

“© 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.”

Fuzzy Rule-based Transfer Learning for Label Space Adaptation

Hua Zuo, Guangquan Zhang, Jie Lu

Decision Systems & e-Service Intelligence Lab
Center of Artificial Intelligence
Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, Australia

Witold Pedrycz

Department of Electrical & Computing Engineering
University of Alberta
Edmonton, Canada

Abstract — As the age of big data approaches, methods of massive scale data management are rapidly evolving. The traditional machine learning methods can no longer satisfy the exponential development of big data; there is a common assumption in these data-driving methods that the distribution of both the training data and testing data should be equivalent. A model built using today's data will not adequately address the classification tasks tomorrow if the distribution of the data item values has changed. Transfer learning is emerging as a solution to this issue, and many methods have been proposed. Few of the existing methods, however, explicitly indicate the solution to the case where the labels' distributions in two domains are different. This work proposes the fuzzy rule-based methods to deal with transfer learning problems where the discrepancy between the two domains shows in the label spaces. The presented methods are validated in both the synthetic and real-world datasets, and the experimental results verify the effectiveness of the introduced methods.

Keywords — *machine learning, fuzzy rules, transfer learning, classification*

I. INTRODUCTION

With the development of the internet, the application of information technology has infiltrated our daily life; we charge water and electricity bills to our credit card, book hotels online, and invest in the stock market through computer networks. A massive amount of data is generated every day, and the rapid evolution of data has left traditional machine learning methods far behind. There is always a common assumption when applying machine learning techniques: the distribution of the training data (source domain) must be identical with that of the testing data (target domain). This assumption means that models trained using training data will not perform well on the testing data if the distribution of testing data changes. This severely impedes the development of machine learning methods in rapidly changing environments. For instance, if a data model is built for a specific time period, but data distribution changes over time, the model will be left with fewer labeled data items. Although a massive amount of unlabeled data items will still be available, labeling them manually is time consuming, yet

discarding the original model and its resulting information would be wasteful.

Transfer learning has emerged as a solution for the above research issue. Transfer learning is a machine learning method that uses previously accumulated knowledge to help solve current tasks. Many transfer learning methods have been proposed to improve the performance of knowledge transfer across domains, and there are different classifications for these methods. One well-known way of dividing the transfer learning methods was introduced in [1], where the methods are classified into four categories: multi-task learning [2], domain adaptation [3], cross-domain adaptation [4] and heterogeneous learning [5]. In addition, the transfer learning methods are divided according to soft techniques [6]: neural-network-based transfer learning [7], Bayesian transfer learning [8] and fuzzy transfer learning [9].

Fuzzy-based methods have been proposed in recent years because of their powerful ability to deal with imprecise and vague information. Behbood et al. [10] uses a fuzzy neural network and applies a multistep fuzzy bridge refinement strategy to successfully solve bank failure prediction problem. Further, for cross-domain prediction problems, they proposed a method that implements fuzzy spectral feature alignment, based on fuzzy similarity, to help knowledge transfer between domains. Shell and Coupland [11,12] introduced a transfer learning framework based on a fuzzy inference system, which provides guiding significance when transferring knowledge across domains. In a practical application to detect epilepsy using an electroencephalogram, Yang et al. [13] proposed a transductive transfer learning method based on the Takagi-Sugeno-Kang fuzzy logic system for recognizing epileptic electroencephalogram signals.

Although many methods have been proposed to handle transfer learning issues, there are few studies that deal explicitly with cases where the label spaces in two domains differ. In this paper, we concentrate on a specific scenario – domain adaptation for classification tasks – where there is

an obvious discrepancy between the labels' distributions across two domains. The main purpose of this paper is the proposal of fuzzy rule-based approaches to address the knowledge transfer issue in two domains where the distributions of the label spaces are quite different.

The paper is structured as follows. In Section II, some related definitions about transfer learning, and the Takagi-Sugeno fuzzy model are introduced. Section III describes the research issues and the proposed methods. The experimental results, including synthetic datasets and a real-word dataset, are presented in Section IV. Finally, Section V concludes the paper and outlines future work.

II. PRELIMINARY

This section gives some definitions of transfer learning and outlines the basic learning model applied in this paper: the Takagi-Sugeno fuzzy model.

A. Definitions

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

Definition 2 (Task) [1]: A task is denoted by $T = \{Y, f(\cdot)\}$, where $Y = \{y_1, \dots, y_m\}$ is a label space, and $f(\cdot)$ is an objective predictive function.

Definition 3 (Transfer Learning) [1]: Given a source domain D_s and learning task T_s , a target domain D_t and 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 and T_s where $D_s \neq D_t$ or $T_s \neq T_t$.

Definition 4 (Domain Adaptation) [1]: Domain adaptation is a category of inductive transfer learning in which $F_t = F_s$, but $P_t(X) \neq P_s(X)$.

B. The Takagi-Sugeno Fuzzy Model

A Takagi-Sugeno fuzzy model is represented in the forms of fuzzy rules [14]:

$$\text{if } \mathbf{x} \text{ is } A_i(\mathbf{x}, \mathbf{v}_i), \text{ then } y \text{ is } L_i(\mathbf{x}, \mathbf{a}_i) \quad i = 1, \dots, c \quad (1)$$

Each fuzzy rule contains two components: one condition $A_i(\mathbf{x}, \mathbf{v}_i) = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x} - \mathbf{v}_j\|}{\|\mathbf{x} - \mathbf{v}_i\|} \right)^{\frac{2}{m-1}}}$, where \mathbf{x} is the input variable

and \mathbf{v}_i is the center of the clusters. $L_i(\mathbf{x}, \mathbf{a}_i) = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n$ is one conclusion, which is determined by the linear function of the input variable \mathbf{x} .

The construction of a Takagi-Sugeno fuzzy model includes two steps: forming the conditions A_1, \dots, A_c through fuzzy clustering, and optimizing the coefficient of the linear functions $L_i(\mathbf{x}, \mathbf{a}_i)$.

III. PROBLEM STATEMENT AND METHODS

In this section, we first identify the transfer learning problem, and then propose the fuzzy rule-based methods to resolve knowledge transfer issues in the scenario.

A. Problem Statement

A well-performing model can be built for the classification tasks in the source domain when plenty of labeled data is available. Let the source data be $\mathcal{S} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, where $\mathbf{x}_k \in R^n$, $y_k \in \{0, 1\}$, $k = 1, \dots, N$. The constructed Takagi-Sugeno fuzzy model is:

$$\text{if } \mathbf{x}_k \text{ is } A_i(\mathbf{x}_k, \mathbf{v}_i), \text{ then } y_k \text{ is } L_i(\mathbf{x}_k, \mathbf{a}_i) \quad i = 1, \dots, c \quad (2)$$

Therefore, when the input data is \mathbf{x}_k , the output can be calculated as

$$y_k = \sum_{i=1}^c A_i(\mathbf{x}_k, \mathbf{v}_i) L_i(\mathbf{x}_k, \mathbf{a}_i) \quad (3)$$

To clearly demonstrate the architecture of this fuzzy rule-based model, it also could be shown as a network, which is more concise and visible. Fig. 2 shows the structure of the constructed model using two input variables and three rules as an example.

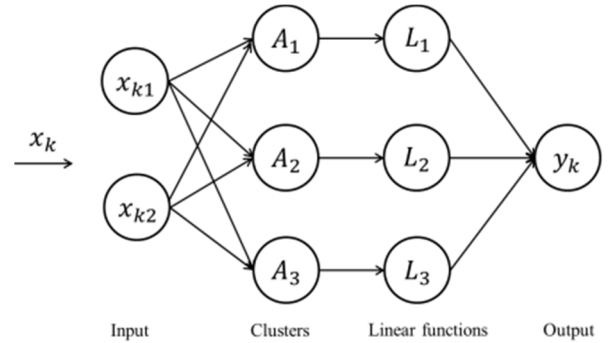


Fig.1 Architecture of the Takagi-Sugeno Fuzzy Model

We also have the target dataset $\mathcal{H} = \{\mathcal{H}_L, \mathcal{H}_U\} = \{(\mathbf{x}'_1, y'_1), \dots, (\mathbf{x}'_{N'_t}, y'_{N'_t}), \mathbf{x}'_{N'_t+1}, \dots, \mathbf{x}'_{N'}\}$, where \mathcal{H}_L is the labeled dataset and \mathcal{H}_U is the unlabeled dataset. A supposed ideal model for the target domain is

$$\text{if } \mathbf{x}'_k \text{ is } A_i(\mathbf{x}'_k, \mathbf{v}_i), \text{ then } y'_k \text{ is } L_i(\mathbf{x}'_k, \mathbf{a}_i) \quad i = 1, \dots, c \quad (4)$$

Thus the output can be calculated by

$$y'_k = \sum_{i=1}^c A_i(\mathbf{x}'_k, \mathbf{v}_i) L_i(\mathbf{x}'_k, \mathbf{a}_i) \quad (5)$$

Since the number of labeled data in \mathcal{H}_L is insufficient, the ideal model (fuzzy rules) (5) for the target domain can't be derived. However, because the data in the source and target domains are not from identical distributions, the model built using the source data cannot be applied to solve the classification tasks in the target domain.

Based on the discrepancy between the source data and target data, two different scenarios are considered. In the first scenario, the input data in two domains come from the same input distributions, i.e., datasets $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and $\{\mathbf{x}'_1, \dots, \mathbf{x}'_{N'}\}$ follow the identical statistical distributions, but there is a gap between the labels' distributions across two domains.

The first scenario is a special case in transfer learning, thus we discuss a more generalized case in the second scenario, where the distributions of the input and output variables are both different in the source and target domains. Figs. 2-4 give two examples to illustrate the two scenarios.

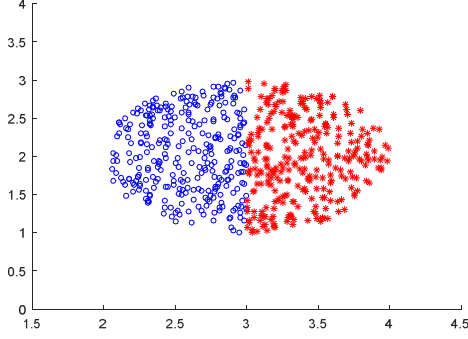


Fig.2 Dataset 1

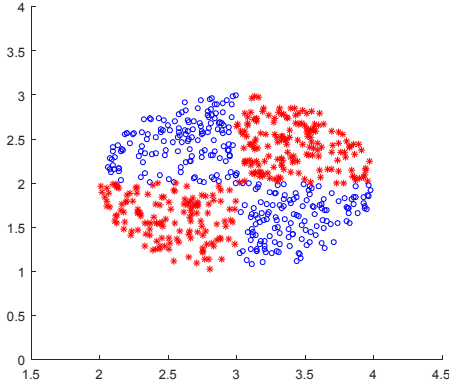


Fig.3 Dataset 2

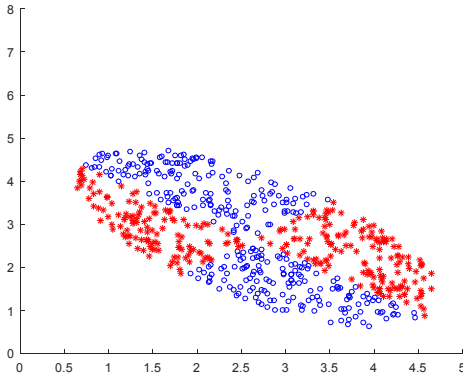


Fig.4 Dataset 3

Figs 2-4 give three datasets with labels expressed by different colors and marker types, blue circle and red asterisk. Comparing the datasets in Fig. 2 (source data) and Fig. 3 (target data), we find that they follow the same input distributions, but the labels' distributions are not identical. To illustrate the second scenario, the datasets displayed in Fig. 2 (source data) and Fig. 4 (target data) are compared, and it is observed that the gap between the source data and

target data comes from two aspects: the different distributions of the input variables and changes in the label spaces.

B. Knowledge Transfer in Fuzzy Rule-based Models

Two methods are proposed to deal with the knowledge transfer problems in the two scenarios mentioned above. These methods are intended to adjust the existing model of the source domain to make it more compatible with the classification tasks in the target domain.

In the first scenario, the discrepancy between the source data and target data is the distribution of the label space. Through the construction procedures of the Takagi-Sugeno fuzzy model, we know that the centers of the clusters (i.e., the conditions) are dominated by the input variables' distributions, and the concluding sections of the fuzzy rules are governed by the input variables and coefficients of the linear functions. The input variables of source data and target data are derived from the same distribution, which indicates that the source and target data could almost lead to clusters with the same centers. Therefore, we aim to change the linear functions of the existing model so that the modified fuzzy rules are suitable for the target data.

We change the conclusion of the fuzzy rules in a nonlinear way, and show the consequential structure of the modified fuzzy model in Fig. 5. This figure also follows the example in Fig. 1.

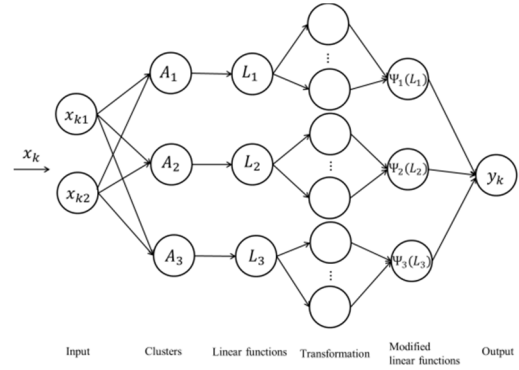


Fig. 5. Modified fuzzy model for the first scenario

When the input is \mathbf{x}'_k , the output \mathbf{t}'_k can be calculated by

$$\mathbf{t}'_k = \sum_{i=1}^c A_i(\mathbf{x}'_k, \mathbf{v}_i) \Psi_i(L_i(\mathbf{x}'_k, \mathbf{a}_i)) \quad (6)$$

where \mathbf{v}_i and \mathbf{a}_i are the centres of the clusters and coefficients of the linear functions in the source domain's model. Ψ_i is the mapping that transforms the linear function L_i .

$$S_1 = \frac{1}{N'_t} \sqrt{\sum_{k=1}^{N'_t} (\mathbf{t}'_k - \mathbf{y}'_k)^2} \quad (7)$$

In more generalized situations, we combine the approach proposed in our previous paper [15] (changing the input space) with an approach that changes the linear functions to deal with cases where both the distributions of input and label spaces are not matched in the source and target domains. Fig. 6 gives the structure of the model using the combined approach. The mappings that change the input and label spaces consist of sigmoid functions in a similar way and the parameters of the sigmoid functions are learned by maximizing this model's performance on the labeled target data.

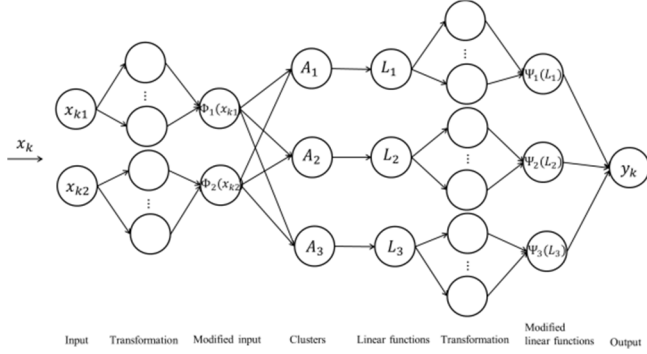


Fig. 6 Modified fuzzy model for the second scenario

When the input is \mathbf{x}'_k , the output \mathbf{h}'_k can be calculated by

$$\mathbf{h}'_k = \sum_{i=1}^c A_i(\Phi(\mathbf{x}'_k), \mathbf{v}_i) \Psi_i(L_i(\Phi(\mathbf{x}'_k), \mathbf{a}_i)) \quad (8)$$

where $\Phi = [\Phi_1 \ \Phi_2]$, and Φ_1 and Φ_2 are the transformation functions for the input variables x'_{k1} and x'_{k2} , respectively. The parameters of $\Phi = [\Phi_1 \ \Phi_2]$, and $\Psi = [\Psi_1 \ \Psi_2 \ \Psi_3]$ are determined by minimizing the objective function below.

$$S_2 = \frac{1}{N'_t} \sqrt{\sum_{k=1}^{N'_t} (\mathbf{h}'_k - \mathbf{y}'_k)^2} \quad (9)$$

IV. EXPERIMENTS

Both synthetic and real-world datasets were used to evaluate the proposed methods.

Prior to showing the experimental results, the indexes of the models' performance are given. The accuracy of source model on the source data is shown as Q , $Q1$ represents the accuracy of source model on the target data \mathbf{H}_U , $Q2$ is the accuracy of the target model trained with few labeled target data on the target data \mathbf{H}_U , and $Q3$ denotes the accuracy of the modified source model based on the proposed method on the target data \mathbf{H}_U . Please note that the construction of the target model using only a few labeled target data items is intended to show that the model's performance using less training data cannot be guaranteed.

A. Synthetic datasets

In this subsection, the two experiments using synthetic datasets were designed to validate the effectiveness of the presented methods for the two different test cases in Section III.

Experiment 1:

This experiment concentrates on the first case where the difference in source and target data is only evident in the label's distributions, thus we apply the first method and the corresponding model shown in Fig. 5.

The source and target data are displayed in Figs. 7 and 8, and are generated following the identical distribution. The data labeled '0' are represented with blue circle and the data labeled '1' are shown as red asterisk. The distributions of the labels are arranged quite differently in the two domains.

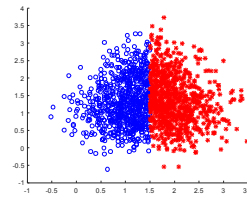


Fig. 7 Source data

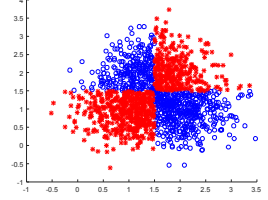


Fig. 8 Target data

In addition, the number of clusters, i.e., the number of fuzzy rules, is an important parameter that is difficult to determine and highly dependent on the dataset. In this case, we adopt different numbers of clusters and compare the results to find the optimal. All the models' construction applies five-fold cross-validation, and the performance of these models and the results of Q , $Q1$, $Q2$ and $Q3$ are shown in Table I and Table II.

TABLE I. RESULTS WITH 3 CLUSTERS (FUZZY RULES) (%)

	Q	$Q1$	$Q2$	$Q3$
1	99.67	39.46	81.62	84.26
2	99.00	39.66	81.96	81.35
3	98.67	38.99	70.54	76.35
4	99.00	39.59	71.89	90.88
5	99.33	39.32	53.04	82.57
mean	99.13	39.41	71.81	83.08
stand deviation	0.38	0.27	11.76	5.26

TABLE II. RESULTS WITH 4 CLUSTERS (FUZZY RULES) (%)

	Q	$Q1$	$Q2$	$Q3$
1	99.00	39.73	68.51	75.68
2	95.33	37.77	58.04	82.16
3	95.00	37.50	59.46	76.42
4	99.33	39.39	66.22	75.07
5	95.00	39.05	50.20	74.59
mean	96.73	38.69	60.49	76.78
stand deviation	2.23	1.00	7.25	3.08

From the results, we can see that the mean value of Q reached 99.13% and 96.73%, and the standard deviation was very small, which means a good model for the source

domain has been achieved. However, the mean value for $Q1$ was also small; 39.41% for 3 clusters and 38.69% for 4 clusters. This indicates that prediction tasks in the target domain cannot be solved by the source domain's existing model. The mean value for $Q2$ was also not very high, and this verifies the assumption of this work: that models constructed with little training data have a high probability of performing badly. Yet, the mean values for $Q3$ were always greater than the mean values for $Q1$ and $Q2$. This indicates that the model trained using the proposed method is superior, not only to the source model but also to the target model trained with few labeled target data.

We noticed that the standard deviations for $Q2$ and $Q3$ were not small; 11.76% and 5.26% with 3 clusters and 7.25% and 3.08% with 4 clusters. This phenomenon is reasonable and expected, because few labeled target data were available to train the new model – even less was available for the testing data. Obviously, the standard deviation will decline if more labeled target data are available, but this is not consistent with the assumption that the labeled target data is insufficient to build a well-performed model for prediction tasks in the target domain.

Experiment 2:

In the second experiment, we deal with situations where the source data and target data are different in both the input and labels' distributions.

The source data and target data are shown in Figs. 9 and 10. Similarly, data with disparate labels are represented by points in contrasting colors and marker types, blue circles and red asterisk. The source data shows the same pattern as that of the first experiment, but the target data is quite contrasted.

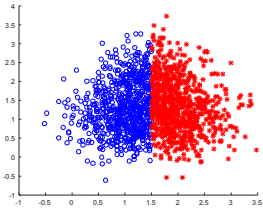


Fig. 9 Source data

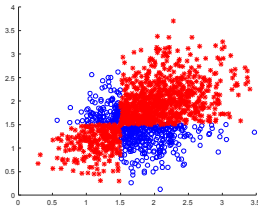


Fig.10 Target data

Likewise, five-fold cross-validation was applied to build the models. The experimental results are shown in Tables III and IV.

Similarly, the results show that the proposed method improves the performance of the existing source model when classifying the data in the target domain. The standard deviation for $Q3$ was not small; 4.77% with 3 clusters and 11.28% with 4 clusters. As with Experiment 1, the number of labeled target data was too low. Another possible reason is the large gap between the source data and target data.

TABLE III. RESULTS WITH 3 CLUSTERS (FUZZY RULES) (%)

	Q	$Q1$	$Q2$	$Q3$
1	99.67	62.30	82.77	79.46
2	99.00	62.70	55.41	88.38
3	98.67	62.30	69.26	90.14
4	99.00	62.64	77.36	81.35
5	99.33	62.09	72.03	81.42
mean	99.13	62.41	71.36	84.15
stand deviation	0.38	0.26	10.32	4.77

TABLE IV. RESULTS WITH 4 CLUSTERS (FUZZY RULES) (%)

	Q	$Q1$	$Q2$	$Q3$
1	99.00	62.64	81.22	68.45
2	95.33	61.28	45.81	74.12
3	95.00	58.85	57.50	85.74
4	99.33	62.50	85.00	86.35
5	95.00	59.86	84.19	60.14
mean	96.73	61.03	70.74	74.96
stand deviation	2.23	1.65	17.96	11.28

B. Real-world dataset

A public dataset, called the banknote authentication dataset, from the Machine Learning Repository was used to verify the effectiveness of the proposed methods. The aim of this dataset is to distinguish between genuine and forged 'banknote-like' specimens. The data were obtained by extracting the features of the images of the genuine and forged banknotes using a wavelet transform tool.

The features of this dataset are shown in Table V.

TABLE V. TABLE TYPE STYLES

Feature 1	Variance of wavelet transformed image
Feature 2	Skewness of wavelet transformed image
Feature 3	Curtosis of wavelet transformed image
Feature 4	Entropy of wavelet transformed image

All the features were normalized, and the dataset was split into two datasets based on the "skewness of wavelet transformed image" feature. The data with this feature value larger than 0.5 fell within the source domain; the remaining data were allocated to the target domain. In total, the source domain contained 960 labeled data items in the source domain, but the target domain only contained 30 labeled data items and 320 unlabeled items. Since the input distributions in two domains differ, the method in Fig.6 is applied to solve this problem. Five-fold cross-validation was used to construct all the models. The experimental results are shown in Table VI.

TABLE VI. RESULTS OF BANKNOTE AUTHENTICATION (%)

	Q	Q1	Q2	Q3
1	94.79	52.42	55.15	83.64
2	96.35	53.03	40.91	87.88
3	95.83	53.33	68.48	85.45
4	96.88	53.03	69.09	63.94
5	94.27	51.52	53.03	74.24
mean	95.63	52.67	57.33	79.03
stand deviation	1.08	0.72	11.79	9.89

In analyzing the results shown in Table VI, it is clear that the mean value for $Q2$ was only 52.67%, which indicates the model built using previous data is invalid. The mean value for $Q2$ was also small, at 57.33%, and even worse, the standard deviation was not small, at 11.79%. The low accuracy and high standard deviation for $Q2$ is testament to our assumption in this work: that a small amount of labeled data cannot guarantee the performance of the constructed models. The performance of the proposed method is reflected in the values for $Q3$, which show a mean value of 79.03% and a standard deviation of 9.89%. The mean value for $Q3$ was not greater than that of $Q1$ and $Q2$. This suggests the proposed model is superior to both the previous model and the model built only with target data.

V. CONCLUSION AND FUTURE WORK

This paper proposes fuzzy rule-based methods to deal with special scenarios in domain adaptation where label distributions in two domains are different. The results from experiments using both synthetic and real-world datasets verify that the presented methods improve the performance of the source model in solving prediction tasks in the target domain.

This work focuses solely on domain adaptation problems, where the dimensionality of the input space in the source and target domain is the same. A more challenging problem – cross-domain adaptation – will be explored in a future work.

ACKNOWLEDGMENT

The work presented in this paper was supported by the Australian Research Council (ARC) under Discovery Grant DP140101366.

REFERENCES

- [1] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345-1359, 2010.
- [2] Y. Yan, E. Ricci, R. Subramanian, G. Liu, O. Lanz, and N. Sebe, "A multi-task learning framework for head pose estimation under target motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 1070-1083, 2016.
- [3] H. Zuo, G. Zhang, V. Behbood and J. Lu, "Feature spaces-based transfer learning", IFSA World Congress and EUSFLAT Annual Meeting (IFSA-EUSFLAT), 2015.
- [4] D. Bollegala, T. Mu, and J. Y. Goulermas, "Cross-domain sentiment classification using sentiment sensitive embeddings," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, pp. 398-410, 2016.
- [5] P. Yang, H. Davulcu, Y. Zhu, and J. He, "A generalized hierarchical multi-latent space model for heterogeneous learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, pp. 3154-3168, 2016.
- [6] 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.
- [7] Y. Huang, M. Hu, X. Yu, T. Wang, and C. Yang, "Transfer learning of deep neural network for speech emotion recognition," in *Chinese Conference on Pattern Recognition*, 2016, pp. 721-729.
- [8] T. T. Joy, S. Rana, S. K. Gupta, and S. Venkatesh, "Flexible transfer learning framework for Bayesian optimisation," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2016, pp. 102-114.
- [9] H. Zuo, G. Zhang, V. Behbood, J. Lu, W. Pedrycz, and T. Zhang, "Fuzzy transfer learning in data-shortage and rapidly changing environments," in *Uncertainty Modelling in Knowledge Engineering and Decision Making: Proceedings of the 12th International FLINS Conference (FLINS 2016)*, 2016, p. 175.
- [10] 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, pp. 1917-1935, 2015.
- [11] J. Shell and S. Coupland, "Towards fuzzy transfer learning for intelligent environments," in *International Joint Conference on Ambient Intelligence*, 2012, pp. 145-160.
- [12] J. Shell and S. Coupland, "Fuzzy transfer learning: methodology and application," *Information Sciences*, vol. 293, pp. 59-79, 2015.
- [13] C. Yang, Z. Deng, K.-S. Choi, and S. Wang, "Takagi-Sugeno-Kang transfer learning fuzzy logic system for the adaptive recognition of epileptic electroencephalogram signals," *IEEE Transactions on Fuzzy Systems*, 2016.
- [14] M. L. Hadjili and V. Wertz, "Takagi-Sugeno fuzzy modeling incorporating input variables selection," *IEEE Transactions on fuzzy systems*, vol. 10, pp. 728-742, 2002.
- [15] 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*, 2016.