed model is much better than existed benchmarks.

V. CONCLUSIONS AND FURTHER STUDY

In the current era of big data, the unlabeled nature of many very large datasets is a big obstacle to the application of traditional machine learning techniques. How to use existing labeled datasets to help classify or predict these unlabeled datasets is consequently a significant and challenging problem. To overcome the limitations of the existing approaches, we propose a fuzzy feature fusion method to map the original feature space onto a new feature space such that the heterogeneous feature spaces of the source and target domains can be mapped onto the same feature space. Based on this feature fusion technique, a novel fuzzy based approach is proposed to handle heterogeneous unsupervised domain adaptation. The proposed approach, FFF-GFK, significantly reduces the probability of causing negative transfer, ensuring that FFF-GFK achieves much better performance than the benchmark methods.

Based on the FFF-GFK, our future works include: 1) determining how to automatically select the dimension of the mapped feature space, and 2) investigating how to avoid negative transfer from the theoretical perspective. The first aspect aims to make FFF-GFK a non-parameter approach, so that it can be widely used in many datasets without parameter adjustments. The second aspect will provide a theoretical foundation for the proposed HeUDA method.

ACKNOWLEDGMENT

The work presented in this paper was supported by Australian Research Council (ARC) under discovery grants DP140101366 and DP170101632.

REFERENCES

- [1] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359.
- [2] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, G. Zhang, Transfer learning using computational intelligence: A survey, *Knowledge-Based Syst.* 80 (2015) 14–23.
- [3] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, *IEEE Trans. Neural Networks.* 22 (2011) 199–210.
- [4] J. Jiang, C. Zhai, A two-stage approach to domain adaptation for statistical classifiers, in: *Proc. ACM Conf. Conf. Inf. Knowl. Manag.*, ACM, 2007: pp. 401–410.
- [5] B. Gong, K. Grauman, F. Sha, Learning kernels for unsupervised domain adaptation with applications to visual object recognition, *Int. J. Comput. Vis.* 109 (2014) 3–27.
- [6] W. Li, L. Duan, D. Xu, I.W. Tsang, Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (2014) 1134–1148.
- [7] M. Xiao, Y. Guo, Feature space independent semi-supervised domain adaptation via kernel matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (2015) 54–66.
- [8] Y.R. Yeh, C.H. Huang, Y.C.F. Wang, Heterogeneous domain adaptation and classification by exploiting the correlation subspace, *IEEE Trans. Image Process.* 23 (2014) 2009–2018.
- [9] Y. Shi, L. Angeles, F. Sha, Information-theoretical learning of discriminative clusters for unsupervised domain adaptation, in: *Proc. Int. Conf. Mach. Learn.*, 2012: pp. 1079–1086.
- [10] J.S. Muller, P. von Bunau, F.C. Meinecke, F.J. Kiraly, K.R. Muller, The stationary subspace analysis toolbox, *J. Mach. Learn. Res.* 10 (2011) 3065–3069.
- [11] B. Kulis, K. Saenko, T. Darrell, What you saw is not what you get: Domain adaptation using asymmetric kernel transforms, in: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2011: pp. 1785–1792.
- [12] T. Kanamori, S. Hido, M. Sugiyama, A least-squares approach to direct importance estimation, *J. Mach. Learn. Res.* 10 (2009) 1391–1445.
- [13] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, J. Lu, Fuzzy regression transfer learning in Takagi-Sugeno fuzzy models, *IEEE Trans. Fuzzy Syst.* (2016).
- [14] V. Behbood, J. Lu, G. Zhang, W. Pedrycz, Multistep fuzzy bridged refinement domain adaptation algorithm and its application to bank failure prediction, *IEEE Trans. Fuzzy Syst.* 23 (2015) 1917–1935.
- [15] H. Zuo, G. Zhang, V. Behbood, J. Lu, W. Pedrycz, T. Zhang, Fuzzy transfer learning in data-shortage and rapidly changing environments, in: *Uncertain. Model. Knowl. Eng. Decis. Mak. Proc. 12th Int. FLINS Conf.*, World Scientific, 2016: pp. 175–180.
- [16] X. Shi, Q. Liu, W. Fan, P.S. Yu, Transfer across completely different feature spaces via spectral embedding, *IEEE Trans. Knowl. Data Eng.* 25 (2013) 906–918.
- [17] C. Wang, S. Mahadevan, Heterogeneous domain adaptation using manifold alignment, in: *Proc. Int. Jt. Conf. Artif. Intell.*, 2011: pp. 1541–1546.
- [18] H. V. Nguyen, H.T. Ho, S. Member, V.M. Patel, DASH-N: Joint hierarchical domain adaptation and feature learning, *IEEE Trans. Image Process.* 24 (2015) 5479–5491.
- [19] Q. Wu, Z. Wang, F. Deng, Z. Chi, D.D. Feng, Realistic human action recognition with multimodal feature selection and fusion, *IEEE Trans. Syst. Man, Cybern. Syst.* 43 (2013) 875–885.