A Hybrid Transfer Learning Method for Intelligent Fault Diagnosis of Machines under Variable Operation Conditions

Zhaojun Du, Bin Yang, Yaguo Lei*, Xiwei Li, Naipeng Li
Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University
Xi'an, Shaanxi, 710049, China
* yaguolei@mail.xjtu.edu.cn

***Abstract*—Intelligent fault diagnosis has been a research hotspot in recent years. However, most of the works are conducted based on the hypothesis that training and testing data subject to the same distribution. In engineering scenarios, machines usually work under variable operation conditions, which results in the data from different conditions subject to distribution discrepancy. Since transfer learning is able to reuse the related knowledge across different domains, a hybrid transfer learning method (HTLM) is proposed to utilize the diagnosis knowledge obtained from one operation condition to complete the diagnosis tasks under other conditions. In the method, transfer component analysis is firstly used to extract fault features with small distribution discrepancy from the cross-domain samples. After that, the features learned from the source domain help train a classifier by Tradaboost algorithm to improve its diagnosis accuracy on the target domain. The effectiveness of the proposed method is verified by a set of laboratory bearing datasets, in which the data under one operation condition are used to help identify the health states of bearings under another condition. The results indicate that the proposed HTLM is able to achieve a higher diagnosis accuracy than other diagnosis methods.**

Keywords—intelligent fault diagnosis; machines; transfer learning; transfer component analysis; Tradaboost.

I. INTRODUCTION

Recently, with the rapid development of machine learning, intelligent fault diagnosis of machines has made substantial achievements [1-3]. In most of these works, there is a basic hypothesis that the training and testing data have the same distribution. However, this hypothesis can hardly be guaranteed in engineering scenarios because machines usually work under variable conditions. With the change of motor load and speed, the collected data will exhibit different vibration amplitude and frequency characteristics [4], which makes the collected data have distribution discrepancy. Thus, it is difficult for the diagnosis model trained under variable conditions to achieve high accuracy. As far as we are aware, investigating on fault diagnosis methods under variable operation conditions are faced with two main problems: (1) Lack of sufficient labeled training data; (2) Existence of data distribution discrepancy among different operation conditions [5]. Plenty of monitoring data are collected during the operation of machines, but only few of them can be used as labeled data to train the diagnosis model, while labeling data manually is time-consuming and labor-intensive for practical use; Moreover, due to the distribution discrepancy, a diagnosis model trained with the data under one operation condition is unable to achieve high accuracy in identifying the health states under other operation conditions, which weakens the generalization of the trained diagnosis model.

Transfer learning (TL) [6] is introduced to work out these problems. There are two datasets analyzed in TL tasks. One represents the source domain, and the other represents the target domains [7]. Source domain data collected from one operation condition is used to provide prior diagnosis knowledge, and target domain data are collected under another condition, in which the health states of the unlabeled data need to be identified. Because of the different operation conditions, there is distribution discrepancy between the collected source and target domain data [8]. A diagnosis model is trained with the two-domain data and TL methods for the health states identification via the target domain data. By using TL methods, the distribution discrepancy problem in extracting fault features can be tackled, and the performance [9-10] of the diagnosis model can be substantially increased in comparison with traditional diagnosis methods. In recent years, researchers have got some achievements in machine fault diagnosis field. For instance, Shen et al. [11] applied Singular Value Decomposition and Tradaboost algorithm to identify the faults of rolling bearings under various conditions. Pan et al. [12] extracted fault features by Deep Belief Network, and utilized Tradaboost to train the classification model, which is applied to the diagnosis of high-voltage circuit breaker. Xie et al. [13] utilized transfer component analysis (TCA) to extracted the low-dimensional features subject to similar distribution, and applied it to the diagnosis of gearbox under variable operation conditions. From the literature review, Tradaboost is essentially a algorithm to train classifier, and TCA is a single-layer feature extraction. There is a problem that both of them cannot achieve high diagnosis accuracy while importing the raw vibration signals into the diagnosis model directly, since the two methods separate the procedures of feature extraction and classifier training. Therefore, it is essential to propose a TL-based method for fault diagnosis which the entire training process is based on TL ideas.

This work is partly supported by National Nature Science Foundation of China under Grant No. 61673311 and NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization No. U1709208.

To bridge the above problem, a hybrid transfer learning method (HTLM) is proposed in this paper for fault diagnosis of machines under variable operation conditions. In this method, TCA is firstly employed to extract the low dimensional features of data in the source and target domains, in which the extracted data are subject to similar distribution. Next, several weak classifiers are trained and the ultimate strong classifier is integrated based on Tradaboost algorithm. The classification of the extracted target domain features is finally utilized as the diagnosis result. The proposed HTLM is validated by the bearing datasets acquired from a multi-stage transmission test rig. The comparison results of the different diagnosis methods show that the proposed method could be used in fault diagnosis under variable operation conditions, and the diagnosis is able to achieve a higher accuracy.

II. THE PROPOSED DIAGNOSIS METHOD

A. Diagnosis Tasks based on TL

Let $\mathcal{D}^{\mathrm{S}} = \{\boldsymbol{X}_{\mathrm{S}}, P(\boldsymbol{X}_{\mathrm{S}})\}$ represents the source domain and $\mathcal{D}^{\mathrm{T}} = \{\boldsymbol{X}_{\mathrm{T}}, P(\boldsymbol{X}_{\mathrm{T}})\}$ represents the target domain. $\boldsymbol{X}_{\mathrm{S}} = \{(\boldsymbol{x}_1^{\mathrm{S}}, y_1^{\mathrm{S}}), \ldots, (\boldsymbol{x}_m^{\mathrm{S}}, y_m^{\mathrm{S}})\}$ is the labeled data collected under one operation condition (OC1), and $\boldsymbol{X}_{\mathrm{T}} = \{(\boldsymbol{x}_1^{\mathrm{T}}, y_1^{\mathrm{T}}), \ldots, (\boldsymbol{x}_l^{\mathrm{T}}, y_l^{\mathrm{T}}), \boldsymbol{x}_{l+1}^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}} | l \le n\}$ is the partially labeled data collected under another operation condition (OC2), in which the health states of unlabeled data of $\boldsymbol{X}_{\mathrm{T}}$ need to be identified. The samples of the source and target domains respectively has the marginal distribution of $P(\boldsymbol{X}_{\mathrm{S}})$ and $P(\boldsymbol{X}_{\mathrm{T}})$ which $P(\boldsymbol{X}_{\mathrm{S}}) \ne P(\boldsymbol{X}_{\mathrm{T}})$.

Assume that there is a set of data $\boldsymbol{X}_{\mathrm{T}}$ collected under OC2, in which plenty of the samples in $\boldsymbol{X}_{\mathrm{T}}$ are unlabeled. Due to lack of labeled data, the classifier $h_{\mathrm{T}}(\cdot)$ trained with $\boldsymbol{X}_{\mathrm{T}}$ is hard to achieve high accuracy. Since the labeled source domain data $\boldsymbol{X}_{\mathrm{S}}$ from OC1 have similar fault information, these data could be utilized to help train the classifier. During the training process, however, it is necessary to make $\boldsymbol{X}_{\mathrm{S}}$ and $\boldsymbol{X}_{\mathrm{T}}$ have less distribution discrepancy. In this TL-based diagnosis task, a diagnosis model needs to be constructed, in which the model is able to extract cross-domain features subject to similar distribution from the source and the target domains. Moreover, the classifier $h_f(\cdot)$ trained with the extracted features should be capable of effectively identifying the health states of unlabeled target domain samples $\{\boldsymbol{x}_{l+1}^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}}\}$.

B. The proposed HTLM

This section details the proposed diagnosis method, and the training process is illustrated as follows. The diagnosis process of HTLM is shown in Fig. 1.

1) Data Collection. The labeled data collected under OC1 are treated as source domain datasets $\boldsymbol{X}_{\mathrm{S}} = \{(\boldsymbol{x}_1^{\mathrm{S}}, y_1^{\mathrm{S}}), \ldots, (\boldsymbol{x}_m^{\mathrm{S}}, y_m^{\mathrm{S}})\}$. And the partially labeled data collected under OC2 are regarded as target domain datasets $\boldsymbol{X}_{\mathrm{T}} = \{(\boldsymbol{x}_1^{\mathrm{T}}, y_1^{\mathrm{T}}), \ldots, (\boldsymbol{x}_l^{\mathrm{T}}, y_l^{\mathrm{T}}), \boldsymbol{x}_{l+1}^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}} | l \le n\}$, in which $\{y_1^{\mathrm{T}}, \ldots, y_l^{\mathrm{T}}\}$ are the few labels of target domain samples.

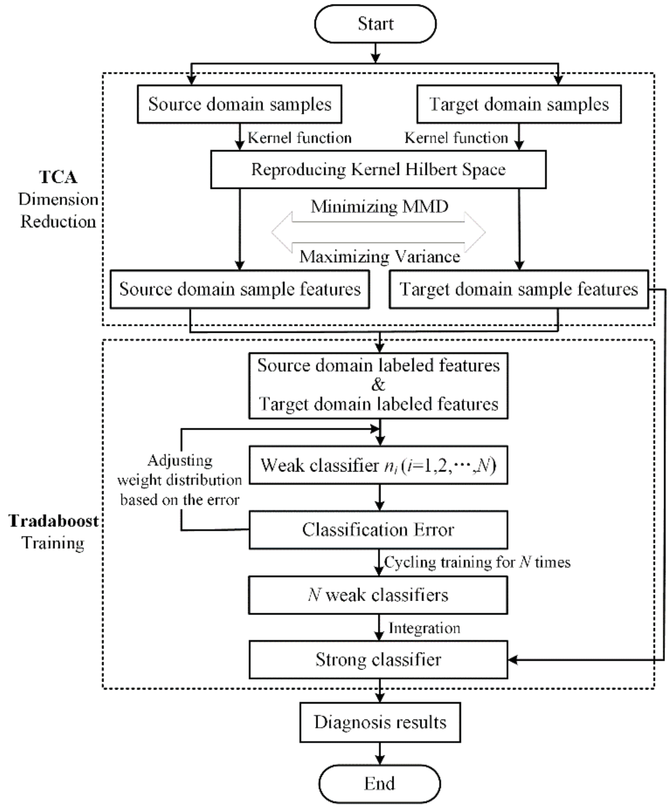

Fig. 1. The diagnosis process of the proposed method.

2) Fault Feature Extraction. The collected samples $\boldsymbol{X}_{\mathrm{S}} = \{\boldsymbol{x}_1^{\mathrm{S}}, \boldsymbol{x}_2^{\mathrm{S}}, \ldots, \boldsymbol{x}_m^{\mathrm{S}}\}$ and $\boldsymbol{X}_{\mathrm{T}} = \{\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \ldots, \boldsymbol{x}_n^{\mathrm{T}}\}$ are firstly mapped into Reproducing Kernel Hilbert Space (RKHS) [14] through Gaussian kernel function $\Phi(\cdot)$. Next, the maximum mean discrepancy (MMD) [15] between $\boldsymbol{X}_{\mathrm{S}}$ and $\boldsymbol{X}_{\mathrm{T}}$ is minimized, while MMD metric is defined by

$$Dist(\boldsymbol{X}_{\mathrm{S}}, \boldsymbol{X}_{\mathrm{T}}) = \left\| \frac{1}{m} \sum_{i=1}^{m} \Phi(\boldsymbol{x}_i^{\mathrm{S}}) - \frac{1}{n} \sum_{j=1}^{n} \Phi(\boldsymbol{x}_j^{\mathrm{T}}) \right\|_{\mathcal{H}}^{2} \tag{1}$$

Considering the kernel matrix $\boldsymbol{K} \in \Re^{(m+n)\times(m+n)}$ and the coefficient matrix $\boldsymbol{C} \in \Re^{(m+n)\times(m+n)}$ shown as

$$\boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}_{\mathrm{S,S}} & \boldsymbol{K}_{\mathrm{S,T}} \\ \boldsymbol{K}_{\mathrm{T,S}} & \boldsymbol{K}_{\mathrm{T,T}} \end{bmatrix} \tag{2}$$

$$C_{ij} = \begin{cases} \dfrac{1}{m^2} & x_i, x_j \in X_{\mathrm{S}} \\ \dfrac{1}{n^2} & x_i, x_j \in X_{\mathrm{T}} \\ -\dfrac{1}{mn} & otherwise \end{cases} \tag{3}$$

in the p-dimensional RKHS, MMD between $\boldsymbol{X}_{\mathrm{S}}$ and $\boldsymbol{X}_{\mathrm{T}}$ could be defined by

$$Dist\left(\boldsymbol{X}_{\mathrm{S}},\boldsymbol{X}_{\mathrm{T}}\right)=\mathrm{trace}\left(\boldsymbol{WKCKW}^{\mathrm{T}}\right) \tag{4}$$

Combining the idea of minimizing MMD and maximizing variance, the object function of TCA could be defined by

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \mathrm{trace}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{KCKW}\right)+\mu\mathrm{trace}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\right) \\ \text{s.t.} \quad & \boldsymbol{W}^{\mathrm{T}}\boldsymbol{KHKW}=\boldsymbol{I}_p \end{aligned} \tag{5}$$

where $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{KHKW}$ is the constructed covariance matrix, $\boldsymbol{H}=\boldsymbol{I}_{m+n}-1/(m+n)\mathbf{1}\mathbf{1}^{\mathrm{T}}$ is central matrix, $\mathbf{1}\in\Re^{m+n}$ is the column vector whose elements all equal to 1. And $\mu>0$ is a tradeoff parameter, $\boldsymbol{I}_p\in\Re^{p\times p}$ is an identity matrix.

The solution of formula (5) is based on Kernel Fisher Discriminant Analysis (KFD) [16]. The optimal solutions $\boldsymbol{W}^*$ is consist of the p leading eigenvectors of $\left(\boldsymbol{KCK}+\mu\boldsymbol{I}\right)^{-1}\boldsymbol{KHK}$. Then the dimension reduction features $\boldsymbol{X}_{\mathrm{S}}'=\left\{\boldsymbol{x}_1^{\mathrm{S}}',\boldsymbol{x}_2^{\mathrm{S}}',\ldots,\boldsymbol{x}_m^{\mathrm{S}}'\right\}$ and $\boldsymbol{X}_{\mathrm{T}}'=\left\{\boldsymbol{x}_1^{\mathrm{T}}',\boldsymbol{x}_2^{\mathrm{T}}',\ldots,\boldsymbol{x}_n^{\mathrm{T}}'\right\}$ could be extracted by $\boldsymbol{W}^*$.

3) Classifier Training. Plenty of labeled source domain features $\boldsymbol{X}_{\mathrm{S}}'=\left\{\left(\boldsymbol{x}_1^{\mathrm{S}}',y_1^{\mathrm{S}}\right),\ldots,\left(\boldsymbol{x}_m^{\mathrm{S}}',y_m^{\mathrm{S}}\right)\right\}$ and few labeled target domain features $\boldsymbol{X}_{\mathrm{T}}'=\left\{\left(\boldsymbol{x}_1^{\mathrm{T}}',y_1^{\mathrm{T}}\right),\ldots,\left(\boldsymbol{x}_l^{\mathrm{T}}',y_l^{\mathrm{T}}\right)\right\}$ are combined as the training datasets. While Decision Tree [17] is taken as the basic classifier, a classifier $h_f(\cdot)$ is trained by Tradaboost algorithm, in which the training process of Tradaboost is shown in Table I.

TABLE I. TRADABOOST ALGORITHM

Input: source domain data $\boldsymbol{X}_{\mathrm{S}}'=\left\{\left(\boldsymbol{x}_1^{\mathrm{S}}',y_1^{\mathrm{S}}\right),\ldots,\left(\boldsymbol{x}_m^{\mathrm{S}}',y_m^{\mathrm{S}}\right)\right\}$, target domain data $\boldsymbol{X}_{\mathrm{T}}'=\left\{\left(\boldsymbol{x}_1^{\mathrm{T}}',y_1^{\mathrm{T}}\right),\ldots,\left(\boldsymbol{x}_l^{\mathrm{T}}',y_l^{\mathrm{T}}\right),\boldsymbol{x}_{l+1}^{\mathrm{T}}',\ldots,\boldsymbol{x}_n^{\mathrm{T}}'\middle| l\le n\right\}$, weak classifiers number N.

Initialization: Sample weights $\boldsymbol{W}^1=\left(w_1^1,w_2^1,\ldots,w_{m+l}^1\right)$ and weight adjustment parameter $\beta=1/\left(1+\sqrt{2ln(l/N)}\right)$.

Training: For $k=1,2,\ldots N$:

1) Set up weight distribution $\boldsymbol{p}^k=\boldsymbol{W}^k/\sum_{i=1}^{m+l}w_i^k$;

2) Train a classifier h_k with $\boldsymbol{X}_{\mathrm{S}}'$, $\boldsymbol{X}_{\mathrm{T}}'$ and $\boldsymbol{p}^k$;

3) Calculate classification error rate of h_k on $\boldsymbol{X}_{\mathrm{T}}'$ by

$$\mathrm{Error}_k=\sum_{i=m+1}^{m+l}\frac{w_i^k\left(h_k\left(\boldsymbol{x}_i\right)\oplus y_i\right)}{\sum_{i=m+1}^{m+l}w_i^k}$$

4) Set up another weight adjustment parameter β_k by

$$\beta_k=\frac{\mathrm{Error}_k}{1-\mathrm{Error}_k}$$

5) Update sample weights by

$$w_i^{k+1}=\begin{cases} w_i^k\left(1-\beta\right)^{\left(h_k(x_i)\oplus y_i\right)} & i=1,2,\ldots,m \\ w_i^k\left(1-\beta_k\right)^{\left(h_k(x_i)\oplus y_i\right)} & i=m+1,\ldots,m+l \end{cases}$$

Output: Integrate the latter half weak classifiers into

$$h_f\left(\boldsymbol{x}\right)=\sum_{k=\left\lceil\frac{N}{2}\right\rceil}^{N}\frac{\left(1-\mathrm{Error}_k\right)h_k\left(\boldsymbol{x}\right)}{\sum_{k=\left\lceil\frac{N}{2}\right\rceil}^{N}\left(1-\mathrm{Error}_k\right)}$$

Remark: Error_k should be 0.5 at most. If the calculated error is greater than 0.5, setting it to 0.5, then continuing to the next step.

4) Diagnosis Results Output. The extracted target domain features $\boldsymbol{X}_{\mathrm{T}}'=\left\{\boldsymbol{x}_1',\boldsymbol{x}_2',\ldots,\boldsymbol{x}_n'\right\}$ are imported into the classifier $h_f(\cdot)$, and the output classification result is the identified health states of the target domain samples.

III. FAULT DIAGNOSIS OF BEARINGS UNDER VARIABLE OPERATION CONDITIONS

A. A brief introduction to Datasets

The datasets shown in Table II consist of bearing vibration signals obtained from a multi-stage transmission test rig [18]. This test rig consists of three main parts: shaft gearbox, planetary gearbox and magnetic brake. The samples in the datasets have four types of health states: (1) the normal condition (N); (2) the inner race fault (IRF); (3) the outer race fault (ORF); (4) the roller fault (RF). The sampling frequency is set as 12,800 Hz. There are three different operation conditions: (1) motor speed 1200 rpm - load 1; (2) motor speed 1200 rpm - load 2; (3) motor speed 2400 rpm - load 1. There are 800 samples for each datasets and each sample contains 1200 data points.

TABLE II. DATASETS DESCRIPTION

Datasets	Bearing specs	Health states	No. of samples	Motor speed	Load
A	LDK UER204	N	800 (200×4)	1200 rpm	Load 1
		IRF			
		ORF			
		RRF			
B		N	800 (200×4)	1200 rpm	Load 2
		IRF			
		ORF			
		RF			
C		N	800 (200×4)	2400 rpm	Load 1
		IRF			
		ORF			
		RF			

According to Table I, we build two diagnosis tasks A→B and A→C, which respectively represent bearing fault diagnosis tasks under variable load conditions and variable speed conditions. Datasets A is considered as source domain data that the samples and corresponding labels could provide diagnosis knowledge. Datasets B and C are target domain data, and the purpose of the two tasks is to identify the health states of samples in datasets B and C.

B. Diagnosis Task A to B (Variable Loads)

Aiming at fault diagnosis of bearings under variable load conditions, the task A→B is constructed to verify the proposed method, and the result of HTLM is compared with those of Convolutional Neural Networks (CNN), Tradaboost, TCA, and Joint Distribution Adaptation (JDA) [19]. For HTLM, the tradeoff parameter μ and the kernel parameter σ^2 are 0.05 and 2 respectively. The dimension of extracted features is 150. The number of weak classifiers is 40, and the labeled samples in the target domain are only 5% (40 labeled samples).

It is shown in Fig. 2 that the proposed HTLM achieves an accuracy of 87.69%, which is the best result among the results of the five methods above. The instance-based TL method, i.e., Tradaboost obtains the lowest accuracy of 33.20% because it is unable to address raw vibration signals. The feature-based TL method TCA and JDA have lower accuracy than CNN, which are respectively 68.65%, 58.63% and 71.28%. Since TCA and JDA are both single-layer domain adaptation methods, the extracted features can hardly reflect fault information of the collected raw signals, which explains their under-performances. The comparison results proves that the proposed HTLM is able to accurately diagnosing faults under variable load conditions.

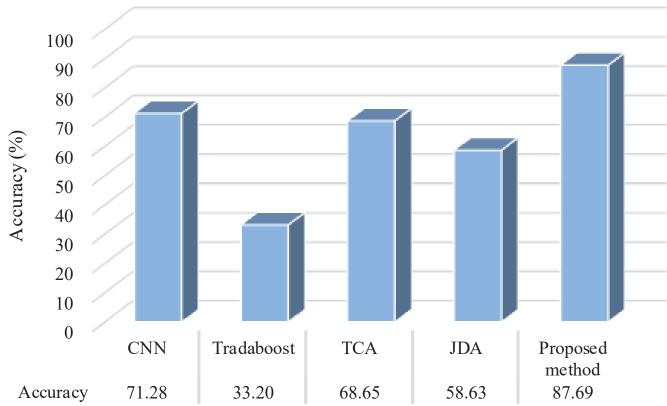

Fig. 2. Classification accuracy (%) for diagnosis task A→B.

C. *Diagnosis Task A to C (Variable speeds)*

The diagnosis task A→C is under the condition of variable speeds, and the compared methods in this task are the same as Sec. III-B. For our proposed method, the tradeoff parameter μ and the kernel parameter σ^2 are respectively 0.05 and 2. The dimension of extracted features is 50. The number of weak classifiers is 40, and the labeled samples in the target domain are only 5% (40 labeled samples).

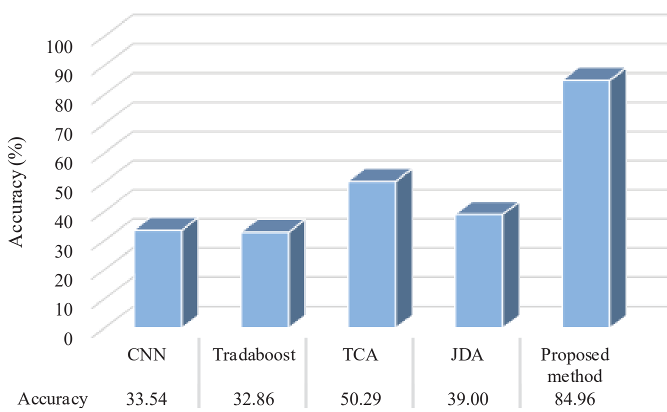

Fig. 3. Classification accuracy (%) for diagnosis task A→C.

As shown in Fig. 3, in the diagnosis task A→C, the accuracies of five methods are lower than those of task A→B, because the distribution discrepancy of samples in different domains are much larger, which makes the diagnosis task under variable speed conditions more complex than that under variable load conditions. In this diagnosis task, the classification accuracy of our proposed method is still the highest one. CNN and Tradaboost achieve similar low accuracy about 33%. TCA and JDA respectively obtain the accuracies of 50.29% and 39.00%. It can be seen from the comparison results that although all the compared methods are not fully capable of dealing with the diagnosis task A→C, our proposed method could still achieve a higher accuracy. Thus, it indicates that the proposed HTLM is suitable for fault diagnosis under variable speed conditions. Moreover, combining with the results in Sec. III-B, it proves that our proposed HTLM is able to achieve high accuracy in intelligent fault diagnosis of bearings under variable operation conditions.

IV. Conclusion

This paper proposes a HTLM to address the problems in the fault diagnosis under variable operation conditions, which could identify the health states under one operation condition with the help of the data collected under other operation conditions. In this method, TCA is firstly applied to extract the cross-domain features subject to similar distribution, and a classifier is trained with all labeled features and corresponding sample weights by Tradaboost algorithm next. The proposed diagnosis method is verified by the laboratory bearing datasets. The testing results indicate that the proposed method is able to achieve a higher diagnosis accuracy in comparison with CNN, Tradaboost, TCA and JDA, which indicates that the proposed HTLM is more effective in fault diagnosis under variable operation conditions, including variable load and variable speed conditions. In future work, for a higher diagnosis accuracy, the deep-layer domain adaptation methods will be considered for constructing a diagnosis model.

Acknowledgment

This research was supported by National Natural Science Foundation of China (61673311) and NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709208).

References

[1] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, “Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data,” Mech. Syst. Signal Process, vol. 72-73, pp. 303-315, May 2016.

[2] L. Guo, N. P. Li, F. Jia, Y. G. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” Neurocomputing, vol. 240, pp. 98-109, May 2017.

[3] F. Jia, Y. G. Lei, L. Guo, J. Lin, and S. B. Xing, “A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines,” Neurocomputing, vol. 272, pp. 619-628, Jan. 2018.

[4] Z. J. Qiao, Y. G. Lei, N. P. Li, “Applications of stochastic resonance to machinery fault detection: A review and tutorial,” Mech. Syst. Signal Process, vol. 122, pp. 502-536, May 2019.

[5] B. Yang, Y. G. Lei, F. Jia, and S. B. Xing, “An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings,” Mech. Syst. Signal Process, vol. 122, pp. 692-706, May 2019.

[6] S. J. Pan, Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

[7] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” IEEE Trans. Neural Netw., vol. 22, no. 2, pp. 199-210, Feb. 2011.

[8] B. Yang, Y. G. Lei, F. Jia, and S. B. Xing, “A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines”, 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), China, Aug. 2018, pp 35–40.

[9] A. K. Waljee, B. Liu, K. Sauder, et al. “Predicting corticosteroid - free endoscopic remission with vedolizumab in ulcerative colitis,” Aliment. Pharmacol. Ther., vol. 47, no. 6, pp. 763-772, Mar. 2018.

[10] A. K. Waljee, B. Liu, K. Sauder, et al. “Predicting Corticosteroid-Free Biologic Remission with Vedolizumab in Crohn’s Disease,” Inflamm. Bowel Dis., vol. 24, no. 6, pp. 1185-1192, Jun. 2018.

[11] F. Shen, C. Chen, R. Q. Chen, and Robert X. Gao, “Bearing fault diagnosis based on SVD feature extraction and transfer learning classification,” In Proceedings of 2015 Prognostics and System Health Management Conference (PHM 2015), China, Jan. 2016, pp 1–6.

[12] Y. Pan, F. Mei, H. Y. Miao, J. Y. Zheng, K. D. Zhu, and H. Y. Sha, “An Approach for HVCB Mechanical Fault Diagnosis Based on a Deep Belief Network and a Transfer Learning Strategy”, J. Electr. Eng. Technol., vol. 14, no. 4, pp. 407-419, Jan. 2019.

[13] J. Y. Xie, L. B. Zhang, L. X. Duan, and J. J. Wang, “On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on Transfer Component Analysis,” 2016 IEEE International Conference on Prognostics and Health Management, Canada, Aug. 2016, pp 1–6.

[14] D. Alpay, P. Jorgensen, and D. Volok, “Relative reproducing kernel Hilbert spaces,” Proceedings Amer. Math. Soc., vol. 142, no. 11, pp. 3889-3895, Nov. 2014.

[15] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schoelkopf, and A. Smola, “A kernel two-sample test,” J. Mach. Learn. Res., vol. 13, pp. 723-773, Mar. 2012.

[16] Q. S. Liu, H. Q. Lu, and S. D. Ma, “Improving Kernel Fisher Discriminant Analysis for face recognition,” IEEE Trans. Circuits Syst. Video Tech., vol. 14, no. 1, pp. 42-49, Jan. 2004.

[17] S. R. Safavian, D. Landgrebe, “A survey of decision tree classifier methodology,” IEEE Trans. Syst. Man Cybern., vol. 21, no. 3, pp. 660-674, May/Jun. 1991.

[18] B. Wang, Y. G. Lei, N. P. Li, and N. B. Li, “A hybrid prognostics approach for estimating remaining useful life of rolling element bearings,” IEEE Trans. Reliab., in press, DOI: 10.1109/TR.2018.2882682.

[19] M. S. Long, J. M. Wang, G. G. Ding, J. G. Sun, and S. Y. Philip, “Transfer feature learning with joint distribution adaptation,” In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Australia, Dec. 2013, pp. 2000-2007.