ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Practice article

A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains

Ping Ma [a], Hongli Zhang [a,*], Wenhui Fan [b], Cong Wang [a]

[a] College of Electrical Engineering, Xinjiang University, Urumqi, China
[b] Department of Automation, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 4 September 2018
Received in revised form 23 August 2019
Accepted 28 August 2019
Available online xxxx

Keywords:
Domain adaptation
Weighted transfer component analysis
Bearing fault diagnosis
Diverse domains

ABSTRACT

In the current research, the diagnosis process of fault diagnosis models is based on an assumption that the same feature distribution exists between training data and testing data. Regrettably, in real applications, datasets are often from diverse domains; in this case, the traditional diagnostic models lack adaptability. To address this issue, this work proposed a transfer diagnosis framework based on domain adaptation, in that the model trained by the labeled data can be applied to diagnose a new but similar target data. An improved domain adaptation algorithm-weighted transfer component analysis (WTCA) is embedded into this framework. Five fault datasets of bearing are used to demonstrate the applicability and practicability of the proposed diagnosis framework. The results show that the proposed diagnosis framework achieves high accuracy in the transfer tasks of bearing fault diagnosis across diverse domains.

1. Introduction

Rolling bearings, as an essential component in rotating machinery, which are widely used in auto-manufacturing, electricity, chemical engineering, etc. The health state of bearings will inevitably affects safes operations of rotating machinery and the costs of production [1,2]. Over the past decade, many intelligent signal processing techniques and machine learning technologies were proposed to analyze the bearing vibration signal and to make diagnostic decisions intelligently, these methods achieved satisfactory results [3,4].

In general, in the previous study, the fault diagnosis framework consists of three steps: data collection, feature extraction, and fault classification (Fig. 1(a)), among which, feature extraction is the key step, which can extract more useful information of vibration signals for fault diagnosis [5–8]. In these steps, some advanced machine learning methods, such as principal component analysis (PCA) [9,10], kernel principal component analysis (KPCA) [11,12], manifold learning and some improved methods based on manifold learning [13–16], such as kernel locality preserving projection (KLPP) [17], have been applied in feature extraction for classification and presented good performance. Despite the considerable success of those methods, traditional fault diagnosis framework has two main latent disadvantages before being applied flexibly in the industry. (1) Most of existing feature extraction and classification models are under the assumption of having the same distribution between training data and testing data. In the current research, the extracted data have the same distribution, the diagnostic model is learned with the training data, and does not guarantee the success on the use of testing data for diverse application domains. Thus, when the distribution of the training data and testing data differs, the diagnostic model often needs to be relearned from scratch for new diagnosis tasks. (2) The working condition variability and diversity of failures of the rotating machine often lead to fewer or even unobtainable target fault data. Datasets are often from diverse domains, and different distributions between the training data and testing data often occur. In this case, the traditional machine learning methods have limitations of fault extraction and classification.

Therefore, developing a new learning method that can reduce the distribution of training data and testing data from diverse domains is necessary to facilitate diagnosing a new but similar testing task, instead of reconstructing and re-training a new diagnosis model from scratch. This diagnosis model is more practical and flexible and can be deployed in a variety of applications. Transfer learning is a new approach [18], which has broad application prospect and wide applicability spanning various fields [19–21]. Feature based transfer, is a mainstream branch of transfer learning technology, which have been applied in image classification [22–24]. In intelligent fault diagnosis, few studies have applied feature based transfer into the diagnosis framework for diverse domain tasks, in which transfer learning methods are combined with neural network to learn feature and

* Corresponding author.
E-mail address: xjdq_mp@foxmail.com (H. Zhang).

https://doi.org/10.1016/j.isatra.2019.08.040
0019-0578/

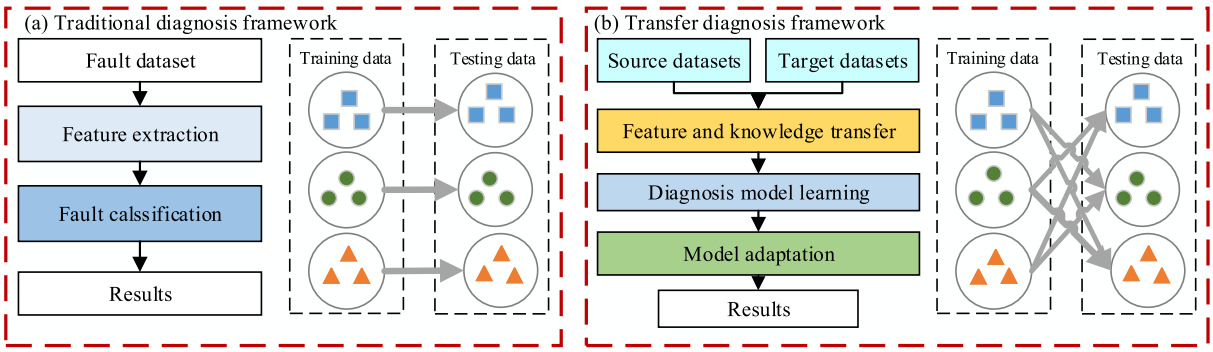

Fig. 1. Intelligent fault diagnosis framework.

transfer the model in the processing of the network training [25, 26]. These researches indicate that feature based transfer method is a promising way to deal with the fault diagnosis of bearings across diverse domain tasks.

Transfer component analysis (TCA) is a typical method in feature based transfer [27]. In contrast with traditional machine learning methods, when the source domain labels are available but the target source data labels are unavailable, TCA aims to reduce the difference in marginal distributions of the different dataset by leveraging the transferable features or knowledge from the source domain (Fig. 1(b)) [28]. This method projected two different domain data onto a specific subspace, which can reduce dramatically the distance in marginal distributions and preserve data properties. TCA reduces the difference in the marginal distributions of two different domain datasets, but it ignores the difference in the conditional distributions. The difference in both the marginal distributions and conditional distributions between different domain datasets need to be minimized, which is crucial to achieve robust distribution adaptation. Therefore, in this paper, based on TCA, an weighted transfer component analysis (WTCA) method is proposed to reduce the difference in the marginal distributions and in the conditional distributions between different datasets.

In this paper, a diagnosis framework based on domain adaptation is proposed to address the need for real industrial application. A novelty domain adaptation algorithm WTCA is introduced to improve the domain adaptability between different domain datasets. The main novelties and contributions of this paper are organized as follows:

(1) A transfer diagnosis framework is introduced to bearing fault diagnosis across diverse domains, in which the learned diagnosis model from source domain datasets can be transferred effectively to a new but similar target domain dataset.

(2) In the transfer diagnosis framework, an improved domain adaptation algorithm WTCA is proposed to reduce the differences in both the marginal distributions and conditional distributions between different datasets, improve the capability of domain adaptation.

(3) The transfer diagnosis tasks are based on five fault datasets of two bearings, i.e., the dataset across different experimental positions, the dataset across diverse fault severity levels, the dataset across diverse fault types, the dataset across diverse operating conditions and the dataset across diverse experiment setups. The results show that the proposed transfer diagnosis framework based on WTCA has strong domain adaptation ability that achieves high accuracy in comparison with other fault diagnosis framework.

The organization of this paper is presented as follows. The theory of WTCA is proposed in Section 2. Section 3 describes the processing of bearing fault diagnosis across diverse domains. In Section 4, the experimental result analysis is given. Conclusions are drawn in Section 5.

2. Weighted transfer component analysis

The hypothesis of the transfer learning is that what had been learned can be adapted to accomplish the new task, building upon and adapting the previous knowledge. The definitions of transfer learning are first presented.

A domain D consists of a D-dimensional feature space X of input and its marginal probability distribution $P(X)$, where $X = \{x_1, \ldots, x_n\} \in \chi$ is a training dataset, $D = \{X, P(X)\}$. $T = \{Y, f(X)\}$ consists of a label space Y and a predictive function $f(X)$, where $Y = \{y_1, \ldots, y_c\} \in Y$ is a training dataset label and $f(X) = Q(Y|X)$ represents the conditional probability distribution. Given a source domain data D_S and a target domain data D_T with corresponding learning task T_S and T_T, feature transfer is used to facilitate the learning process of target predictive function $f_T(X)$ in D_T by using the knowledge and information in D_S and T_S, where $D_S \neq D_T(X_s \neq X_T \vee P_S(X_S) \neq P_T(X_T))$ or $T_S \neq T_T(Y_s \neq Y_T \vee Q_S(Y_S|X_S) \neq Q_T(Y_T|X_T)$.

Given two datasets X_S, X_T, $P_S(X_S) \neq P_T(X_T)$, a transformation ϕ exists such that $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$, where ϕ is a nonlinear mapping function in a reproducing kernel Hilbert space H. The empirical Maximum Mean Discrepancy (MMD) is the distance measure comparing the distance of two marginal distributions $P(X_S)$ and $P(X_T)$, which can be measured by the distance between the source domain and target domain. This can be represented as follows:

$$MMD_H^2(X_S, X_T) = \left\| 1/n_S \sum_{i \in X_S} \phi(x_{S_i}) - 1/n_T \sum_{i \in X_T} \phi(x_{T_i}) \right\|_H^2 \quad (1)$$

where n_S and n_T are the number of two domain samples, respectively. H is a reproducing kernel Hilbert space. From the equation, it can be seen that the empirical estimation of the discrepancy for two distributions can be translated as the distance between the two data distributions in reproducing kernel Hilbert space.

In a fault diagnosis task, given a labeled source dataset $X_S = \{x_{S_i}, y_{S_i}\}_{i=1}^{n_S}$ and a unlabeled target dataset $X_T = \{x_{T_i}, y_{T_i}\}_{i=1}^{n_T}$, $X_s = X_T$, $Y_s = Y_T$, $P_S(X_S) \neq P_T(X_T)$, $P_S(X_S) \neq P_T(X_T)$, $Q_S(Y_S|X_S) \neq Q_T(Y_T|X_T)$. The detailed description of domain adaptation methods TCA and WTCA are presented as follows.

2.1. Transfer component analysis (TCA)

TCA aims to minimize simultaneously the discrepancy between two marginal distributions $P_S(\phi(X_S))$ and $P_T(\phi(X_T))$. Instead of finding the nonlinear transformation ϕ explicitly, this problem can be transform as a kernel learning problem by merit of the kernel trick, (i.e., $K = [\phi(x_i)^T \phi(x_j)]$), the distance shown in Eq. (1)

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

can be written as follows:

$$MMD(X_S, X_T) = \left\| {}^{1}/_{n_S} \sum_{x_i \in X_S} \phi(x_{S_i}) - \frac{1}{n_T} \sum_{x_j \in X_T} \phi(x_{T_i}) \right\|_H^2 = tr(KL), \quad where \quad K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix} \tag{2}$$

where $tr(\cdot)$ represents the trace of a matrix, K is a $(n_S + n_T) \times (n_S + n_T)$ kernel matrix, $K_{S,S}$, $K_{S,T}$ and $K_{T,T}$ respectively are the kernel matrices defined by k on the data in the source domain, cross domain, and target domains, and L can be calculated as

$$L_{ij} = \begin{cases} 1/n_S^2 & (x_i, x_j \in X_S, i, j \in [1, 2, \ldots, n]) \\ 1/n_T^2 & (x_i, x_j \in X_T, i, j \in [1, 2, \ldots, n]) \\ 1/n_S n_T & (other) \end{cases}. \tag{3}$$

TCA maps the features of two domain datasets into the same kernel space through a unified kernel function. The kernel matrix K can be decomposed to the empirical kernel map. The empirical kernel map features can be transforms to low m-dimensional space by using the matrix $\tilde{W} \in R^{(n_S+n_T)\times m}$. The resultant kernel matrix can be calculated as follows:

$$\tilde{K} = (KK^{-1/2}\tilde{W})(\tilde{W}KK^{-1/2}) = KWW^TK, \tag{4}$$

where $W = K^{-1/2}\tilde{W}$. The distance between the two datasets can be defined as

$$MMD(X_S, X_T) = tr((KWW^TK)L) = tr(W^TKLKW). \tag{5}$$

A regularization term $tr(W^TW)$ is used to control the complexity of W, which is used to avoid the rank deficiency of the denominator. The kernel learning problem can be rewritten as

$$\begin{array}{ll} \min_W & tr(W^TKLKW) + \mu tr(W^TW) \\ s.t. & W^TKHKW = I \end{array}, \tag{6}$$

where μ is a trade-off parameter, which is used to guarantee that this problem could be well-defined. $I \in R^{m\times m}$ represents the identity matrix. $H = I^{n_S+n_T} - (1/(n_S + n_T))\mathbf{1}\mathbf{1}^T$ is the centering matrix, $\mathbf{1} \in R^{n_S+n_T}$ is an all 1's column vector, and $I^{n_S+n_T} \in R^{(n_S+n_T)\times(n_S+n_T)}$ is the identity matrix. $W^TKHKW = I$ is used to avoid the trivial solution $W = 0$.

Combining the above observations, the purpose of TCA is that latent space spanned by the learned samples preserves the variance of the data and minimizes the marginal distributions between different domains as much as possible. The optimization problem (6) is transformed into the trace optimization problem is efficiently solved.

2.2. Weighted transfer component analysis (WTCA)

TCA reduces the difference in the marginal distributions of different datasets, but it ignores the difference in the conditional distributions. To achieve robust domain adaptation, minimizing the difference in the conditional distributions is crucial. Note that the labeled source domain data the unlabeled target domain data make the condition distributions $Q_S(Y_S|X_S)$ and $Q_T(Y_T|X_T)$ not being modeled directly.

In this section, an improved domain adaption method WTCA is proposed. Using the true source and pseudo target labels, the class-conditional distributions $Q_S(x_S|y_S = c)$ and $Q_T(x_t|y_t = c)$ can be matched essentially, each class $c \in \{1, \ldots, C\}$ in the label set C. The MMD is modified to measure the distance of $Q_S(x_S|y_S = c)$ and $Q_T(x_T|y_T = c)$,where its objective is to minimize

$$Dist(X_S, X_T) = \left\| {}^{1}/_{n_S^{(c)}} \sum_{x_i \in D_S^{(c)}} \phi(x_{S_i}) - {}^{1}/_{n_T^{(c)}} \sum_{x_j \in D_T^{(c)}} \phi(x_{S_j}) \right\|_H^2 = \sum_{c=1}^{C} tr(W^TKL_1K^TW), \tag{7}$$

where $D_S^{(c)} = \{x_i : x_i \in D_S \vee y_i = c\}$ and $D_T^{(c)} = \{x_j : x_j \in D_T \vee \hat{y}_j = c\}$ present the samples with the class c in the source and target data, respectively, y_i is the true label of x_i, and $\hat{y}_j$ presents the pseudo label of x_j. The matrices L_1 involving class labels can be reformulated as

$$(L_1)_{ij} = \begin{cases} \frac{1}{n_S^{(c)} n_S^{(c)}}, & x_i, x_j \in D_S^{(c)} \\ \frac{1}{n_T^{(c)} n_T^{(c)}}, & x_i, x_j \in D_T^{(c)} \\ \frac{-1}{n_S^{(c)} n_T^{(c)}}, & \begin{cases} x_i \in D_S^{(c)}, x_j \in D_T^{(c)} \\ x_j \in D_S^{(c)}, x_i \in D_T^{(c)} \end{cases} \\ 0, & otherwise \end{cases}. \tag{8}$$

Then, the objective function of linear discriminate analysis is added to improve the separability of class labels, which can be written as:

$$\min_W atr\left(\frac{W^TKS_bKW}{W^TKS_tKW}\right) \tag{9}$$

where a is the equilibrium coefficient. S_b represents the between-class scatter matrix, and S_t is the total scatter matrix, which can be obtained as follow

$$\begin{aligned} S_b &= \sum_{c=1}^{C} (u^{(c)} - u)(u^{(c)} - u)^T \\ S_w &= \sum_{c=1}^{C} \sum_{l=1}^{n_c} (x_l^{(c)} - u^{(c)})(x_l^{(c)} - u^{(c)})^T \\ S_t &= \sum_{c=1}^{C} (x_l - u)(x_l - u)^T = S_w + S_b \end{aligned} \tag{10}$$

where S_w is the within-class scatter matrix, n_c and $u^{(c)} = \frac{1}{n_c}\sum_{l=1}^{n_c} x_l^{(c)}$ are the number and mean of samples belonging to the class c, respectively. x_l is the lth sample, $u = \frac{1}{n}\sum_{t=1}^{n} x_t$ is the mean of the total sample.

The aim of WTCA is to minimize the differences in both the marginal distributions and conditional distributions of different datasets, and improve the capability of domain adaptation. Thus, the final optimization problem of WTCA is defined as:

$$\begin{array}{ll} \min_W & \sum_{c=1}^{C} tr(W^TKL_1KW) + atr\left(\frac{W^TKS_bKW}{W^TKS_tKW}\right) + utr(W^TW) \\ s.t. & W^TKHKW = I_m \end{array}. \tag{11}$$

The optimization problem (11) can be transformed to the trace optimization problem.

Proof. The diagonal matrix containing Lagrange multipliers is applied to Eq. (11), it can be transformed as follows:

$$tr(W^T(K\sum_{c=1}^{C} L_1K + a(S_t - S_b) + uI)W) - tr((W^TKHKW - I)Z). \tag{12}$$

Setting $\partial W/\partial t = 0$, Eq. (12) can be transformed as

$$(K\sum_{c=1}^{C} L_1 K + a(S_t - S_b) + uI)W = KHKWZ. \quad (13)$$

Multiplying both sides on the left by W^T, the following can be obtained:

$$\min_W tr((W^T KHKW)^{\dagger} W^T (K\sum_{c=1}^{C} L_1 K + a(S_t - S_b) + \mu I)W). \quad (14)$$

Then, the following trace optimization problem can be obtained:

$$\max_W tr(W^T (K\sum_{c=1}^{C} L_1 K + a(S_t - S_b) + \mu I)W)^{-1} W^T KHKW, \quad (15)$$

when $C = 1$ and $a = 0$, WTCA degenerates to TCA.

With WTCA, more accurate labeling for the target data can be obtained. It can alternately improve the labeling quality through iteration when use this labeling as a pseudo target label. The procedure for TCA and WTCA are summarized as follows:

Algorithm 1: TCA

Input: Training data: $X_S = \{(x_{S_1}, y_{S_1}),...,(x_{S_m}, y_{S_C})\}$, testing data : $X_T = \{x_{t_1},...,x_{S_{n_t}}\}$.

Output: Transformation matrix W.

Begin

1. Construct K, L, and $\backslash H$.
2. Solve the optimization problem in Eq.(6) and construct W.
3. Return W.
4. Train a standard classifier f on $\{(W^T K_i, y_i)\}_{i=1}^{n_s}$ to obtain predict target labels $\{y_j = f(W^T K_j)\}_{j=n_s+1}^{n_s+n_t}$.

Algorithm 2:WTCA

Input: Training data: $X_S = \{(x_{S_1}, y_{S_1}),...,(x_{S_m}, y_{S_C})\}$ testing data $X_T = \{x_{t_1},...,x_{S_{n_T}}\}$.

Output: Transformation matrix W, adaptive classifier f.

Begin

1. Construct L_1, train a classifier f on $\{(W^T K_i, y_i)\}_{i=1}^{n_s}$ to obtain initial pseudo target labels $\{\hat{y} = f(W^T K_j)\}_{j=n_s+1}^{n_s+n_t}$.
2. **Repeat:**
   3. Solve the optimization problem in Eq.(11) and construct W.
   4. Train a standard classifier f on $\{(W^T K_i, y_i)\}_{i=1}^{n_s}$ to update pseudo target labels $\{y_j = f(W^T K_j)\}_{j=n_s+1}^{n_s+n_t}$.
   5. Construct L_1 by Eq.(8).
6. **Until** the end of the iteration.
7. Return an adaptive classifier f trained on $\{(W^T K_i, y_i)\}_{i=1}^{n_s}$.

3. Fault diagnosis procedure based on domain adaptation

In real industrial processes conditions, abominable operating conditions often lead to fewer or even unobtainable target fault data in bearing fault diagnosis, which, in turn, often lead to a large domain discrepancy between source domain and target domain; the traditional diagnosis frameworks are limited in real industrial application. Consequently, a novel transfer diagnosis framework-based domain adaptation is proposed for bearing fault diagnosis on diverse transfer tasks.

3.1. Fault diagnosis procedure

The fault diagnosis procedure of our proposed diagnosis framework can be summed up into three steps. (1) Data collection. The vibration signal can be divided into some segments. For each segment, the time and frequency domain features of labeled samples and unlabeled samples are extracted as a high-dimensional mixed-domain feature dataset. (2) Feature selection and domain adaptation. The high-dimensional mixed-domain feature set of source and target samples is compressed into a m-dimensional eigenvector by using the domain adaptation method, where m is the embedding dimension. (3) Diagnosis model learning and testing. The m-dimensional eigenvectors of the source domain as the training samples input into classifier to produce the trained classification model, which can diagnose the target domain dataset effectively. The scheme of the proposed transfer diagnosis framework is described briefly in Fig. 2.

3.2. Comparison studies

In this work, the proposed method will be compared with several fault diagnosis methods: (1) K-NN; (2) PCA; (3) KLPP; (4) TCA, and (5) WTCA (this work), among which (1)–(3) are the standard machine learning diagnosis methods and (4)–(5) are the transfer learning-based diagnosis methods.

In (1), the former 29 popular time and frequency domain statistical features from vibration signal in literature [29] are extracted to form source dataset and target dataset; using the labeled data to train the classifiers and the unlabeled data can be diagnosed by this model. In the traditional feature extraction methods PCA and KLPP, the 29 statistical features are extracted to form a high-dimensional dataset, which consists of the labeled source data and unlabeled target data. PCA and KLPP process the dimension reduction of the high-dimensional dataset. Similarly, in the transfer learning methods TCA and WTCA, extracted 29 statistical features to form a high-dimensional dataset, TCA and WTCA are used to conduct the unsupervised domain adaptation and dimension reduction of the labeled data and unlabeled data. For (2)–(5), the classification model is trained with low-dimensional features in source domain, and the low-dimensional features in target domain can be diagnosed. In KLPP, TCA, and WTCA, the kernel type is selected from RBF. Most existing methods use trial and error to determinate the parameters in these methods. In our study, the adaptation regularization parameter μ is set to 0.01. The equilibrium coefficient a is 0.001. The kernel bandwidth λ of RBF and the embedded dimension m of domain adaptation are optimized with the strategy of trial and error.

4. Experiments

In this section, experiments for four different conditions of bearings were conducted to demonstrate the efficiency and practical value of the proposed transfer diagnosis framework for fault diagnosis.

The bearing data are obtained from the Bearing Data Center of the Case Western Reserve University (CWRU). The dataset are obtained from a fault injection test of 6205-SKF deeply grooved ball bearings and the electro-discharge machining is applied to set a single point failure with the fault diameters 0.007 inches, 0.014 inches and 0.021 inches. The collected signal of bearing

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

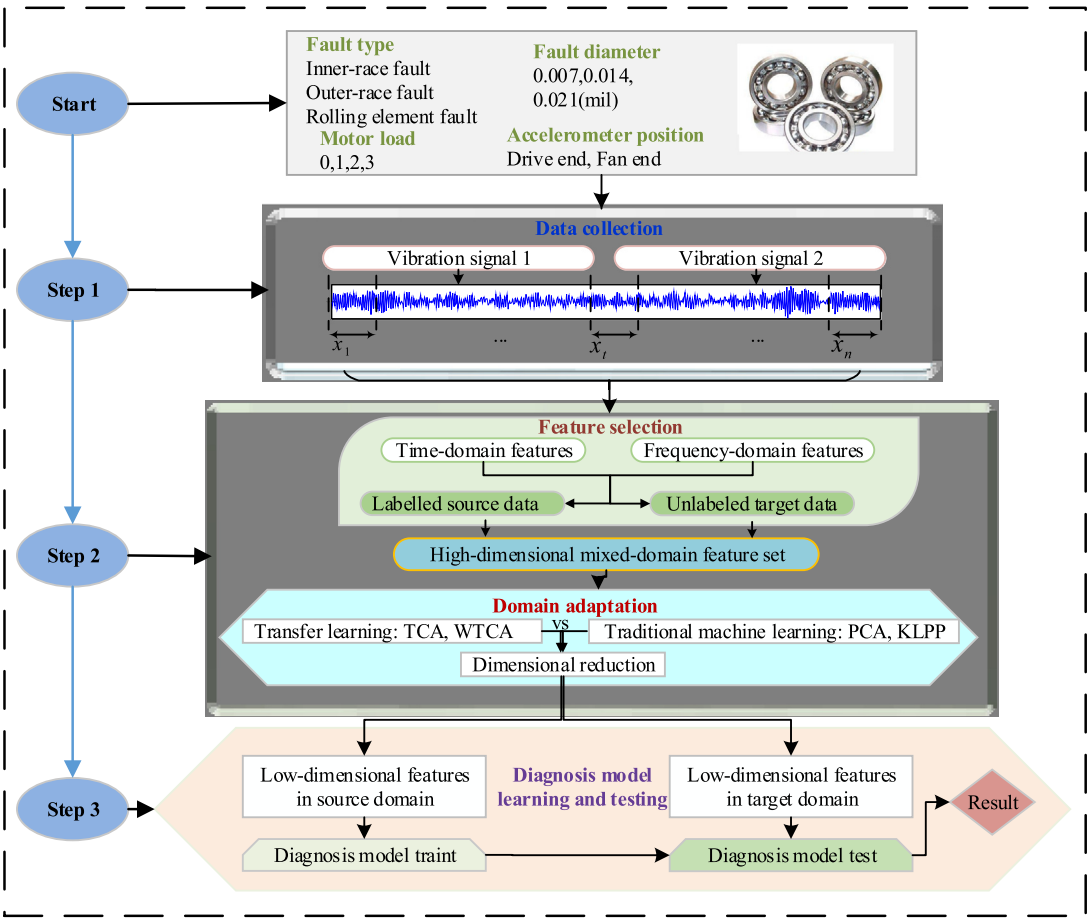

Fig. 2. Flow chart of the proposed diagnosis framework based on domain adaptation of transfer learning.

consists of four state: the inner race fault (IR), rolling element fault (RE), outer race fault (OR) and normal. In the drive end (DE) and fan end (FE) of the motor housing, 3 accelerometers are placed at the 12 o'clock position, respectively. The test bearing supports the motor shaft and motor loads of 0 to 3 horsepower (hp). The sampling frequency is 12 kHz for drive end and fan end bearing experiments, and the experimental setup shown in Fig. 3. The length of each segment of vibration signal is set to 1024.

4.1. Data description and transfer task establishment

This section considers the following four transfer task of fault diagnosis: (1) Designed a transfer task across the different experimental positions. (2) Designed a transfer task across the diverse fault types or diverse fault severities. (3) Designed a transfer task that diagnoses fault severity levels across diverse motor loads. (4) Designed a transfer task that diagnoses fault types across diverse motor loads.

(1) Task 1: a transfer task across the different experimental positions

The experiments are performed with bearing data collected from two different experimental positions. For single-point faults with the same severity levels/fault type, the vibration data of three fault types/three severity levels are collected by two accelerometers, which are installed in the DE and FE of the motor housing, respectively.

Task 1 has twelve datasets named 'A' to 'K' and the samples contained in each dataset is differ. Twelve transfer tasks are performed using these datasets for demonstration: A → B, B → A, C → D, D → C, E → F, F → E and G → H, H → G, I → J, J → I, K → L, L → K. Tasks A → B, B → A, C → D, D → C, E → F, F → E aim to classify diverse fault severities across different experimental positions. For clarity, A → B is defined as the transfer task from source dataset A to target dataset B. The source dataset A contains the three fault diameters of RE, which are collected from the accelerometers installed in the DE of the motor housing, while the target dataset B consisted of the same samples that are collected from the accelerometers installed in the FE of the motor housing. Tasks G → H, H → G, I → J, J → I, K → L, L → K aim to classify diverse fault severities across different experimental positions. Similarly, in G → H, the source dataset G contains three fault types (i.e., RE, IR, and OR) with fault diameters of 0.007 inches, which are collected from the accelerometers installed in the DE of the motor housing, while the target dataset H consisted of the same samples collected from the accelerometers installed in the FE of the motor housing. Each dataset has 180 samples and every fault diameters or fault type has 60 samples. The details of the designed transfer tasks across different experimental positions are shown in Table 1.

(2) Task 2: a transfer task across the diverse fault types or diverse fault severities

In most existing studies, samples with the same fault severity but different fault types or samples with the same fault type but different fault severities are also regarded as distinct categories. These fault signals have various characteristics. In real industry diagnosis application, those distinct categories will lead to a large domain discrepancy; therefore, sufficient data are needed which are used to train the diagnosis model of distinct categories. Therefore, this section aims to verify the performance of the

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

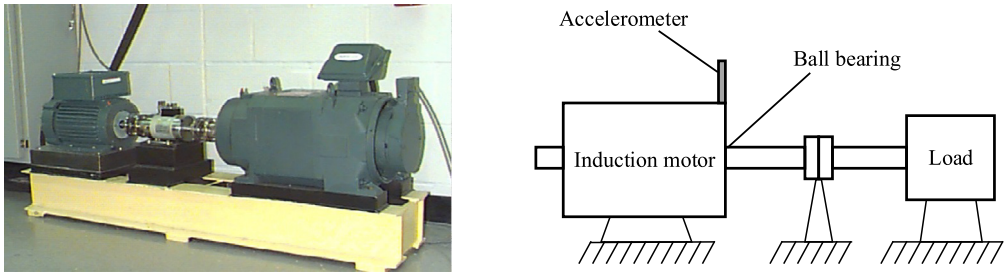

Fig. 3. The bearing test rig.

Table 1
Details of the designed transfer task across the different experimental positions.

Classified diverse fault severities and diverse fault types across different experimental positions (under load 2)											
Classified diverse fault severities across different experimental positions (0.007, 0.014, 0.021)						Classified diverse fault types (RB, IR, OR) across different experimental positions					
Transfer tasks	Fault types	Source domain	Target domain	Labeled source samples	Unlabeled target samples	Transfer tasks	Fault diameters	Source domain	Target domain	Labeled source samples	Unlabeled target samples
A→B	RE	FE	DE	180	180	G→H	0.007	FE	DE	180	180
B→A		DE	FE	180	180	H→G		DE	FE	180	180
C→D	IR	FE	DE	180	180	I→J	0.014	FE	DE	180	180
D→C		DE	FE	180	180	J→I		DE	FE	180	180
E→F	OR	FE	DE	180	180	K→L	0.021	FE	DE	180	180
F→E		DE	FE	180	180	L→K		DE	FE	180	180

Table 2
Details of the designed transfer task across the diverse fault types and diverse fault severities.

Classified fault severities/ types across diverse fault type/ severities (under load 2, in drive end)									
Classified fault severity levels across diverse fault					Classified fault types across diverse fault severity				
Transfer tasks	Source domain (fault types)	Target domain (fault types)	Labeled source samples	Unlabeled target samples	Transfer tasks	Source domain (fault diameters)	Target domain (fault diameters)	Labeled source samples	Unlabeled target samples
A→C	RE	IR	180	180	G→I	0.007	0.014	180	180
C→A	IR	RE	180	180	I→G	0.014	0.007	180	180
C→E	IR	OR	180	180	I→K	0.014	0.021	180	180
E→C	OR	IR	180	180	K→I	0.021	0.014	180	180
A→E	RE	OR	180	180	G→K	0.007	0.021	180	180
E→A	OR	RE	180	180	K→G	0.021	0.007	180	180

proposed transfer learning methods across the diverse fault types and diverse fault severities in this dataset.

Task 2 has twelve transfer tasks are performed for demonstration: A → C, C → A, C → E, E → C, A → E, E → A and G → I, I → G, I → K, K → I, G → K, K → G. Tasks A → C, C → A, C → E, E → C, A → E, E → A aim to classify diverse fault severities across different fault types. For instance, in A → C, the source dataset A contains the three fault diameters of RE, while the three fault diameters of IR form the target dataset B. Tasks G → I, I → G, I → K, K → I, G → K, K → G aim to classify diverse fault types across different fault severities. Similarly, in G → I, the dataset G contains three fault types with the fault diameters of 0.007 inches. The target dataset I is formed by the three fault types with the fault diameters of 0.014 inches. The details of the designed transfer task across the diverse fault types and diverse fault severities are shown in Table 2.

(3) Task 3: a transfer task that diagnoses fault severity levels across diverse motor loads

In task 3, the vibration signal of the bearings under varying loading conditions are collected to create operating conditions close to reality. The experiment is conducted under three motor loads that range from 0 to 3 horsepower (hp). This task aim to explore the ability of proposed transfer methods for fault severity level classification across diverse operating conditions.

For a bearing fault, three fault severity level data are collected from load 0, 1, and 2, which are regarded as datasets 0, 1, and 2, respectively. Take the RE fault as an example; six transfer tasks are performed for demonstration: 0→1, 1→0, 1→2, 2→1, 1→3, 2→3. For instance, in 0→1, the source dataset 0 contains the three fault severities of RE fault under load 0, while the three fault severities of RE under load 1 form the target dataset 1. Correspondingly, for IR and OR faults, six transfer tasks are performed. This task aims to classify diverse fault severities across diverse operating conditions. The details of the designed transfer task across diverse operating conditions are shown in Table 3.

(4) Task 4: a transfer task that diagnoses fault types across diverse motor loads

In task 4, similar to task 3, it aims to explore the ability of the proposed transfer methods for fault type classification across diverse operating conditions. Under the same fault severity levels, three fault type data collected from loads 0, 1, and 2 are regarded as datasets 0, 1, and 2, respectively. Take fault severity 0.007 as an example; six transfer tasks are performed for demonstration: 0→1, 1→0, 1→2, 2→1, 1→3, 2→3. For instance, in 0→1, the source dataset 0 contains three fault types with 0.007 fault

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

Table 3
Details of the designed transfer task of fault severities classification across diverse operating conditions.

Classified different severities of fault type across diverse operating conditions (under load 2, in drive end)										
Classified severity levels of RE fault			Classified severity levels of IR fault			Classified severity levels of OR fault				
Transfer task	Source domain	Target domain	Dataset: Transfer task	Source domain	Target domain	Dataset: Transfer task	Source domain	Target domain	Labeled source samples	Unlabeled target samples
1: 0→1	load 0	load 1	7: 0→1	load 0	load 1	13: 0→1	load 0	load 1	180	180
2: 1→0	load 1	load 0	8: 1→0	load 1	load 0	14: 1→0	load 1	load 0	180	180
3: 0→2	load 0	load 2	9: 0→2	load 0	load 2	15: 0→2	load 0	load 2	180	180
4: 2→0	load 2	load 0	10: 2→0	load 2	load 0	16: 2→0	load 2	load 0	180	180
5: 1→2	load 1	load 2	11: 1→2	load 1	load 2	17: 1→2	load 1	load 2	180	180
6: 2→1	load 2	load 1	12: 2→1	load 2	load 1	18: 2→1	load 2	load 1	180	180

Table 4
Details of the designed transfer task of fault severities classification across diverse operating conditions.

Classified different severities of fault type across diverse operating conditions (under load 2, in drive end)										
Classified faults types with the fault diameter of 0.007			Classified faults types with the fault diameter of 0.014			Classified faults types with the fault diameter of 0.021				
Transfer task	Source domain	Target domain	Dataset: Transfer task	Source domain	Target domain	dataset: Transfer task	Source domain	Target domain	Labeled source samples	Unlabeled target samples
1: 0→1	load 0	load 1	7: 0→1	load 0	load 1	13: 0→1	load 0	load 1	180	180
2: 1→0	load 1	load 0	8: 1→0	load 1	load 0	14: 1→0	load 1	load 0	180	180
3: 0→2	load 0	load 2	9: 0→2	load 0	load 2	15: 0→2	load 0	load 2	180	180
4: 2→0	load 2	load 0	10: 2→0	load 2	load 0	16: 2→0	load 2	load 0	180	180
5: 1→2	load 1	load 2	11: 1→2	load 1	load 2	17: 1→2	load 1	load 2	180	180
6: 2→1	load 2	load 1	12: 2→1	load 2	load 1	18: 2→1	load 2	load 1	180	180

Table 5
Properties of algorithms for dimensionality reduction.

Algorithm	Parametric	Parameters	Computational
PCA	Yes	None	$O(D^3)$
KLPP	No	$K(\cdot,\cdot)$	$O((n_S+n_T)^3)$
TCA	No	$K(\cdot,\cdot)$, μ	$O(mD^2+(n_S+n_T)^2)$
WTCA	No	$K(\cdot,\cdot)$, μ	$O(imD^2+iC(n_S+n_T)^2+iD(n_S+n_T))$

severity under load 0, while the same samples under load 1 form the target dataset 1. Correspondingly, for fault severities of 0.014 and 0.021, six transfer tasks are performed respectively. This task aims to classify diverse fault types across diverse operating conditions. The details of the designed transfer task are shown in Table 4.

4.2. General properties

Three properties of the proposed algorithm and others comparison algorithms: (1) the parametric nature of the mapping between the high-dimensional and the low-dimensional space, (2) the main free parameters that have to be determined, (3) the computational complexities are listed in Table 5.

The property 1 in Table 5 shows that most algorithms are non-parametric. These algorithms are not possible to reconstruct the low-dimensional data to the original high-dimensional data because they not have a direct mapping from high-dimensional to the low-dimensional space. For property 2, except for PCA, other three nonlinear algorithms need to determine the free parameters that directly influence the cost function. It can be noted that TCA and WTCA have additional regularization parameter μ. The free parameters make the algorithm more flexible, whereas it is necessary to optimize to improve the performance of the dimensionality reduction algorithm.

The computational complexities of these algorithms are analyzed using the big O notation. The computational complexity of a algorithm is consisting of two parts: (1) dataset properties. It consists of the number of datapoints n ($n = n_S + n_T$) and dataset dimensionality D; (2) algorithm parameters, the target dimensionality m and the iterations i. The eigenanalysis of the $D \times D$ covariance matrix is the computationally most demanding part of PCA, which is performed using a power method in $O(D^3)$. Through a kernel transform, KLPP perform an eigenanalysis of an $(n_S+n_T)\times(n_S+n_T)$ matrix using a power method in $O((n_S+n_T)^3)$.

The computational complexities of TCA and WTCA are detailed as follows: (1) TCA: $(n_S+n_T)^2$ for constructing the MMD matrices, $O(mD^2)$ for solving the generalized eigendecomposition problem. (2) WTCA: $C(n_S+n_T)^2$ for constructing the MMD matrices, $iD(n_S+n_T)$ for all remaining steps, All computational complexity of WTCA is $O(imD^2 + iC(n_S + n_T)^2 + iD(n_S + n_T))$. The computational complexities show that WTCA impose demands on computational resources. In this study, the selected features of bearing signal are composed into a dataset within a relatively low-dimensional space, and the number of this dataset is small. The analysis process has low complexity and the application of the WTCA is considered valuable.

4.3. Result discussion

(1) Task 1: designed transfer task across the different experimental positions

In task 1, the kernel bandwidth λ of RBF and the embedded dimension m of domain adaptation of twelve tasks are shown in Fig. 4. For each experiment, using the basic K-nearest neighbor classifier bearing faults using the features extracted from these algorithms, repeated 10 times and then report the average classification accuracy. The accuracy of classification is obtained by calculating the percentage of correctly classified samples to total samples. The diagnosis results of these twelve tasks are presented in Fig. 5 and Table 6. It can be observed that the transfer diagnosis

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

Table 6
Diagnosis accuracy (%) on transfer task 1 with five methods.

Methods	A→B	B→A	C→D	D→C	E→F	F→E	G→H	H→G	I→J	J→I	K→L	L→K	Average
NN	44.9	36.1	66.7	66.7	46.7	58.3	65.6	46.7	39.4	35.0	66.7	56.1	52.4
KLPP	36.7	31.1	67.8	66.7	66.7	59.4	65.0	66.7	55.6	40.0	66.7	52.8	56.2
PCA	37.2	41.1	67.7	55.6	66.7	41.1	56.1	66.7	32.8	48.9	55.6	33.3	50.1
TCA	60.6	48.3	72.2	73.3	98.9	100	80.0	86.7	58.3	52.2	96.7	100	77.3
WTCA	**63.9**	**46.1**	**98.3**	**100**	**100**	**100**	**100**	**96.1**	**81.7**	**80.6**	**100**	**100**	**88.9**

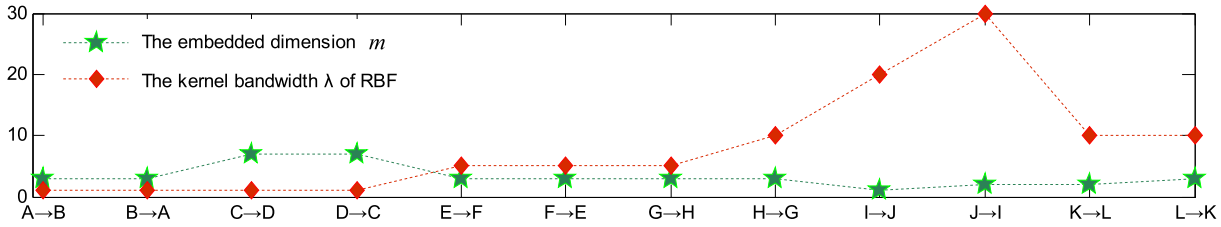

Fig. 4. Value of kernel bandwidth λ of RBF and embedded dimension m on transfer task 1.

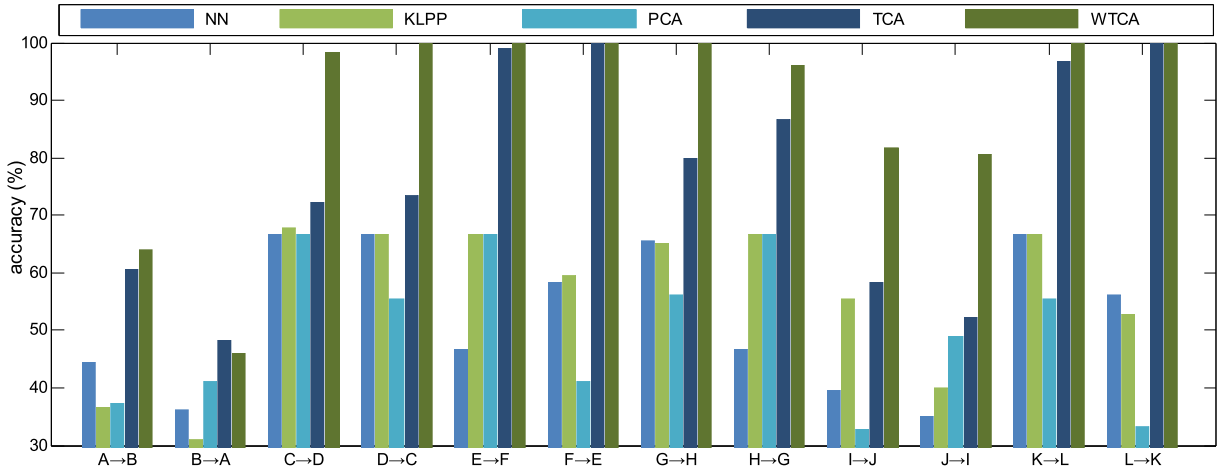

Fig. 5. Comparison of the performance of five methods on transfer task 1.

framework based on TCA and WTCA in this work significantly outperforms the other three standard machine learning methods. The domain adaptation ability of WTCA is demonstrated by significantly improved accuracies under different transfer tasks. Compared with the other three methods, the diagnosis performance is much improved because of the transfer framework with domain adaptation. Furthermore, the average accuracy of TCA and WTCA are 77.3% and 88.9%, respectively and transfer improvement of 24.9% and 36.5%, compared with the baseline KNN of 52.4%. It also explains that compared with TCA, WTCA has a better domain adaptation performance by considering jointly the marginal and conditional distributions of dataset.

To make a clearer presentation of the effectiveness of the proposed framework, feature visualization is used to analyze the high performance of the proposed model. To facilitate comparison of these methods, PCA is used to transform the extracted low-dimensional features into a two-dimensional latent space. This method is used because it has few parameters and is easy to implement. In this transfer task, for example task L → K, the domain adaptation across different experimental positions is realized. In Fig. 6, three categories in two domains are regarded as labels 1, 2, and 3, respectively. S and T represent the samples in source and target domains, respectively. For example, the L1-S means the samples of categories 1, in the source domain. As can be seen in Fig. 6(a), under the transfer framework based on WTCA, the three categories of the source and target samples are separated well, and the same category between the source and target domains is well-overlapped. Similarly, as shown in Fig. 6(b), the distribution of the same category between the source and target domains is very close using the transfer framework based on TCA, which makes the category of the target domain fault easy to classify. In Fig. 6(c), under the standard framework based on KLPP, most of the three categories of the source and target samples are separated, while the data distribution between the source and target domains have considerable differences, which will lead to degraded classification results of target domain samples. The analysis demonstrated that the proposed transfer framework is capable of domain adaption, which achieves high accuracy for fault diagnosis across different experimental positions.

(2) Task2: designed transfer task across the diverse fault types or diverse fault severities

Task 2 aims to classify diverse fault severities across different fault types and classify diverse fault types across different fault severities. Fig. 7 presents the kernel bandwidth λ of RBF and the embedded dimension m of domain adaptation of twelve tasks of transfer task 2. The diagnosis results of twelve tasks are presented in Fig. 8 and Table 7.

Several phenomena can be noted. First, the transfer framework based on WTCA can significantly improve the accuracy of classification. The average accuracy of the transfer framework with WTCA is 78.9%, which doubles the accuracy of the baseline KNN which is 33.3%. Second, in the three standard methods, the diagnosis performance varies on different tasks. For example, in

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

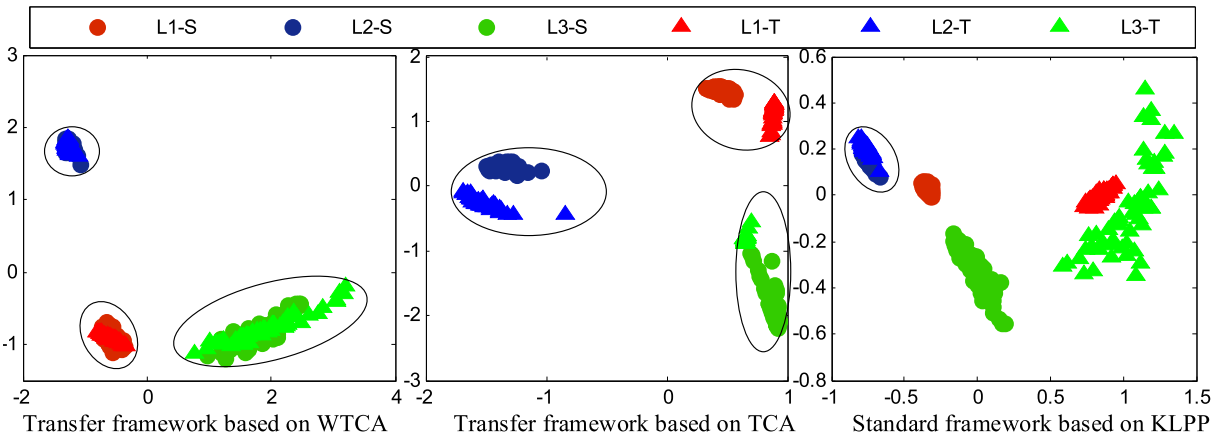

Fig. 6. Domain adaptation visualization in task L → K.

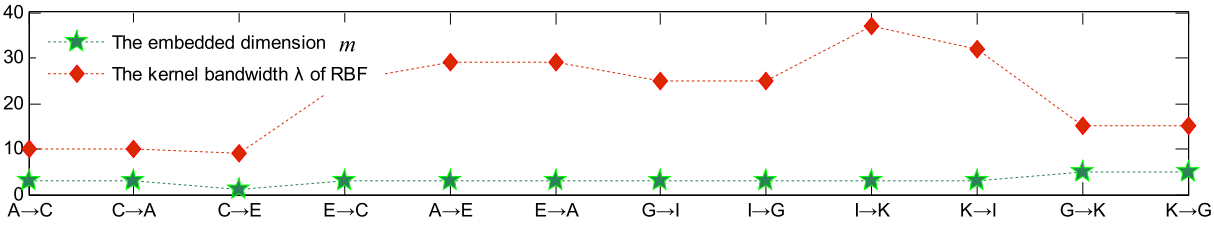

Fig. 7. Value of kernel bandwidth λ of RBF and embedded dimension m on transfer task 2.

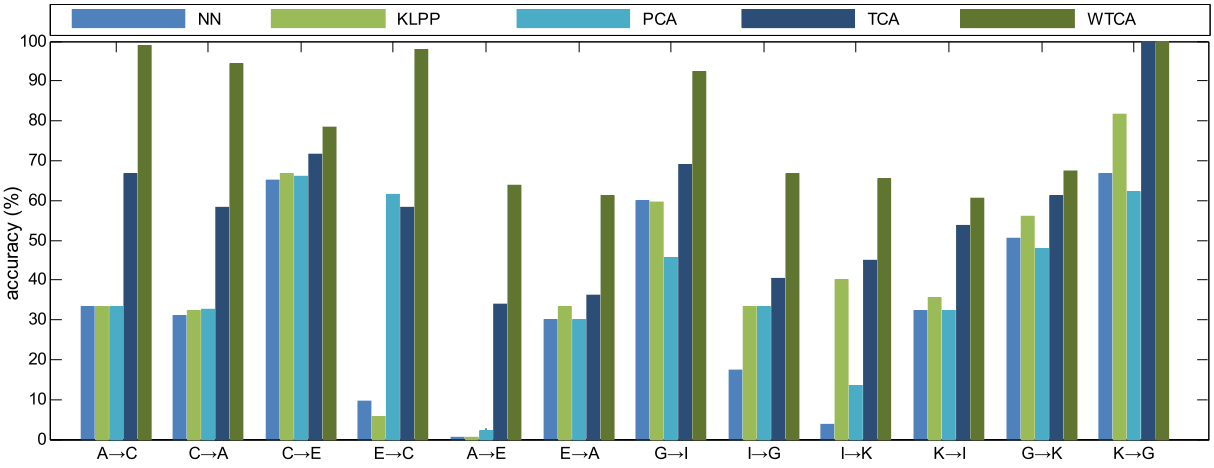

Fig. 8. Comparison of the performance of five methods on transfer task 2.

Table 7
Diagnosis accuracy (%) on transfer task 2 with five methods.

Methods	A→C	C→A	C→E	E→C	A→E	E→A	G→I	I→G	I→K	K→I	G→K	K→G	Average
NN	33.3	31.1	65.0	9.44	0.56	30.0	60.0	17.2	3.90	32.2	50.6	66.7	33.3
KLPP	33.3	32.2	66.7	5.56	0.56	33.3	59.4	33.3	40.0	35.6	56.1	81.7	39.8
PCA	33.3	32.8	66.1	61.7	2.18	30.0	45.6	33.3	13.3	32.2	47.8	62.2	38.4
TCA	66.7	58.3	71.7	58.3	33.9	36.1	68.9	40.6	45.0	54.9	61.1	100	57.9
WTCA	**98.9**	**94.4**	**78.3**	**97.8**	**63.9**	**61.1**	**92.3**	**66.7**	**65.6**	**60.6**	**67.2**	**100**	**78.9**

the method of NN, the classification accuracies of task A → E and E → A are 9.44% and 30.0%, respectively, while the results in tasks G → K and K → G are 50.6% and 66.7%, respectively. This result is because the operating conditions of A and E are closer than the G-H, and the data of G and K have a more similar feature space, which achieves higher diagnostic accuracy. It also reveals that the traditional diagnosis framework relies considerably on the same distribution of training data and testing data. Third, compared with the transfer task 1, the classification accuracies of transfer task 2 are reduced in most cases, because the distribution difference of dataset in task 2 is bigger than task 1. In this case, the transfer framework with WTCA, significantly promotes the adaptation and classification ability of most tasks, especially under the transfer tasks of diverse fault types.

Task C → A adapts the distribution across diverse fault types. As shown in Fig. 9(a), the transfer framework based on WTCA,

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

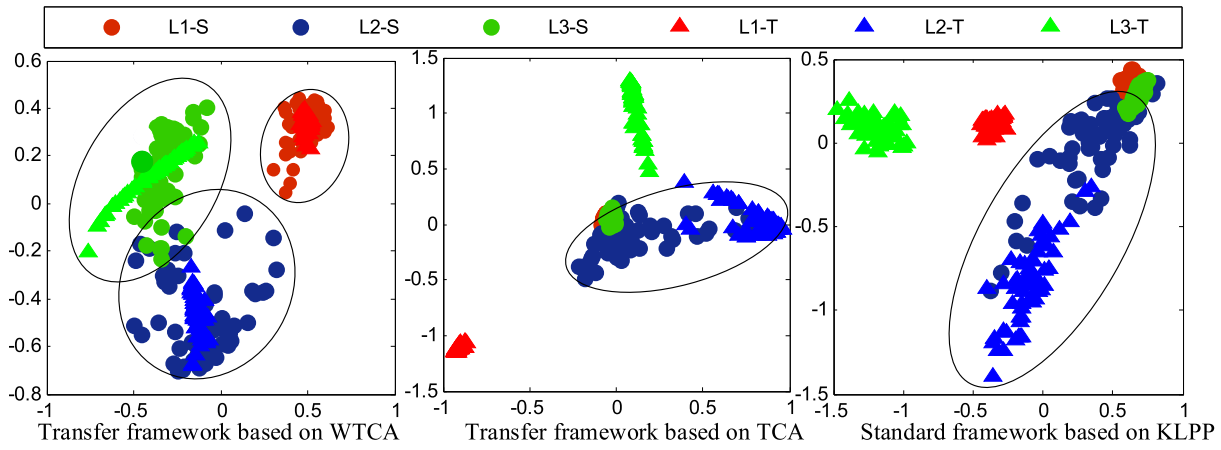

Fig. 9. Domain adaptation visualization in task C → A.

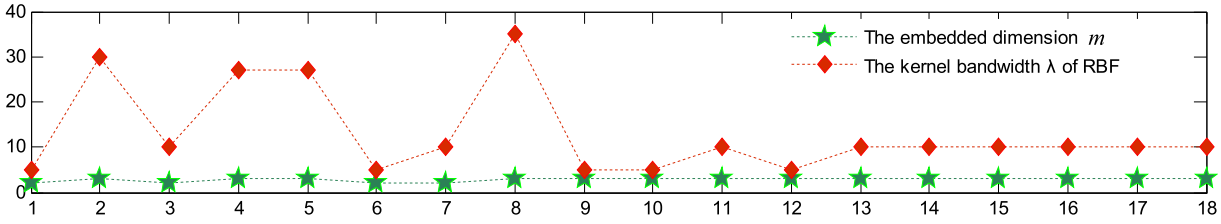

Fig. 10. The value of kernel bandwidth λ of RBF and embedded dimension m on transfer task 3.

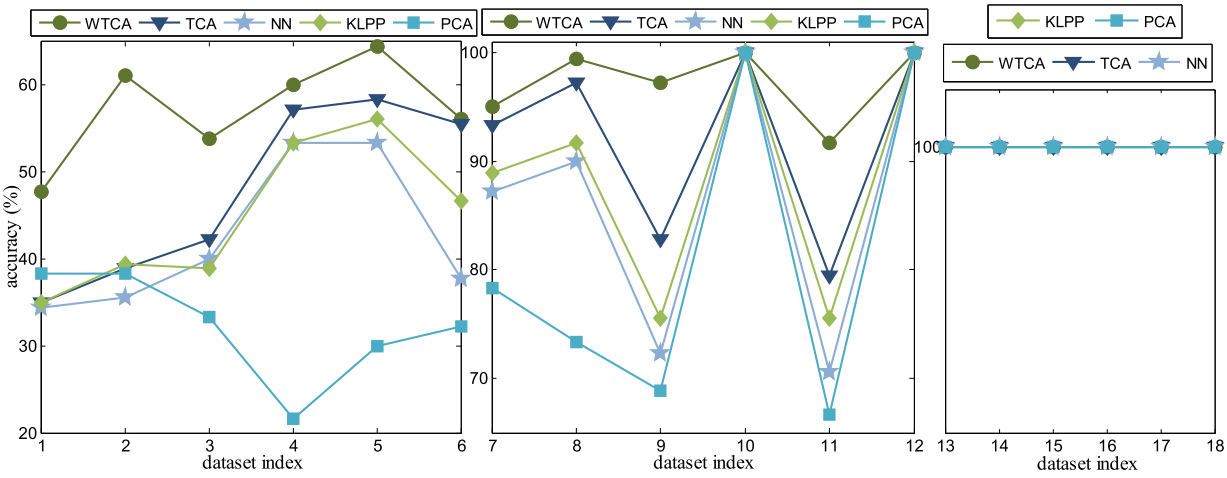

Fig. 11. Comparison of the performance of five methods on transfer task 3.

most of the samples with the same category are aligned well, while a small part of the category 2 and category 3 in the source are mixed. In Fig. 9(b), category 2 and category 3 target away from the corresponding distribution in the source domain. This observation is reasonable because TCA only considers the marginal distribution of dataset, of which the domain adaptability is limited when the conditional distributions are different across domains. In contrast, in Fig. 9(c), under the standard framework based on KLPP, the three categories in source domain are mixed and the distribution of same category source domain and target domain have considerable differences, which explains the low accuracy as shown in Table 7.

(3) Task3: designed transfer task that diagnoses fault severity levels across diverse motor loads

In task 3, the vibration signal of bearing under varying loading conditions are collected to explore the ability of the proposed transfer methods. Fig. 10 presents the kernel bandwidth λ of RBF and the embedded dimension m of domain adaptation of eighteen tasks of transfer task 3. Fig. 11 and Table 8 present the diagnosis results of those tasks.

In Fig. 11, the first figure represents the classification accuracy of three severity levels of RE fault across diverse operating conditions. The diagnosis performance of the traditional diagnosis framework and transfer framework are low on these tasks, but the transfer framework still makes some transfer improvement. Some differences could be observed in the feature distributions of source domain and target domain in these tasks. In contrast, the classification accuracy of three severity levels of IR fault across diverse operating conditions in the second figure, which have a higher diagnostic accuracy than the first one, it shows that the samples in source and target domain share a closer feature space. It is worth mentioning that compared with the standard framework, transfer learning maintained high accuracy in the

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

Table 8
Diagnosis accuracy (%) on transfer task 3 with five methods.

Methods	1	2	3	4	5	6	7	8	9	10	11	12	13–18	Average
NN	34.4	35.6	40.0	53.3	53.3	37.8	87.2	90.0	72.3	100	70.6	100	100	76.4
KLPP	35.0	39.4	38.9	53.3	56.1	46.7	88.9	91.7	75.6	100	75.6	100	100	77.8
PCA	38.3	38.3	33.3	21.7	30.0	32.2	78.3	73.3	68.9	100	66.7	100	100	71.2
TCA	35.0	38.9	42.2	57.2	58.3	55.6	93.3	97.2	82.8	100	79.4	100	100	80.0
WTCA	**47.8**	**61.1**	**53.9**	**60.0**	**64.4**	**56.1**	**95.0**	**99.4**	**97.2**	**100**	**91.7**	**100**	**100**	**84.1**

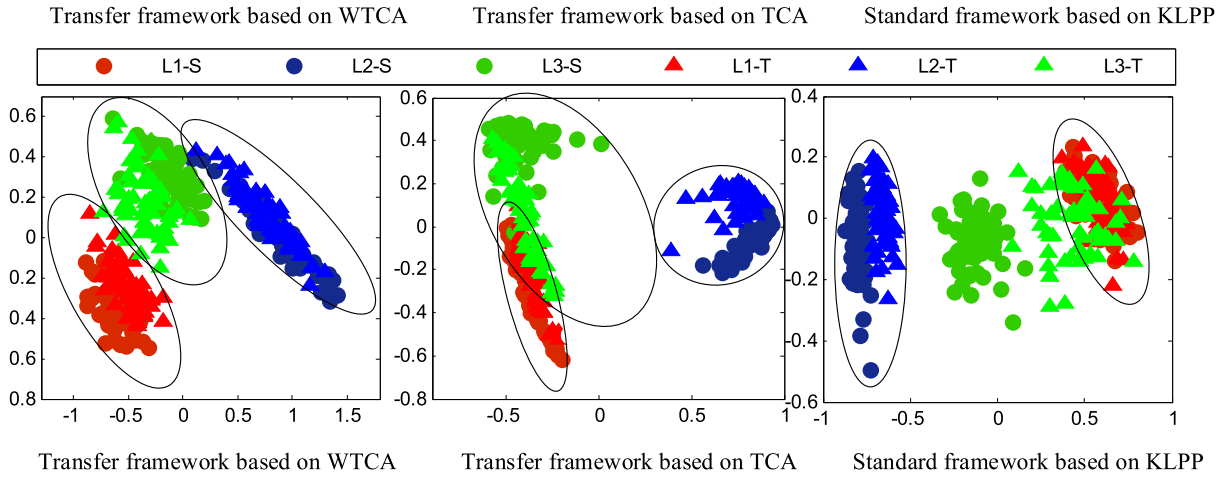

Fig. 12. Domain adaptation visualization in dataset 9: 0→2.

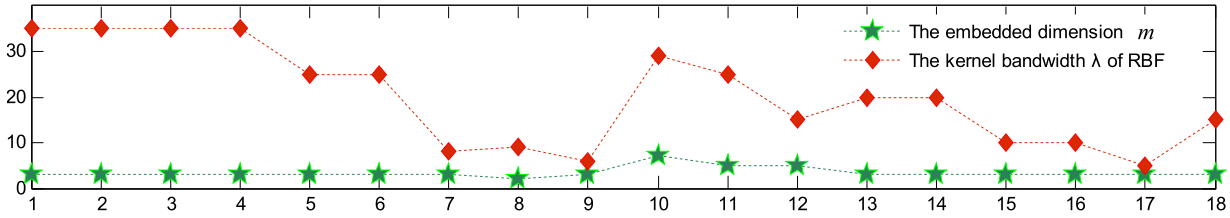

Fig. 13. Value of kernel bandwidth λ of RBF and embedded dimension m on transfer task 4.

two cases. It is interesting that all these diagnosis frameworks obtained the highest diagnostic accuracy results on the diverse severity levels classification of OR fault across diverse operating conditions, where the source and target data under different operating conditions could be substantially similar.

In this transfer task, dataset 9: 0→2 adapts the distribution across different operating conditions. In Fig. 12(a)–(c), most of the samples with different categories in source domain are separated well using all these diagnosis frameworks, while a small part of category 1 and category 3 in the target domain using TCA and KLPP are mixed. In Fig. 12(c), under the standard framework based on KLPP, category 1 in two domains and category 3 in the target domain are mixed, which explains the low accuracy as shown in Table 8.

(4) Task 4: designed transfer task that diagnoses fault types across diverse motor loads

Task 4, compared with task 3, aims to classify diverse fault types across different working operations. The kernel bandwidth λ of RBF and embedded dimension m of domain adaptation of transfer task 4 are shown in Fig. 13. Diagnosis results of twelve tasks are presented in Fig. 14 and Table 9. As can be observed in Table 9, the lowest average accuracy of these diagnosis frameworks are 77.8%, which explains that the data between the source domain and target domain are close and share similar feature spaces in these tasks. The domain adaptation ability of the transfer framework based on WTCA is demonstrated by significantly improved accuracies under different transfer tasks, the average accuracy is 96.7%, which is 14.8% higher than the diagnosis method based on KNN. Compared with KLPP and PCA, TCA has a better domain adaptation performance by considering jointly the marginal distribution of the dataset. This analysis demonstrated that the transfer framework for fault diagnosis present superior ability to the standard diagnosis framework, which shows strong feature learning and feature transferability capacity in the bearing fault diagnosis across different experimental positions.

In this transfer task, dataset 17: 1–2 is used to realize the domain adaptation across different operating condition. Fig. 15(a)–(b) show that under the transfer frameworks, the three categories samples of source and target domains are separated, and the same category samples in the source domain and target domain overlap well. In Fig. 15(c), under the standard framework based on KLPP, most of the three categories of the source and target samples are separated, while a small part of the category 1 in the source domain and category 3 in target domain are mixed. These phenomena explain the accuracies of dataset 17: 1–2 as shown in Table 9.

4.4. Transfer task across different experiment setups

To further verify the performance of the proposed method, the proposed method is applied to diagnose the datasets obtained from two different experiment setups.

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

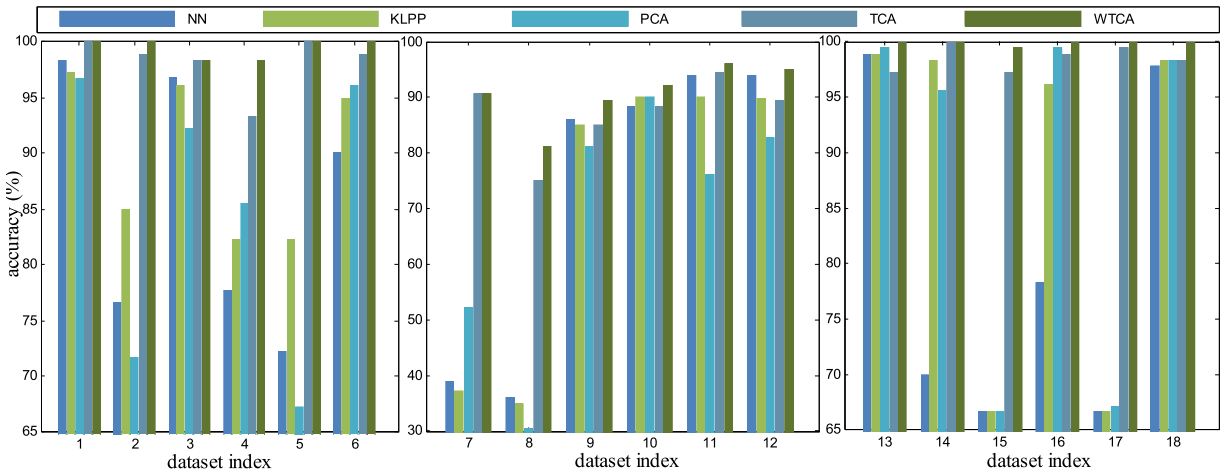

Fig. 14. Comparison of the performance of five methods on transfer task 4.

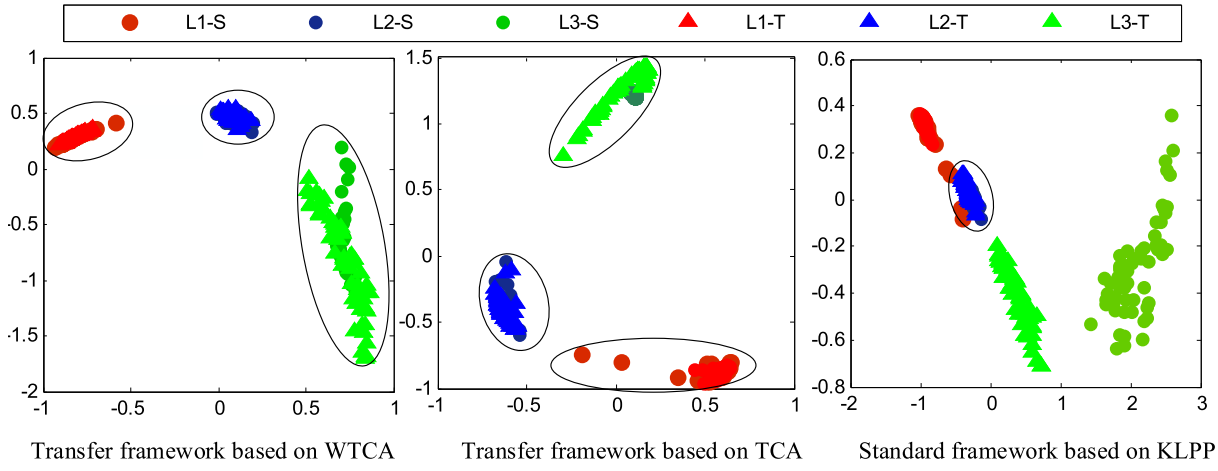

Fig. 15. Domain adaptation visualization in dataset 17: 1→2.

Table 9
Diagnosis accuracy (%) on transfer task 4 with five methods.

Methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Average
NN	98.3	76.7	96.7	77.8	72.3	90.0	38.9	36.1	86.1	88.3	93.9	93.9	98.9	98.3	66.7	96.7	66.7	98.3	81.9
KLPP	97.2	85.0	96.1	82.2	82.2	95.0	37.2	35.0	85.0	90.0	90.0	89.8	99.4	95.6	66.7	99.4	67.2	98.3	82.8
PCA	96.7	71.7	92.2	85.6	67.2	96.1	52.2	30.6	81.1	90.0	76.1	82.8	98.9	70.0	66.7	78.3	66.7	97.8	77.8
TCA	100	98.9	98.3	93.3	100	90.6	75.0	85.0	88.3	94.4	89.4	89.4	97.2	100	97.2	98.9	99.4	98.3	94.6
WTCA	**100**	**100**	**98.3**	**98.3**	**100**	**90.6**	**81.1**	**89.4**	**92.2**	**96.1**	**96.1**	**95.0**	**100**	**100**	**99.4**	**100**	**100**	**100**	**96.7**

The target domain dataset is obtained from a fault injection test of 6308 rolling bearings with the sampling frequency of 10.24 kHz, which is called to dataset 'M'. The electro-discharge machining is applied to set a single point failure to the IR, RE and OR of the test bearing with the fault diameters 0.012 inches, 0.012 inches, and 0.022 inches, respectively. The test stand shown in Fig. 16.

In this section, the dataset 'H' in transfer task 1 that contains three fault types (i.e., RE, IR, and OR) with fault diameters of 0.007 inches is regard as source domain dataset, while the target dataset 'M' consists three fault types with different fault diameters that are collected from test stand.

The tasks H → M and M → H aim to classify diverse fault types across different experiment setups. The value of kernel bandwidth λ of RBF is 1, and the embedded dimensions m of two tasks are 5 and 3, respectively. The diagnosis result is presented Table 10. in the method of WTCA, the classification accuracies of tasks H → M and M → H are 99.8% and 98.3%, respectively, which outperforms other four methods. Fig. 17 presents the visualization of three categories in source and target domains samples. In Fig. 17(a), under the transfer framework based on WTCA, the samples with same category in source and target domains is well-clustered, and most of samples with different categories are separated. In Fig. 17(b), the distribution of the samples with same category in the source and target domains are close by using TCA. Compared with KLPP and TCA, WTCA has a better domain adaptation performance by considering jointly the marginal and conditional distributions of different datasets. The analysis demonstrated that the proposed transfer framework based on WTCA can achieve high accuracy for fault diagnosis across different experimental setups.

The above results have confirmed the effectiveness of the proposed framework while handling with the fault classification tasks across the diverse domain. These fault diagnosis results

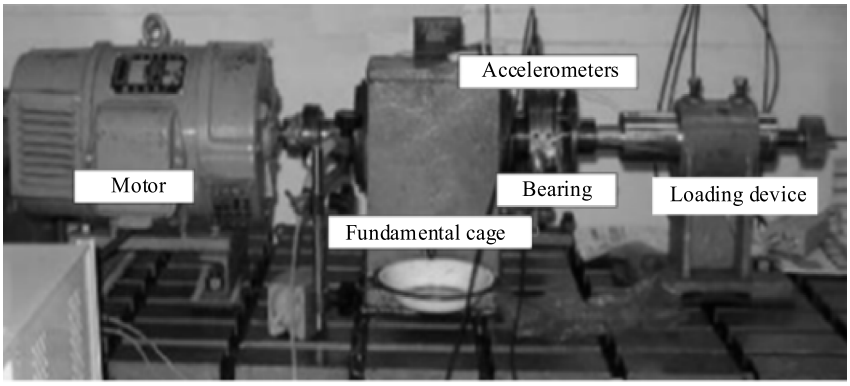

Fig. 16. The bearing test stand.

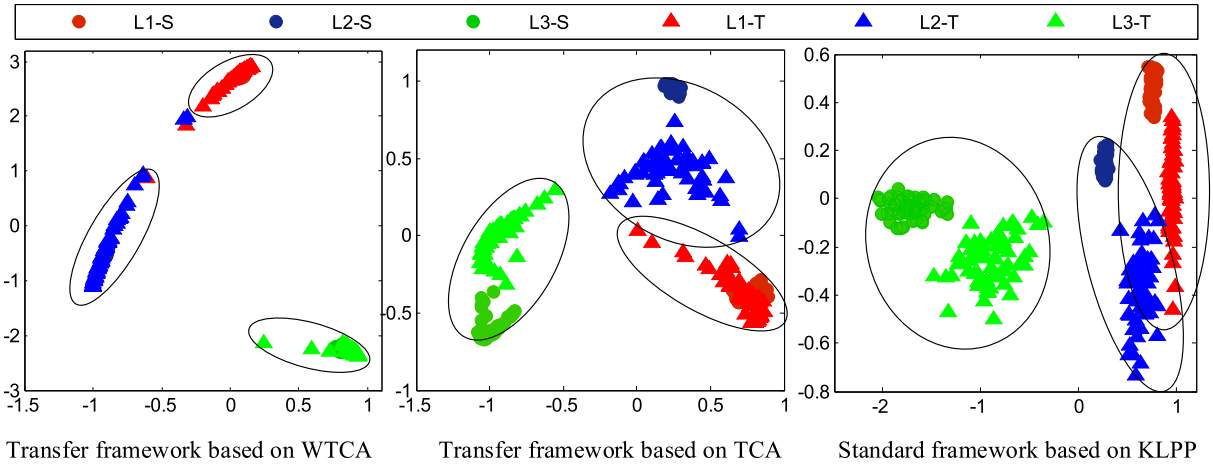

Fig. 17. Domain adaptation visualization in dataset 17: 1→2.

Table 10
Diagnosis accuracy (%) on transfer task across different experiment setups with five methods.

Transfer task	NN	KLPP	PCA	TCA	WTCA
H→M	88.9%	90.6%	76.1%	94.4%	99.8%
M→H	90.0%	91.7%	66.7%	97.2%	98.3%

are obtained from the datasets that are under a reasonable experiment condition; hence, it can be verified that the proposed diagnosis framework has potential for the fault diagnosis across diverse domain task.

Through the above analysis, the proposed fault diagnosis framework can reduce the distribution of the training data and testing data to facilitate diagnosing a new but similar testing task. In real applications, the working condition variability will result in sufficient fault data samples collected under one working condition, while the fault data sample is lacking under other working conditions. In this case, the sufficient fault data samples can be regard as source domain dataset to diagnose the fault data samples under other working conditions. When selecting a dataset as source domain dataset for diagnosis across the diverse domain, it is necessary to select the dataset with sufficient samples and similar to the target domain dataset as the source domain dataset to achieve more accurate diagnosis results.

5. Conclusion

Due to the harsh environment and variability of working conditions in real industrial scene, intelligent fault diagnosis technologies using machine learning method have limited model retraining when the distribution differs between the source domain (model is learned) and the target domain (model is applied). In practical applications, re-learning the model is time-consuming, which also has challenges when labeled data is lacking. For this issue, this paper proposed a transfer framework for bearing fault diagnosis across similar conditions, in which a novel domain adaptation method, that is, WTCA, is used to improve the domain adaption capability for the source domain and target domain. Compared with the other traditional diagnosis frameworks, the advantage of the proposed transfer framework is illustrated. Through considerable contrast tasks on four bearing datasets, the result demonstrated that the proposed transfer framework have higher accuracies than the state-of-the-art methods and its superiority over these methods is verified with the visualization.

Compared with common diagnosis algorithms, the proposed WTCA increase the domain adaptation ability between two domain data, which ensures that the learned diagnosis model from source domain datasets can be transferred effectively to new but similar applications. The transfer framework based on WTCA could be beneficial in industrial application. Therefore, some classified model can be combined and the information about the target domain can be used to improve the accuracy of the proposed framework. In the future, this method can be used to the unbalanced distribution of variable conditions of the gearbox.

Please cite this article as: P. Ma, H. Zhang, W. Fan et al., A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. ISA Transactions (2019), https://doi.org/10.1016/j.isatra.2019.08.040.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Science Foundation of China (No 51767022 and No. 51575469), the Outstanding Doctor Graduate Student Innovation Project ensemble, China (No. XJUBSCX-2016017) and the Graduate Student Innovation Project of Xinjiang Uygur Autonomous Region, China (No. XJGRI2017006).

References

[1] Qin Y. A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis. IEEE Trans Ind Electron 2018;65(3):2716–26.
[2] Glowacz A, Glowacz W, Glowacz Z, Kozik, J. Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. Measurement 2018;113:1–9.
[3] Fei CW, Choy YS, Bai GC, Tang GC. Multi-feature entropy distance approach with vibration and acoustic emission signals for process feature recognition of rolling element bearing faults. Struct Health Monit 2018;17(2):156–68.
[4] Jia XD, Jin C, Buzza M, Di Y. A deviation based assessment methodology for multiple machine health patterns classification and fault detection. Mech Syst Signal Process 2018;99:244–61.
[5] Gao ZW, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. IEEE Trans Ind Electron 2015;62(6):3757–67.
[6] Su ZQ, Tang BP, Ma JH, Deng L. Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier. Measurement 2014;48:136–48.
[7] Jia F, Lei YG, Lin J, Zhou X. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 2016;72:303–15.
[8] Moosavian A, Khazaee M, Ahmadi H, Khazaee M. Fault diagnosis and classification of water pump using adaptive neuro-fuzzy inference system based on vibration signals. Struct Health Monit 2015;14(5):402–10.
[9] De Moura EP, Souto CR, Silva AA, Irmão MAS. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. Mech Syst Signal Process 2011;25(5):1765–72.
[10] Wang TZ, Qi J, Xu H, Wang YD. Fault diagnosis method based on FFT-RPCA-SVM for cascaded-multilevel inverter. ISA Trans 2016;60:156–63.
[11] Wong PK, Yang ZX, Vong CM, Zhong JH. Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. Neurocomputing 2014;128:249–57.
[12] Liu ZW, Guo W, Hu JH, Ma WS. A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM. ISA Trans 2017;66:249–61.
[13] Tang BP, Song T, Li F, Deng L. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. Renew Energy 2014;62:1–9.
[14] Wang X, Zheng Y, Zhao ZZ, Wang JP. Bearing fault diagnosis based on statistical locally linear embedding. Sensors 2015;15(7):16225–47.
[15] Li ZX, Yan XP, Jiang Y, Li Q. A new data mining approach for gear crack level identification based on manifold learning. Mechanics 2012;18(1):29–34.
[16] Xue XM, Zhou JZ. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. ISA Trans 2017;66:284–95.
[17] Sun C, Zhang ZS, He ZJ, Chen BQ. Novel method for bearing performance degradation assessment—A kernel locality preserving projection-based approach. Inst. Mech. Eng. C 2014;228(3):548–60.
[18] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59.
[19] Khatami A, Babaie M, Tizhoosh HR, Khosravi A. A sequential search-space shrinking using CNN transfer learning and a radon projection pool for medical image retrieval. Expert Syst Appl 2018;100:224–33.
[20] Qureshi AS, Khan A, Zameer A, Usman A. A wind power prediction using deep neural network based meta regression and transfer learning. Appl Soft Comput 2017;58:742–55.
[21] Mun S, Shin M, Shon S, Kim W. DNN transfer learning based non-linear feature extraction for acoustic event classification. IEICE Trans Inf Syst 2017;100(9):2249–52.
[22] Long MS, Wang JM, Ding GG, Sun JG, Yu PS. Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE international conference on computer vision. 2013. p. 2200–07.
[23] Ghifary M, Kleijn WB, Zhang MJ, Balduzzi D, W. Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In: 14th European conference on computer vision 2016 Oct 11–14. Amsterdam, Netherlands: Springer Cham; 2016, p. 597–613.
[24] Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: The IEEE conference on computer vision and pattern recognition 2017.
[25] Lu WN, Liang B, Cheng Y, Meng DS, Yang J, Zhang T. Deep model based domain adaptation for fault diagnosis. IEEE Trans Ind Electron 2017;64(3):2296–305.
[26] Wen L, Gao L, Li XY. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Trans Syst Man Cybern Syst 2017;49(1):136–44.
[27] Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 2011;22(2):199–210.
[28] Wei Y, Zhang Y, Yang Q. Learning to Transfer. arXiv preprint arXiv:1708.05629, 2017.
[29] Li YF, Liang XH, Lin JH, Chen YJ, Liu JX. Train axle bearing fault detection using a feature selection scheme based multi-scale morphological filter. Mech Syst Signal Process 2018;101:435–48.

Ping Ma received the PHD degree from the school of electrical engineering, Xinjiang University, Urumqi, China. Her current research interests include control, dynamics and fault diagnosis.

Hongli Zhang received the PhD degree from Beijing Institute of Technology in 2009. He is currently a professor in Xinjiang University, Urumqi, China. His current research interests include control, dynamics and fault diagnosis.

Wenhui Fan received the PhD degree from Zhejiang University in 1998. He is currently a professor in Tsinghua University, Beijing, China. His current research interests include control, dynamics and fault diagnosis.

Cong Wang currently a PhD candidate at Xinjiang University, Urumqi, China. Her research interests include machine learning, intelligent computing, and system modeling and control.