Domain Selection of Transfer Learning in Fuzzy Prediction Models

Hua Zuo
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Hua.Zuo@uts.edu.au

Guangquan Zhang
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Guangquan.Zhang@ualberta.ca

Witold Pedrycz
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
wpedrycz@ualberta.ca

Jie Lu
Centre for Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia
Jie.Lu@uts.edu.au

***Abstract*—Transfer learning has emerged as a solution for the cases where little or no labeled data are available in the training process. It leverages the previously acquired knowledge (a source domain with a large amount of labeled data) to facilitate solving the current tasks (a target domain with little labeled data). Many transfer learning methods have been proposed, and especially fuzzy transfer learning method, which is based on fuzzy systems, has been developed because of its capability to deal with the uncertainty in transfer learning. However, there is one issue with fuzzy transfer learning that has not yet been resolved: the domain selection problem, which is heavily depended on the knowledge transfer method and the applied prediction model. In this work, we explore the domain selection problem in Takagi-Sugeno fuzzy model when multiple source domains are accessible, and define the similarity between the source and target domains to provide guidance for the domain selection. The experiments on synthetic datasets are designed to simulate the situations of multiple sources in transfer learning, and demonstrate the rationality of the proposed similarity in selecting the source domain for the target domain. Further, the real-world datasets are used to validate the proposed domain adaptation method, and verify its capability in solving practical situations.**

***Index Terms*—transfer learning, fuzzy systems, machine learning, multiple domains, domain selection**

I. Introduction

The achievement of machine learning has deeply affected in many areas, such as business management [1], biology [2], medical imaging [3], and computer vision [4]. However, the training process of many machine learning models require a large amount of labled data, which is difficult or even impossible to be obtained in the new emerging area. Additionally, the rapidly changing environment makes the models outdated quickly, and building a new model from scratch is time-consuming and expensive.

Transfer learning addresses the problem of how to leverage previously acquired knowledge to improve the efficiency and accuracy of learning in one domain that in some way relates to the original domain [5]. Many work has been studied in the transfer learning area from both the theory and application aspects, and some survey papers are presented in a specific area, such as activity recognition [6], reinforcement learning [7], and computational intelligence [8].

These methods have gained great achievement in dealing with domain adaptation problems but ignore the inherent phenomenon of uncertainty, a crucial factor during the knowledge transfer process [9]. There is a clear co-dependency between the level of certainty in learning a task and the amount of information that is available; problems with too little information have a high degree of uncertainty. The few labeled data in the target domain makes only a finite amount of information can be extracted, and leads to a a high degree of uncertainty. However, the introduction of fuzzy systems into transfer learning has shown promising results in overcoming this problem.

Many research work dedicates to the integration of fuzzy logic with transfer learning, and this area has drawn considerable attention. Deng et al. [10] developed the transfer learning methods in fuzzy models, including the Mamdani-Larsen-type fuzzy system and the Takagi-Sugeno-Kang fuzzy model, by defining two new objective functions. Further, their research results are also applied on the practical applications, for example, the recognition of electroencephalogram signals in environments with a data shortage. Behbood et al. [11] proposed a fuzzy-based transfer learning approach to long-term bank failure prediction models with source and target domains that have different data distributions. Liu et al. [12] focus on the unsupervised heterogeneous domain adaptation problem, and presented a novel transfer learning model via n-dimensional fuzzy geometry and fuzzy equivalence relations.

Despite these advancements in fuzzy system-based transfer learning methods, there is still a main issue that has not been resolved: how to select an appropriate domain for the target domain when multiple source domains are available?

Some of our own previous research has focused on developing the domain adaptation ability of fuzzy rule-based models in regression tasks [13]. A set of algorithms are proposed for two different scenarios, where the datasets from the source domain and target domain are in homogeneous and heterogeneous spaces [14], separately. In this paper, based on our previous work, we will explore the capability of fuzzy systems in handling knowledge transfer problems when multiple source domains are available. The contribution of this work is developing a novel source domain selection method based on the Takagi-Sugeno fuzzy model, which improves the effectiveness of transfer learning in the situation with multiple source domains.

The reminder of this paper is structured as followed. Section II presents the preliminaries, including some important definitions in transfer learning, and the Takagi-Sugeno fuzzy model, the foundation of our domain selection and adaptation method. Section III details the procedures of implementing domain selection in the transfer learning problem with multiple sources. The experiments in Sections IV and V validate the effectiveness of the proposed method using both synthetic and real-world datasets. The final section concludes the paper and outlines future work.

II. Preliminaries

This section begins with some basic definitions of transfer learning, followed by an introduction to the Takagi-Sugeno fuzzy model, which is the basic prediction model applied in our domain selection and domain adaptation methods.

A. Definition

Definition 1 (Domain) [5]: A domain is denoted by $D = \{F, P(X)\}$, where F is a feature space, and $P(X), X = \{x_1, x_2, ..., x_n\}$ are the probability distributions of the instances.

Definition 2 (Task) [5]: A task is denoted by $T = \{Y, f(\cdot)\}$, where $Y \in R$ is the response, and $f(\cdot)$ is an objective predictive function.

Definition 3 (Transfer Learning) [5]: Given a source domain D_s, a learning task T_s, a target domain D_t, and a learning task T_t, transfer learning aims to improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s an T_s, where $D_s \neq D_t$, or $T_s \neq T_t$.

In brief, transfer learning aims to use knowledge of a domain (from a source domain) to support the construction of prediction model in a new, but related domain (the target domain).

B. Takagi-Sugeno Fuzzy Model

A Takagi-Sugeno (TS) fuzzy model [15] is a commonly used regression model based on the combination of fuzzy rules in an nonlinear way. The TS model consists of c rules with the following representation:

$$\text{if } \boldsymbol{x} \text{ is } A_i(\boldsymbol{x},\ \boldsymbol{v}_i),\ \text{then } y \text{ is } L_i(\boldsymbol{x}, \boldsymbol{a}_i) \qquad i = 1, 2, \cdots, c \quad (1)$$

where $\boldsymbol{v}_i$ are the centers of the clusters that determine the layout of the fuzzy rules, and $\boldsymbol{a}_i$ are coefficients of the linear functions, which defines the action of each rule on the input variables.

The construction of the TS model, a set of fuzzy rules, is based on a labeled dataset $\{(\boldsymbol{x_1}, y_1), ..., (\boldsymbol{x_N}, y_N)\}$ using two procedures. In the first procedure, fuzzy C-means (FCM) is applied to divide the data in an unsupervised learning process, so that clusters are learned and the centers of the clusters are obtained. After getting the clusters, the coefficients of the linear functions, which are defined in each cluster, are computed using the labeled datasets. [16].

III. Domain Selection Method in TS Fuzzy Models

This section presents the method of selecting an appropriate domain for the target domain when multiple source domains are available. The domain selection problem in TS fuzzy model is stated with formulas firstly, then the procedures of selecting source domain are given with details, and at last, a theoretical analysis of the method performance index is included to benefit the illustration and analysis in the experiments.

A. Problem Statement

Consider h source domains, $\boldsymbol{S}_1$, $\boldsymbol{S}_2$,...,$\boldsymbol{S}_h$, with large amounts of labeled data, and a target domain with very little labeled data.

In the source domains, one of the dataset $\boldsymbol{S}_j$ is illustrated as an example:

$$\boldsymbol{S}_j = \{(\boldsymbol{x}_1^{sj}, y_1^{sj}), (\boldsymbol{x}_2^{sj}, y_2^{sj}), ..., (\boldsymbol{x}_{cj}^{sj}, y_{cj}^{sj})\} \qquad j \in \{1, ..., h\} \quad (2)$$

where $(\boldsymbol{x}_k^{sj}, y_k^{sj})$ represents the kth input-output data pair in the jth ($j \in \{1, ..., h\}$) source domain, $\boldsymbol{x}_k^{sj} \in R^n (k = 1, ..., N_{sj})$ is an n-dimensional input variable, the label $y_k^{sj} \in R$ is a continuous variable, and N_{sj} indicates the number of pairs.

The dataset $\boldsymbol{T}$ in the target domain consists of two subsets: one with labels and one without:

$$\begin{aligned}\boldsymbol{T} = \{\boldsymbol{T}_L, \boldsymbol{T}_U\} = \{\{(\boldsymbol{x}_1^t, y_1^t), ..., (\boldsymbol{x}_{N_{t1}}^t, y_{N_{t1}}^t)\},\\ \{\boldsymbol{x}_{N_{t1}+1}^t, ..., \boldsymbol{x}_{N_t}^t\}\}\end{aligned} \quad (3)$$

where $\boldsymbol{x}_k^t \in R^n (i = 1, ..., N_t)$ is an n-dimensional input variable, $y_k^t \in R$ is the label only accessible for the first N_{t1} data. $\boldsymbol{T}_L$ includes the instances with labels, and $\boldsymbol{T}_U$ contains the data without labels. The numbers of instances in $\boldsymbol{T}_L$ and $\boldsymbol{T}_U$ are N_{t1} and $N_t - N_{t1}$, respectively, and satisfy $N_{t1} << N_t$, $N_t << N_{s1}$,..., $N_t << N_{sh}$.

B. Select Domain from Multiple Sources

The method of selecting an appropriate source domain from multiple sources, and implementing knowledge transfer between the source and target domains can be summarized into five steps:

Step 1: Train TS fuzzy model for each source domain, separately.

Since a large amount of labeld data are available, a TS fuzzy model could be built for each source domain. A set of fuzzy rules, correspondingly, is obtained, denoted as $\boldsymbol{R}^{s1}, \boldsymbol{R}^{s2}, ...,$ and $\boldsymbol{R}^{sh}$:

$$\boldsymbol{R}^{sj} = \{r(\boldsymbol{v}_1^{sj}, \boldsymbol{a}_1^{sj}), r(\boldsymbol{v}_2^{sj}, \boldsymbol{a}_2^{sj}), ...r(\boldsymbol{v}_{cj}^{sj}, \boldsymbol{a}_{cj}^{sj})\} \tag{4}$$

where $r(\boldsymbol{v}_i^{sj}, \boldsymbol{a}_i^{sj})$ represents a fuzzy rule in the jth source domain with the center $\boldsymbol{v}_i^{sj}$ and coefficients of linear functions $\boldsymbol{a}_i^{sj}$. And the rule $r(\boldsymbol{v}_i^{sj}, \boldsymbol{a}_i^{sj})$ is represented as:

$$\begin{aligned} &\text{if } \boldsymbol{x}_k^{sj} \text{ is } A_i(\boldsymbol{x}_k^{sj},\ \boldsymbol{v}_i^{sj}), \text{ then } y_k^{sj} \text{ is } L_i(\boldsymbol{x}_k^{sj}, \boldsymbol{a}_i^{sj}) \\ &\qquad\qquad\qquad\qquad i = 1, 2, \cdots, cj \end{aligned} \tag{5}$$

where $j \in \{1, ..., h\}$.

From the obtained fuzzy rules, the centers of the clusters in each domain are also accessible and denoted as $\boldsymbol{V}^{sj}$:

$$\boldsymbol{V}^{sj} = \{\boldsymbol{v}_1^{sj}, \boldsymbol{v}_2^{sj}, ..., \boldsymbol{v}_{cj}^{sj}\} \tag{6}$$

Step 2: Obtain the centers of clusters in target data.

The distribution of target data plays a crucial role in the domain selection and adaptation process. First, the number of data clusters in the target domain is determined using Infinite Gaussian mixture model (IGMM) [17], which explores the structure of data based on distributions. Suppose the number of clusters in the target data is ct. Then, based on the obtained ct, the clustering algorithm FCM is applied to capture the centers of the clusters, denoted as $\boldsymbol{V}^t$:

$$\boldsymbol{V}^t = \{\boldsymbol{v}_1^t, \boldsymbol{v}_2^t, ..., \boldsymbol{v}_{ct}^t\} \tag{7}$$

Step 3: Augment the labeled target data using active learning technique.

The purpose of this procedure is to increase the amount information in the target domain by actively selecting and labeling some of the unlabeled target data. The idea is first evaluating the labeled target data in each cluster, and then labeling some unlabeled target data if the amount of labeled target data is less than a given threshold. The technique of active learning is applied to select the unlabeled data for labeling, and the detailed procedure could be referred in our previous work [18].

Step 4: Calculate the similarity between domains, and select a source domain for the target.

The similarity of a source domain and a target domain is defined as:

$$sim(\boldsymbol{S}_j, \boldsymbol{T}) = 1/(1 + exp(\frac{1}{cj * ct} \sum_{i=1}^{cj} \sum_{k=1}^{ct} ||\boldsymbol{v}_i^{sj} - \boldsymbol{v}_k^t||)) \tag{8}$$

where $\boldsymbol{v}_i^{sj}$ and $\boldsymbol{v}_k^t$ are the centers of clusters in the source and target domains, separately.

Step 5: Modify the rules from the selected source domain to fit the target data.

Suppose the selected source domain is $\boldsymbol{S}_1$, and then the rules in $\boldsymbol{R}^{s1}$ will be modified to be compatible with the target data. The input and output space of the fuzzy rules are changed through the mappings, and the parameters of the mappings are optimized using the labeled target data. The modified fuzzy rules under the mappings are represented as:

$$\begin{aligned} &\text{if } \boldsymbol{x}_k^t \text{ is } A_i(\Phi(\boldsymbol{x}_k^t,\ \boldsymbol{v}_i^{s1})), \text{ then } y_k^t \text{ is } \Psi_i(L_i(\Phi(\boldsymbol{x}_k^t, \boldsymbol{a}_i^{s1})) \\ &\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \cdots, c1 \end{aligned} \tag{9}$$

where Φ and Ψ are the mappings used to change the input and output spaces, separately.

The mappings of Φ and Ψ are constructed by a three layers network with one hidden layer. The transformation of the neurons through the hidden layer modifies the input and output space so that the new rules could fit the target data. The parameters of the mappings are optimized using the labeled target data [13].

C. Performance Index

This section provides the formulation of performance index for the models, so the proposed method is evaluated in a clear way in the experiment parts.

Two types of models are evaluated: the ”no transfer models”, and the ”single source transfer models”. ”No transfer models” mean the models in the source domains are used directly to solve the regression tasks in target domain. ”Single source transfer models” represent that one source model is modified and optimized to fit the target data. All the models will be tested on the unlabeled target dataset $\boldsymbol{T}_U$ to verify the ability of the models in solving regression tasks in the target domain. Next, the performance of these models will be represented with formulas.

The ”no transfer models” contains h models: the prediction model for each source domain, separately. The performance of these models on $\boldsymbol{T}_U$ are calculated as follows:

$$Q_{S_jT} = \sqrt{\frac{1}{N_t - N_{t1}} \sum_{k=1}^{N_t - N_{t1}} \sum_{i=1}^{cj} A_i(\boldsymbol{x}_k^t, \boldsymbol{v}_i^{sj}) L_i(\boldsymbol{x}_k^t, \boldsymbol{a}_i^{sj}) - y_k^t)^2} \tag{10}$$

where $j \in \{1, ..., h\}$.

The prediction accuracy of the ”single source transfer models” on $\boldsymbol{T}_U$ is represented as:

$$P_{S_jT} = \sqrt{\frac{1}{N_t - N_{t1}} \sum_{k=1}^{N_t - N_{t1}} \sum_{i=1}^{cj} (A_i(\Phi(\boldsymbol{x}_k^t, \boldsymbol{v}_i^{sj}) \Psi_i(L_i(\Phi(\boldsymbol{x}_k^t, \boldsymbol{a}_i^{sj}))) - y_k^t)^2} \tag{11}$$

where $j \in \{1, ..., h\}$.

IV. Experiments on Synthetic Datasets

The experiments using synthetic datasets are implemented to explore the performance of the domain selection method. The experiments includes two parts: the first part verifies the rationality of the defined similarity, and the second group of experiments consider the domain selection problem when the multiple source domains have the similar data structure. All the models' construction follows the five-fold cross validation, and the results are shown in the form of $mean \pm variance$.

A. Verify the Rationality of the Defined Similarity

Two groups of experiments are designed to simulate the domain adaptation problem with multiple source domains.

In the first group of experiments, two source domains and a target domain are accessible, and the input data of the datasets in the three domains are illustrated in Fig. 1. The points in blue and yellow are the source data, denoted as "source 1", and "source 2" respectively, and the red points represents the target data. However, the relations between the datasets in domains are quite different. We can see that only "source 1" has the similar data structure with the target data, and the distribution of "source 2" is not identical with the target data.

In the second group of experiments, three source domains are available for the target domain. The input data of the four domains are shown in Fig. 2. Similarly, the target data are represented with red points, and the data in the three source domains , which are denoted as "source 1", "source 2" and "source 3", are represented with points with blue, yellow, and black.

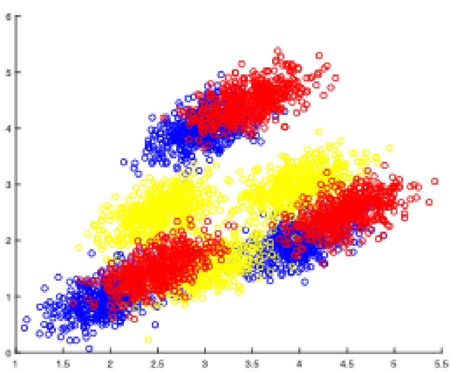

Fig. 1. Input data for datasets 1

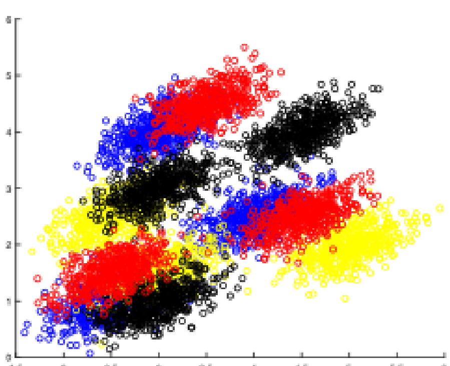

Fig. 2. Input data for datasets 2

The experimental results are shown in Tables I and II. The first column is the index of the source domains, and the second column shows the similarity of individual source domain and the target domain using formula 8. The third and fourth columns include the performance of two types of models on the target data: no transfer models and transfer models.

TABLE I
RESULTS OF DATASET 1

Source Domains	**similarity**	**Performance on target data**	
		No transfer	**Transfer**
Source 1	0.4041	1.3519±0.0000	0.2544 ± 0.0003
Source 2	0.4005	3.6346±0.0000	0.6006 ± 0.0065

TABLE II
RESULTS OF DATASET 2

Source Domains	**similarity**	**Performance on target data**	
		No transfer	**Transfer**
Source 1	0.4034	1.1981±0.0007	0.3611 ± 0.0007
Source 2	0.3989	4.3106±0.0001	0.5143 ± 0.0064
Source 3	0.3940	3.7372±0.0001	0.7076 ± 0.0195

From the results shown in Tables I and II, the performance of the models applying transferring technique is supervisor than the no transfer models, which validates the effectiveness of our transfer learning method. Comparing the similarity in the second column and the results in the last column, we find that the similarity determines the performance of the transfer learning, i.e. the higher the similarity of the source and target domains, the better of the models implementing knowledge transfer. For example, in Table II, "source 1" has the highest similarity with target domain than other two source domains, so the transfer model with "source 1" has the best results in solving the target tasks.

B. Multiple Source Domains with Similar Structure

In the last section, we get the conclusion that if the source domain has the similar data structure with the target domain, then the model will have a good transferring performance. Therefore, in this subsection, we generate multiple source domains, which all have the similar structures with the target domain, but with different distances. The input data of these source domain and the target domain are shown in Fig. 3.

From Fig. 3 we can see that, the data distributions in the three source domains and target domain are quite similar, but the similarity or distance between the source domains and the target domain are different. And The experimental results of implementing transfer learning from the multiple source domains to the target domain are shown in Table III. The results in second and fourth columns further verify that, the source domain with similar structure with target data will lead to a good transferring result.

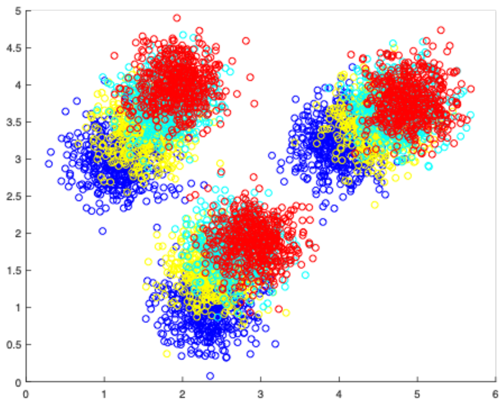

Fig. 3. Input data in multiple domains

TABLE III
RESULTS WITH MULTIPLE SOURCE DOMAINS WITH SIMILAR STRUCTURE

Source Domains	similarity	Performance on target data	
		No transfer	Transfer
Source 1	0.4576	2.1413±0.0000	0.3605 ± 0.0046
Source 2	0.4680	1.1080±0.0000	0.3144 ± 0.0101
Source 3	0.4848	0.3867±0.0000	0.3132 ± 0.0038

V. EXPERIMENTS ON REAL-WORLD DATASETS

In this section, datasets from the real world are used to validate the effectiveness of the proposed domain selection method. Since the studies on regression problems of domain adaptation are scarce, there are no public datasets in these scenarios. In this work, therefore, two datasets from UCI Machine Learning Repository are used and modified to simulate the multiple sources transfer learning scenarios. The way of modifying the datasets is crucial, so a detailed description to the datasets is provided.

The first dataset concerns air quality. We selected two of the existing attributes, temperature and relative humidity, as the input data and chose absolute humidity as the output. All the attributes were normalized, and the dataset was split into two domains based on relative humidity. Data with a relative humidity greater than 0.5 were chosen as the source domain, with 3600 instances, and the remaining data were used to form the target domain, with 1200 instances. Further, the first 2000 source instances are as source domain 1, and the remaining 1600 instances are as source domain 2. All the instances in the source domains are labeled, but only 10 instances in the target domain are labeled.

In the ”Condition based maintenance of naval propulsion plants dataset (CBM)”, fourteen features, such as ship speed, gas turbine shaft torque and so on, are used to predict the gas turbine decay state coefficient. The dataset was split based on the ship speed value. Data with a ship speed of greater than 10 knots formed the source domains, with 7500 instances, the remaining data, 3500 instances, was used for the target domain. Further, the first 4000 source instances are as source domain 1, and the remaining 3500 instances are as source domain 2. All the instances in the source domains are labeled, but only 10 instances in the target domain are labeled.

Although there is a step in the algorithm to find out the number of clusters in target data, it is not easy to determine in the real cases, especially in the high-dimension datasets. Therefore, the number of cluster is set as a hyper-parameter in the following experiments to explore the impact of it to the performance of transfer learning in fuzzy models.

With different numbers of clusters, domain selection and domain adaptation will be implemented. Tables IV and VI show the similarity between the source domains and target domain in two datasets, separately. And Tables V and VII display the transferring results.

TABLE IV
SIMILARITY OF SOURCE AND TARGET”

Number of cluster	Source 1	Source 2
2	0.4074	0.4264
3	0.4042	0.4144
4	0.4010	0.4105
5	0.3954	0.3996
6	0.3967	0.4002
7	0.3948	0.3998

TABLE V
RESULTS OF DATASET ”AIR QUALITY”

Number of clusters	No Transfer		Transfer	
	source 1	source 2	source 1	source 2
2	0.2186 ±0.0000	0.1052 ±0.0000	0.0917 ±0.0000	**0.0910 ±0.0001**
3	0.2176 ±0.0000	0.0984 ±0.0000	0.0922 ±0.0000	**0.0848 ±0.0000**
4	0.2106 ±0.0000	0.0986 ±0.0000	0.1053 ±0.0001	**0.0875 ±0.0000**
5	0.2140 ±0.0000	0.0957 ±0.0000	0.0917 ±0.0000	**0.0847 ±0.0000**
6	0.2098 ±0.0000	0.0949 ±0.0000	0.1044 ±0.0000	**0.0873 ±0.0000**
7	0.2110 ±0.0000	0.0951 ±0.0000	0.1063 ±0.0000	**0.0924 ±0.0000**

Analyzing the results in Tables IV and V, we can see that ”source 2” is closer to the target domain than ”source 1” with all the values of clusters, so the transferring results of using ”source 2” is the best in all the models.

TABLE VI
SIMILARITY OF SOURCE AND TARGET CBM”

Number of cluster	Source 1	Source 2
2	0.1981	0.1982
3	0.1925	0.1926
4	0.1844	0.1845
5	0.1807	0.1890
6	0.1850	0.1671
7	0.1563	0.1886

From the results in Tables IV and V, we can see that the results in Table VII is not always consistent with the similarity

TABLE VII
RESULTS OF DATASET "CBM"

Number of clusters	No Transfer		Transfer	
	source 1	source 2	source 1	source 2
2	12.6452 ±0.0400	11.5698 ±0.0118	**0.3528** ±**0.0040**	0.9865 ±0.4415
3	7.5414 ±4.6245	7.2834 ±0.4401	1.0210 ±1.0283	**0.4839** **0.0751**
4	6.8498 ±0.1799	7.0976 ±0.4232	1.8728 ±0.8535	**1.5630** ±**0.5046**
5	5.7624 ±1.7224	5.0380 ±2.0067	3.4981 ±17.1539	**1.4300** ±**0.2451**
6	5.8152 ±0.1625	5.1217 ±0.1689	**1.1194** ±**0.4887**	2.9732 ±5.7228
7	5.7530 ±0.1272	5.5072 ±0.1799	**2.4206** ±**4.2064**	2.6721 ±2.8046

in Table VI. This may due to the high dimension of the dataset, which makes the similarity calculation imprecise. This also shows the limitation of the proposed domain selection method.

VI. CONCLUSION AND FUTURE STUDY

This work explores the transfer learning problems when multiple source domains are available. A similarity is defined between domains to help select an appropriate source domain from multiple sources for the target domain. And based on the seleced source domain, fuzzy rules are modifies and transferred to support target domain in solving regression tasks. The experiments on both synthetic and real-world datasets validate the effectiveness of the proposed domain selection method in transfer learning.

The domain selection method presented in this paper is a preliminary attempt in dealing with transfer learning in multiple source domains situation. The limitation of the proposed method, imprecise selection result in handling high-dimensional dataset, will be resolved in the future studies.

ACKNOWLEDGMENT

This work was supported by Australian Research Council under DP 170101623.

REFERENCES

[1] J. Heaton, N. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 1, pp. 3–12, 2017.

[2] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Molecular Systems Biology*, vol. 12, no. 7, p. 878, 2016.

[3] S. Hoo-Chang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, p. 1285, 2016.

[4] A. Ioannidou, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris, "Deep learning advances in computer vision with 3d data: A survey," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, p. 20, 2017.

[5] S. J. Pan, Q. Yang *et al.*, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[6] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge and Information Systems*, vol. 36, no. 3, pp. 537–556, 2013.

[7] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *Journal of Machine Learning Research*, vol. 10, no. Jul, pp. 1633–1685, 2009.

[8] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: a survey," *Knowledge-Based Systems*, vol. 80, pp. 14–23, 2015.

[9] J. Shell and S. Coupland, "Fuzzy transfer learning: methodology and application," *Information Sciences*, vol. 293, pp. 59–79, 2015.

[10] Z. Deng, Y. Jiang, H. Ishibuchi, K.-S. Choi, and S. Wang, "Enhanced knowledge-leverage-based tsk fuzzy system modeling for inductive transfer learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, p. 11, 2016.

[11] V. Behbood, J. Lu, G. Zhang, and W. Pedrycz, "Multistep fuzzy bridged refinement domain adaptation algorithm and its application to bank failure prediction." *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 1917–1935, 2015.

[12] F. Liu, J. Lu, and G. Zhang, "Unsupervised heterogeneous domain adaptation via shared fuzzy equivalence relations," *IEEE Transactions on Fuzzy Systems*, 2018.

[13] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Granular fuzzy regression domain adaptation in takagi–sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 847–858, 2018.

[14] H. Zuo, J. Lu, G. Zhang, and W. Pedrycz, "Fuzzy rule-based domain adaptation in homogeneous and heterogeneous spaces," *IEEE Transactions on Fuzzy Systems*, 2018.

[15] M. L. Hadjili and V. Wertz, "Takagi-sugeno fuzzy modeling incorporating input variables selection," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 6, pp. 728–742, 2002.

[16] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Fuzzy regression transfer learning in takagi–sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1795–1807, 2017.

[17] C. E. Rasmussen, "The infinite gaussian mixture model," in *Advances in Neural Information Processing Systems*, 2000, pp. 554–560.

[18] H. Zuo, J. Lu, G. Zhang, and F. Liu, "Fuzzy transfer learning using an infinite gaussian mixture model and active learning," *IEEE Transactions on Fuzzy Systems*, 2018.