ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/328547669

Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data

Article *in* IEEE Transactions on Industrial Electronics · October 2018

DOI: 10.1109/TIE.2018.2877090

CITATIONS
212

READS
1,780

5 authors, including:

Liang Guo
Southwest Jiaotong University
29 PUBLICATIONS **1,374** CITATIONS
SEE PROFILE

Yaguo Lei
Xi'an Jiaotong University
130 PUBLICATIONS **9,696** CITATIONS
SEE PROFILE

Saibo Xing
Xi'an Jiaotong University
13 PUBLICATIONS **1,271** CITATIONS
SEE PROFILE

Tao Yan
Xi'an Jiaotong University
12 PUBLICATIONS **758** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Remaining Useful Life prediction based on adaptive feature representation View project

Project Intelligent fault diagnosis View project

All content following this page was uploaded by Liang Guo on 09 March 2020.

The user has requested enhancement of the downloaded file.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data

Liang Guo, Yaguo Lei, *Member, IEEE,* Saibo Xing, Tao Yan, and Naipeng Li

***Abstract*—The success of intelligent fault diagnosis of machines relies on two conditions. 1) Labeled data with fault information are available. 2) The training and testing data are drawn from the same probability distribution. However, for some machines, it is difficult to obtain massive labeled data. Moreover, even though labeled data can be obtained from some machines, the intelligent fault diagnosis method trained with such labeled data possibly fails in classifying unlabeled data acquired from the other machines due to data distribution discrepancy. These problems limit the successful applications of intelligent fault diagnosis of machines with unlabeled data. As a potential tool, transfer learning adapts a model trained in a source domain to its application in a target domain. Based on transfer learning, we propose a new intelligent method named deep convolutional transfer learning network (DCTLN). DCTLN consists of two modules: condition recognition and domain adaptation. The condition recognition module is constructed by a one-dimensional convolutional neural network (CNN) to automatically learn features and recognize health conditions of machines. The domain adaptation module facilitates the one-dimensional CNN to learn domain invariant features by maximizing domain recognition errors and minimizing probability distribution distance. The effectiveness of the proposed method is verified using six transfer fault diagnosis experiments.**

***Index Terms*—Transfer learning, intelligent fault diagnosis, Convolutional neural network, bearing.**

Manuscript received February 26, 2018; revised May 26, 2018 and August 26, 2018; accepted September 30, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant U1709208 and Grant 61673311, and in part by the Young Talent Support plan of Central Organization Department.

L. Guo, S. Xing, T. Yan, and N. Li are with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China. (email: guoliang@mail.xjtu.edu.cn; xingsaibo@stu.xjtu.edu.cn; yantao_0421@163.com; li3112001096@stu.xjtu.edu.cn).

Y. Lei is with the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China. (Corresponding author contacts, phone: +86 029 8339 5041, e-mail: yaguolei@mail.xjtu.edu.cn).

I. Introduction

INTELLIGENT fault diagnosis is able to handle massive monitoring data and distinguish health conditions of machines [1-3]. In recent years, various intelligent fault diagnosis methods have been studied. Cerrada *et al.* [4] conditions of a gearbox. The proposed method was validated by vibration signals acquired from a gearbox experiment platform which simulated six kinds of gear faults. Rauber *et al.* [5] used an intelligent model to distinguish health conditions of rotating machines. The performance of the used method was validated using a labeled bearing dataset from Case Western Reserve University (CWRU). Lee *et al.* [6] built a convolutional neural network (CNN) to recognize health conditions of machines. An experiment dataset including six kinds of health conditions of chemical vapor deposition machines was applied to demonstrate the effectiveness of the proposed method. Su *et al.* [7] proposed a least square support vector machine based multi-fault diagnosis method. The proposed method was verified via experimental data acquired from a roller bearing fault experiment rig.

Through the literature review, it can be seen that the datasets applied to validate the effectiveness of various intelligent fault diagnosis methods satisfy the following two conditions. 1) Labeled data with fault information are available. 2) The training and testing data are drawn from the same probability distribution. As a matter of fact, it is demonstrated in [8-10] that the success of intelligent fault diagnosis of machines relies on those two conditions. However, for some machines, it is difficult to satisfy those two conditions due to the following problems.

1) Labeled fault data are difficult to be obtained from some machines [11]. Specifically speaking, there are two main reasons causing the lack of labeled fault data. Firstly, machines may not be allowed to run to failure since an unexpected fault usually leads to the breakdown of machines or even catastrophic accidents. In such cases, fault data are impossible to be obtained. Secondly, machines generally go through a long degradation process from health to failure. It means that it is time-consuming and expensive to obtain fault data of machines [12].

2) An intelligent fault diagnosis method trained with labeled data acquired from one machine possibly fails in classifying

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

unlabeled data acquired from the other machines. Although massive labeled data are difficult to be obtained for some machines, they can still be obtained from different but related machines. For example, labeled data of railway locomotive bearings are difficult to be obtained, while the labeled data of motor bearings can be relatively easier to be obtained. However, probability distributions of data acquired from different machines are different [13]. Therefore, the classification performance of intelligent fault diagnosis methods degenerates when the training and testing datasets are acquired from different machines.

The above two problems limit the successful applications of intelligent fault diagnosis of machines where no labeled data is available. To promote the successful applications of intelligent fault diagnosis of machines with unlabeled data, a new intelligent method is in urgent need. For such method, the model trained with labeled data acquired from one machine is able to be generalized to the unlabeled data acquired from the other machines. Transfer learning is able to use the learned knowledge from the source domain to solve a new but related task in the target domain [14, 15]. It is expected to solve the problem that there are no sufficient labeled data to train a reliable intelligent model. An intuitive and commonly-used idea of transfer learning is to obtain a feature representation where the different domains are close to each other while keeping good classification performance on the source data [16-18]. Deep learning is able to learn deep hierarchical representation of the data, which is considered to provide the cross-domain invariant features for transfer learning. Transfer leaning based on these cross-domain invariant features can effectively reduce the discrepancy between the source and the target domains. Based on this idea, deep transfer learning based methods present successful applications in various tasks [19-22].

Currently, several transfer learning based intelligent fault diagnosis methods have been proposed [13, 23, 24]. Lu *et al.* [13] presented a deep model based domain adaptation method for machine fault diagnosis. A gearbox dataset collected under different operation conditions was used to test the performance of the proposed method. Wen *et al.* [23] set up a new deep transfer learning method for fault diagnosis. The validation dataset was acquired from a bearing testbed operating under different loading conditions. Xie *et al.* [24] proposed a transfer analysis based gearbox fault diagnosis method. The performance of the presented method was verified by a gearbox dataset obtained under various operation conditions.

It can be observed from those works that the existing transfer learning based intelligent fault diagnosis methods mainly focus on the transfer between different operation conditions. Those studies demonstrate that transfer learning allows the intelligent fault diagnosis methods to be applicable across datasets acquired from a machine under different operation conditions. Actually, in real-world applications, the massive labeled data are difficult to be obtained from some machines. Therefore, the transfer fault diagnosis between machines is essential and crucial. Only if the transfer fault diagnosis between machines works well, the health conditions of machines with unlabeled data can be recognized by the intelligent method trained with labeled data acquired from the other machines.

Therefore, in order to promote the successful applications of intelligent fault diagnosis of machines with unlabeled data, we propose a new deep transfer learning method named deep convolutional transfer learning network (DCTLN). DCTLN consists of two modules: condition recognition and domain adaptation. The condition recognition module seeks to automatically learn features and accurately recognize health conditions of machines. The domain adaptation module includes a domain classifier and a domain distribution discrepancy metrics term, which makes the learned features to be domain invariant. With those two modules, DCTLN trained with labeled data acquired from one machine is expected to effectively classify the unlabeled data acquired from the other machines. Using three bearing datasets acquired from different machines, six transfer fault diagnosis experiments are conducted to demonstrate the effectiveness of the proposed method. The results indicate that the proposed method improves the recognition accuracies of bearing health conditions about 32.1% compared with traditional methods without transfer learning.

The main insights and contributions of this paper are summarized as follows.

1) A new deep transfer learning method is proposed. The proposed method includes a condition recognition module and a domain adaptation module. In the condition recognition module, a one-dimensional CNN is constructed to learn features from the raw vibration data, and then the health condition classifier can classify the samples based on these features. In the domain adaptation, a domain classifier and a distribution discrepancy metrics are built to help learn domain invariant features.

2) The intelligent fault diagnosis for machines with unlabeled data is explored. Generally, the training and testing dataset should be acquired from one machine for an intelligent fault diagnosis method. In this paper, we explore a new scenario of intelligent fault diagnosis, where the training and testing dataset are acquired from different machines and the data from the monitoring machines are unlabeled. This exploration would promote the practical application of intelligent fault diagnosis.

The rest of this paper is organized as follows. Section II details the transfer learning problem. Section III is the proposed method. In Section IV, the transfer fault diagnosis experiments between three bearing datasets are conducted. Finally, conclusions are drawn in Section V.

II. Transfer Learning Problem

In order to clearly state the problem to be solved, a basic notation on transfer learning, domain, is first introduced. Let χ be a feature space, X be a particular sample and $P(X)$ be a marginal probability distribution. Then a domain is defined by $D=\{\chi, P(X)\}$, where $X=\{x_1, x_2, ..., x_n\} \in \chi$, and x_i is the ith feature term [14]. As shown in Fig. 1, traditional intelligent methods are trained using labeled data, and tested on

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

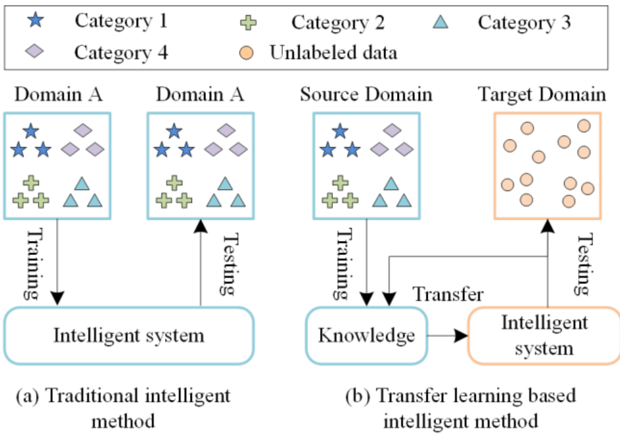


Fig. 1. Intelligent learning systems. (a) Traditional intelligent method. (b) Transfer learning based intelligent method.

the future data with the same feature space and probability distribution. It indicates that the training and testing data come from the same domain. Transfer learning, however, allows the distributions of the training and testing data to be different [14]. Generally, given a source domain D_{S} with labeled training data, and a target domain D_{T} with unlabeled data, transfer learning aims to improve the capacity of the target predictive function in the domain D_{T} using the knowledge learned from the domain D_{S}.

In this paper, an intelligent fault diagnosis method trained with labeled data acquired from one machine is expected to recognize health conditions of other machines with unlabeled data. In other words, the intelligent method is able to accomplish transfer fault diagnosis between machines. More specifically, we denote the labeled data acquired from one machine as the source domain which is referred to as $D_{\mathrm{S}} = \{\{x_{Si}, y_{Si}\}\}$, where $x_{Si} \in \chi_{\mathrm{S}}$ is a data sample and $y_{Si} \in Y_{\mathrm{S}}$ is its corresponding label. Similarly, we denote the unlabeled data acquired from the other machines as the target domain which is $D_{\mathrm{T}} = \{(x_{T_i})\}$, where a data sample x_{Ti} is in χ_{T}. Note that, in this paper, it is assumed that the source and target domains share the same label space. In other words, the source and target domains only differ in their respective data probability distributions. In order to obtain the desirable result of this paper, the proposed intelligent method should be able to learn domain invariant features. Learning Domain invariant features means that the features should be subject to the same or almost the same distribution, no matter the source domain data or target domain data they are learned from. If the features are domain invariant, then the health condition classifier trained with the source domain data is able to effectively classify the features learned from the target domain data. Therefore, learning domain invariant features is a critical procedure for accomplishing transfer fault diagnosis between machines.

III. SUBMISSION OF REVISED MANUSCRIPT FOR REVIEW

In this section, we detail the architecture and training process of the proposed method.

A. Deep convolutional transfer learning network

As shown in Fig. 2, the proposed DCTLN consists of two modules: condition recognition and domain adaptation. Condition recognition is achieved by a one-dimensional CNN. The one-dimensional CNN includes a feature extractor and a health condition classifier. The feature extractor seeks to automatically learn features, and the health condition classifier recognizes health conditions based on the extracted features. Domain adaptation is completed by a domain classifier and a distribution discrepancy metrics. The domain adaptation module is connected to the feature extractor to help the one-dimensional CNN learn domain invariant features.

1) **Condition recognition**. Condition recognition is achieved by a one-dimensional CNN with sixteen layers including one input layer, six convolutional layers, six pooling layers, two fully-connected layers and one output layer. The detailed information of the one-dimensional CNN can be found in Table I. Among those sixteen layers, the first fifteen layers can be thought of a feature extractor, and the last layer is seen as a health condition classifier.

The input layer is built by input vibration signals with a length of L. The next one is the convolutional layer. In the convolutional layer, the convolution kernel convolutes with input data to learn features. Since vibration signals are one-dimensional, the convolution is designed as a one-dimensional operation. Concretely, the one-dimensional convolution is to take a dot product between a kernel $\boldsymbol{w}_{\mathrm{c}} \in R^m$ and the jth segmented signal $s^i_{j-m+1:j} \in R^m$ to obtain convolution features,

$$c_j = \mathrm{Relu}\left(\sum_{i=1}^{n} \mathbf{w}_{\mathrm{c}} * s^i_{j-m+1:j} + b_{\mathrm{c}}\right), \tag{1}$$

where $*$ is a one-dimensional convolution operator, $\boldsymbol{w}_{\mathrm{c}}$ is referred to as the convolution kernel, b_{c} is the corresponding bias, n is the number of kernels, and c_j is the jth output point of the convolutional layer. $\mathrm{Relu}(\cdot)$ is an activation function.

Each convolutional layer is connected with a pooling layer, in which a pooling operation is conducted to reduce the dimension of convolution features and to enable the learned convolution features to be shift invariant. The max pooling function is utilized in the paper, which returns the maximum value within a certain sub-region as follows,

$$p_j = \max\{c_{j\times k:(j+1)\times k}\}, \tag{2}$$

where k is the pooling length, and p_j is the pooling output of the jth point.

After six convolution and pooling operations, the input vibration signals are mapped into features in layer P6. Then, the first fully-connected layer FC1 is flatten from the output of layer P6. The second fully-connected layer FC2 is calculated as follows,

$$f = \sigma((\boldsymbol{w}_{\mathrm{f}})^{\mathrm{T}} s_{\mathrm{m}} + b_{\mathrm{f}}), \tag{3}$$

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

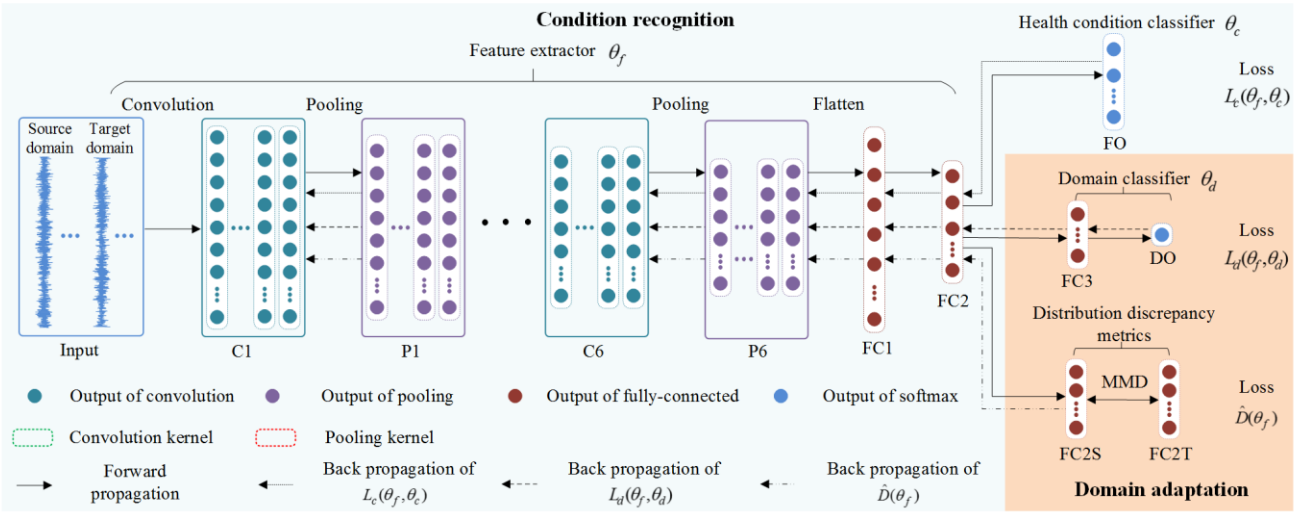


Fig. 2. The structure illustration of the proposed method.

where $\boldsymbol{w}_{\mathrm{f}}$ is the weight matrix connecting two fully-connected layers, b_{f} is the corresponding bias vector, and s_{m} is the input data. Based on the output $f2$ of layer FC2, the health conditions of machines are estimated in the health condition output layer FO through the softmax regression,

$$y=\frac{1}{\sum_{i=1}^{K}e^{((\boldsymbol{w}_i)^{\mathrm{T}}f2+b)}}\begin{bmatrix}e^{((\boldsymbol{w}_1)^{\mathrm{T}}f2+b)}\\ e^{((\boldsymbol{w}_2)^{\mathrm{T}}f2+b)}\\ \vdots\\ e^{((\boldsymbol{w}_k)^{\mathrm{T}}f2+b)}\end{bmatrix}, \tag{4}$$

where $\boldsymbol{w}_i$ is the weight matrix connecting to the ith output neuron, b is the corresponding bias vector, and K denotes the health condition categories.

2) **Domain adaptation**. The domain adaptation module includes a domain classifier and a domain distribution discrepancy metrics.

TABLE I
THE ARCHITECTURE OF THE ONE-DIMENSIONAL CNN

Layer	Symbol	Operator	Parameter size
1	Input	Input signal	L
2	C1	Convolution	$m\times1\times n$
3	P1	Pooling	k
4	C2	Convolution	$m\times n\times n$
5	P2	Pooling	k
6	C3	Convolution	$m\times n\times n$
7	P3	Pooling	k
8	C4	Convolution	$m\times n\times n$
9	P4	Pooling	k
10	C5	Convolution	$m\times n\times n$
11	P5	Pooling	k
12	C6	Convolution	$m\times n\times n$
13	P6	Pooling	k
14	FC1	Fully-connected	/
15	FC2	Fully-connected	/
16	FO	Softmax	/

As shown in Fig. 2, the domain classifier includes two layers: a fully-connected layer FC3 and a domain discrimination output layer DO. The FC3 layer is calculated by (3), where the input data s_{m} is the output of layer FC2. The domain discrimination output layer DO is a binary classifier setting with logistics regression,

$$d=\frac{1}{1+e^{-((\boldsymbol{w}_{\mathrm{d}})^{\mathrm{T}}f3+b_{\mathrm{d}})}}, \tag{5}$$

where $\boldsymbol{w}_{\mathrm{d}}$ is the weight matrix in domain adaptation module, b_{d} is the corresponding bias vector, and $f3$ is the output of layer FC3.

Let $P(f2^{(\mathrm{S})})$ and $Q(f2^{(\mathrm{T})})$ be the probability distributions of FC2S and FC2T, where FC2S and FC2T are the output of layer FC2 with the source domain and target domain data, respectively. The distance between $P(f2^{(\mathrm{S})})$ and $Q(f2^{(\mathrm{T})})$ is calculated by the maximum mean discrepancy (MMD) as follows,

$$D=\left\|\frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}}f2_i^{(\mathrm{S})}-\frac{1}{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}}f2_j^{(\mathrm{T})}\right\|_{\mathrm{H}}^{2}, \tag{6}$$

where n_{s} is the number of training samples from source domain, n_{t} is the number of training samples from target domain, and $\|\cdot\|_{\mathrm{H}}$ is a reproducing kernel Hilbert space (RKHS).

A. Optimization objective

The proposed DCTLN has three optimization objects. 1) Minimize the health condition classification error on the source

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

domain dataset. 2) Maximize the domain classification error on the source and target domain dataset. 3) Minimize the MMD distance between the source and target domain dataset.

1) **Object 1**: In order to accomplish transfer fault diagnosis, DCTLN should be able to recognize health conditions and learn domain invariant features. Specifically, the condition recognition module is designed to recognize health conditions of machines. Therefore, the first optimization object of DCTLN is to minimize the health condition classification error on the source domain data. For a dataset with k health condition categories, the desired objective function can be defined as a standard softmax regression loss,

$$L_c = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} I[y_i = k]\log\frac{e^{((\boldsymbol{w}_j)^T f2+b)}}{\sum_{l=1}^{k} e^{((\boldsymbol{w}_l)^T f2+b)}}\right], \tag{7}$$

where m is the batch size of the training samples, k is the fault categories, and $I[\cdot]$ is an indicator function.

2) **Object 2**: The domain adaptation module is designed to learn domain invariant features. The domain adaptation module includes a domain classifier and a distribution discrepancy metrics. As shown in Fig. 2, the domain classifier is connected with the feature extractor. Based on the theory suggestion on domain adaptation [15], if a domain classifier cannot discriminate features between the source and target domain, the features are domain invariant. Therefore, the second optimization object of DCTLN is to maximize the domain classification error on the source and target domain data. The domain classification loss is defined as,

$$L_d = \frac{1}{m}\sum_{i=1}^{m}(g_i \log d(x_i) + (1-g_i)\log(1-d(x_i))) , \tag{8}$$

where g_i is the ground truth domain label, and $d(x_i)$ denotes the output domain for ith sample, which indicates whether x_i come from the source domain or the target domain. In the training stage, the training dataset is constructed by n_{s} source domain data samples and n_{t} target domain data samples. Therefore, (8) can be rewritten as follows,

$$L_d = \frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} L_d(f2_i^{(\mathrm{S})}) + \frac{1}{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}} L_d(f2_j^{(\mathrm{T})}) , \tag{9}$$

where $f2_i^{(\mathrm{S})}$ and $f2_j^{(\mathrm{T})}$ is the high-level learned features from the source domain data and the target domain data, respectively.

3) **Object 3**: In the proposed DCTLN, the output of layer FC2 is the high-level features which connect to the health condition classifier. In other words, the high-level features directly influence the effectiveness of transfer fault diagnosis. In order to reduce the distribution discrepancy distance between features learned from different domains, the distribution discrepancy distance between the FC2S and FC2T is directly measured. Therefore, the third optimization object of DCTLN is to minimize the distribution discrepancy distance between the source and target domain data. To calculate the distribution distance of high-level learned features between different domains, the practical computation of MMD is written as,

$$\begin{aligned}\hat{D} = &\frac{1}{n_{\mathrm{s}}^2}\sum_{i=1}^{n_{\mathrm{s}}}\sum_{j=1}^{n_{\mathrm{s}}} k(f2_i^{(\mathrm{S})}, f2_j^{(\mathrm{S})}) \\ &+\frac{1}{n_{\mathrm{t}}^2}\sum_{i=1}^{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}} k(f2_i^{(\mathrm{T})}, f2_j^{(\mathrm{T})}) \ , \\ &-\frac{2}{n_{\mathrm{s}} n_{\mathrm{t}}}\sum_{i=1}^{n_{\mathrm{s}}}\sum_{j=1}^{n_{\mathrm{t}}} k(f2_i^{(\mathrm{S})}, f2_j^{(\mathrm{T})})\end{aligned} \tag{10}$$

where $\hat{D}$ is the unbiased estimation of $D(P,Q)$, and $k(\cdot,\cdot)$ is a kernel function. According to [25], the Gaussian radial basis function (RBF), i.e., $k(x,y) = \exp(\|x-y\|^2 / 2\sigma^2)$, maps the original features to a higher dimensional space, and is widely used for MMD calculation. Therefore, in this paper, RBF is chosen to estimate the MMD between domains.

Combining those three optimization objects, the final optimization object can be written as,

$$L = L_c - \lambda L_d + \mu\hat{D}, \tag{11}$$

where the hyper-parameters λ and μ determine how strong the domain adaptation is.

Once the optimization object of the propose method is built, it is convenient to train the proposed method by stochastic gradient descent (SGD) algorithm. As shown in Fig. 2, let θ_f, θ_c and θ_d be the parameters of the feature extractor, health condition classifier and domain classifier, respectively. The loss function equation (11) is rewritten as follows,

$$L(\theta_f^*, \theta_c^*, \theta_d^*) = \min_{\theta_f, \theta_c, \theta_d} L_c(\theta_f, \theta_c) - \lambda L_d(\theta_f, \theta_d) + \mu\hat{D}(\theta_f) \tag{12}$$

Based on the equation (12) and SGD algorithm, the parameters θ_f, θ_c and θ_d are updated as follows,

$$\theta_f \leftarrow \theta_f - \varepsilon\left(\frac{\partial L_c}{\partial \theta_f} - \lambda\frac{\partial L_d}{\partial \theta_f} + \mu\frac{\partial \hat{D}}{\partial \theta_f}\right), \tag{13}$$

$$\theta_c \leftarrow \theta_c - \varepsilon\frac{\partial L_c}{\partial \theta_c}, \tag{14}$$

$$\theta_d \leftarrow \theta_d - \varepsilon\frac{\partial L_d}{\partial \theta_d}, \tag{15}$$

where ε is the learning rate.

When the training process is completed, the health condition classifier is able to correctly classify unlabeled samples in the target domain if the learned features are subject to ambiguous domain categories and small domain discrepancy. For the testing process of DCTLN, the input of DCTLN is unlabeled data from the target domain. DCTLN learns domain invariant features from these data first. Then the health condition classifier predicts the health condition according to the learned domain invariant features.

IV. Experiment Results and Comparisons

B. Dataset

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

As discussed in Section Ⅰ, in order to promote the successful applications of intelligent fault diagnosis of machines with unlabeled data, the intelligent fault diagnosis method trained with labeled data acquired from one machine is expected to classify unlabeled data acquired from the other machines effectively. Therefore, in this section, three datasets acquired from three different but related machines are used to conduct six transfer fault diagnosis experiments.

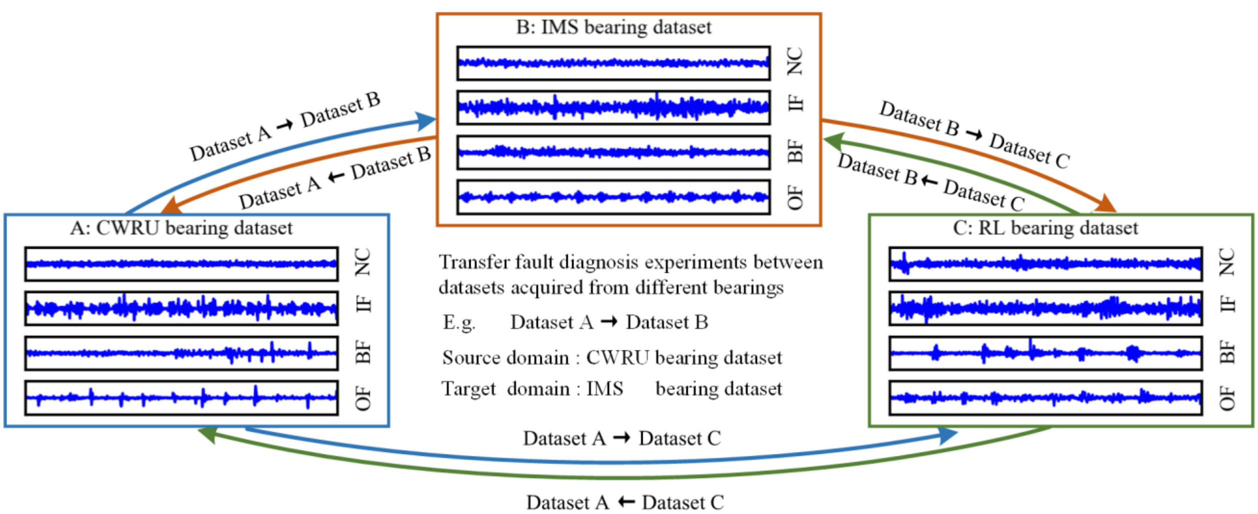
A: CWRU bearing dataset. CWRU bearing dataset was collected from an experiment platform provided by CWRU [26]. On this experiment platform, experiments were conducted using an electric motor, and vibration data were measured from the motor bearings. Faults were introduced separately at the inner raceway, rolling element and outer raceway of bearings. The CWRU bearing dataset is composed of vibration signals acquired from the above-mentioned health conditions. Each health condition, i.e., normal condition (NC), inner race fault (IF), outer race fault (OF) and ball fault (BF), has 1000 data samples. Thus, this dataset consists of 4000 data samples, and the length of each sample is 1200.

B: IMS bearing dataset. IMS bearing dataset was generated by the Center for Intelligent Maintenance Systems (IMS), which was also collected from a bearing experiment platform [27]. Four double row bearings were installed on a shaft. On the bearing housings, accelerometers were installed to acquire vibration signals. Through several run-to-failure experiments, outer race failures, inner race failures and ball failures occurred in three bearings, respectively. We choose the data collected from failure conditions to construct IMS bearing dataset. Therefore, this dataset includes three types of failure conditions and one type of normal condition. A total of 4,000 data samples are in this dataset, and the length of each sample is also 1200.

C: RL bearing dataset. RL bearing dataset was acquired from real railway locomotive (RL) rolling element bearings [1]. The RL bearing was installed on a test bench. An accelerometer was mounted on the outer race of the test bearing to measure its vibration. The health condition types and number of data samples are the same as those of the CWRU and IMS bearing dataset.

TABLE II
VARIOUS BEARING DATASETS

Name	Bearing	Condition	Speed	Load
A	Motor bearing	NC	1750 r/min	2 HP
		IF	1750 r/min	2 HP
		OF	1750 r/min	2 HP
		BF	1750 r/min	2 HP
B	Shaft support bearing	NC	2000 r/min	26.6 kN
		IF	2000 r/min	26.6 kN
		OF	2000 r/min	26.6 kN
		BF	2000 r/min	26.6 kN
C	Railway locomotive bearing	NC	650 r/min	≈ 9.6 kN
		IF	500 r/min	≈ 9.6 kN
		OF	480 r/min	≈ 9.6 kN
		BF	630 r/min	≈ 9.6 kN



Transfer fault diagnosis experiment	Training dataset		Testing dataset
Dataset A → Dataset B	Labeled dataset A: 100%	Unlabeled dataset B: 50%	dataset B: 50%
Dataset B → Dataset A	Labeled dataset B: 100%	Unlabeled dataset A: 50%	dataset A: 50%
Dataset A → Dataset C	Labeled dataset A: 100%	Unlabeled dataset C: 50%	dataset C: 50%
Dataset C → Dataset A	Labeled dataset C: 100%	Unlabeled dataset A: 50%	dataset A: 50%
Dataset B → Dataset C	Labeled dataset B: 100%	Unlabeled dataset C: 50%	dataset C: 50%
Dataset C → Dataset B	Labeled dataset C: 100%	Unlabeled dataset B: 50%	dataset B: 50%

Fig. 3. The six transfer fault diagnosis experiments.

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

TABLE III
RECOGNITION RESULTS OF VARIOUS METHODS

Method	A→B	B→A	A→C	C→A	B→C	C→B	Average
CNN	0.551 ± 0.010	0.571 ± 0.014	0.508 ± 0.009	0.463 ± 0.065	0.506 ± 0.008	0.592 ± 0.018	0.531
TCA [29]	0.288 ± 0.031	0.310 ± 0.053	0.301 ± 0.090	0.322 ± 0.098	0.303 ± 0.063	0.305 ± 0.006	0.304
DAFD [30]	0.383 ± 0.024	0.479 ± 0.008	0.398 ± 0.024	0.540 ± 0.009	0.369 ± 0.016	0.465 ± 0.029	0.439
DDC [25]	0.799 ± 0.086	0.872 ± 0.013	0.716 ± 0.090	0.691 ± 0.071	0.760 ± 0.009	0.703 ± 0.054	0.756
DANN [28]	0.759 ± 0.021	0.880 ± 0.005	0.773 ± 0.024	0.737 ± 0.013	0.769 ± 0.015	0.813 ± 0.046	0.788
DCTLN (Ours)	**0.897 ± 0.018**	**0.899 ± 0.009**	**0.873 ± 0.014**	**0.837 ± 0.010**	**0.824 ± 0.012**	**0.848 ± 0.020**	**0.863**

The vibration signals of those three datasets are shown in Fig. 3, and their detailed information are displayed in Table II. Those three datasets are all bearing vibration signals, while they are acquired from different machines and different operation conditions.

C. Transfer fault diagnosis of DCTLN

As shown in Fig. 3, we evaluate the proposed DCTLN on six transfer fault diagnosis experiments, i.e., A→B, B→A, A→C, C→A, B→C, and C→B. In each transfer fault diagnosis experiment, the part before the arrow represents the source domain, and that after the arrow refers to the target domain. For example, for the transfer fault diagnosis experiment A→B, the CWRU bearing dataset is the source domain, and the IMS bearing dataset is the target domain. We follow the standard evaluating protocol for unsupervised transfer learning tasks. In each experiment, the training dataset includes all the labeled data samples from the source domain data and half of the unlabeled data samples from the target domain. The other half of data samples from the target domain are used for testing.

The detailed parameters for each experiment are set as follows. In the condition recognition module, the sizes of convolution kernel and pooling kernel are set to be 5 and 2, respectively. In the domain adaptation module, the RBF with the bandwidth σ of median pairwise distances on the training data is used to calculate MMD distance between high-level learned features from the source and target domains. As shown in Fig. 4(a), the penalty parameters λ and μ are gradually changed from 0 to 1 using the formula $2/(1+\exp(-10\times p))-1$, and p is the training progress that changes from 0 to 1. The method is trained using the SGD with a learning rate, $0.001/(1+10\times p)^{0.75}$ [28]. The training set and testing set of each experiment are listed in Fig. 3. The batch size is set as 256. Half of each batch is populated by the data samples from the source domain, and the rest is constituted by the ones from the target domain. The training step is set to be 3000. Take transfer fault diagnosis A→B as an example, the loss function of the DCTLN is plotted in Fig. 4(b). As shown in Fig. 4(b), the training loss of the proposed DCTLN converges after about 2000 training epoch.

With those parameters, each transfer fault diagnosis experiment is repeated ten times. The results of six transfer fault diagnosis experiments are shown in Table III. In each

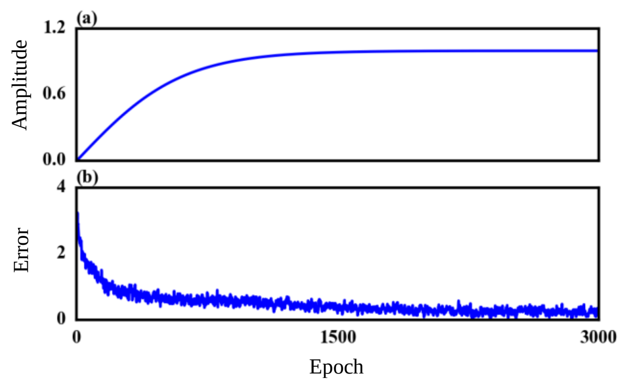


Fig. 4. The penalty parameter and training loss of the proposed DCTLN. (a) Penalty parameter. (b) Training loss.

experiment, all of the transfer fault diagnosis accuracies are over 82% and some are even over 89%. It means that the proposed DCTLN is able to effectively recognize the health conditions of bearings where no labeled data is available.

D. Comparison results

In order to further demonstrate the effectiveness of the proposed DCTLN, five methods are used for comparison on the six transfer fault diagnosis experiments. As shown in Table IV, the five comparison methods are CNN trained only by the source data, transfer component analysis (TCA) [29], deep domain adaptation neural network based fault diagnosis (DAFD) [30], deep domain confusion (DDC) [25], and domain adversarial training of neural networks (DANN) [28]. Based on different comparison purposes, those five methods are classified into three types.

TABLE IV
VARIOUS TRANSFER LEARNING METHODS

Method	Features	Transfer method
CNN	Learned features	No transfer
TCA [29]	Handcrafted features	MMD
DAFD [30]	Frequency spectrum features	MMD
DDC [25]	Learned features	MMD
DANN [28]	Learned features	Adversarial
DCTLN (Ours)	Learned features	Multi adaptation

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

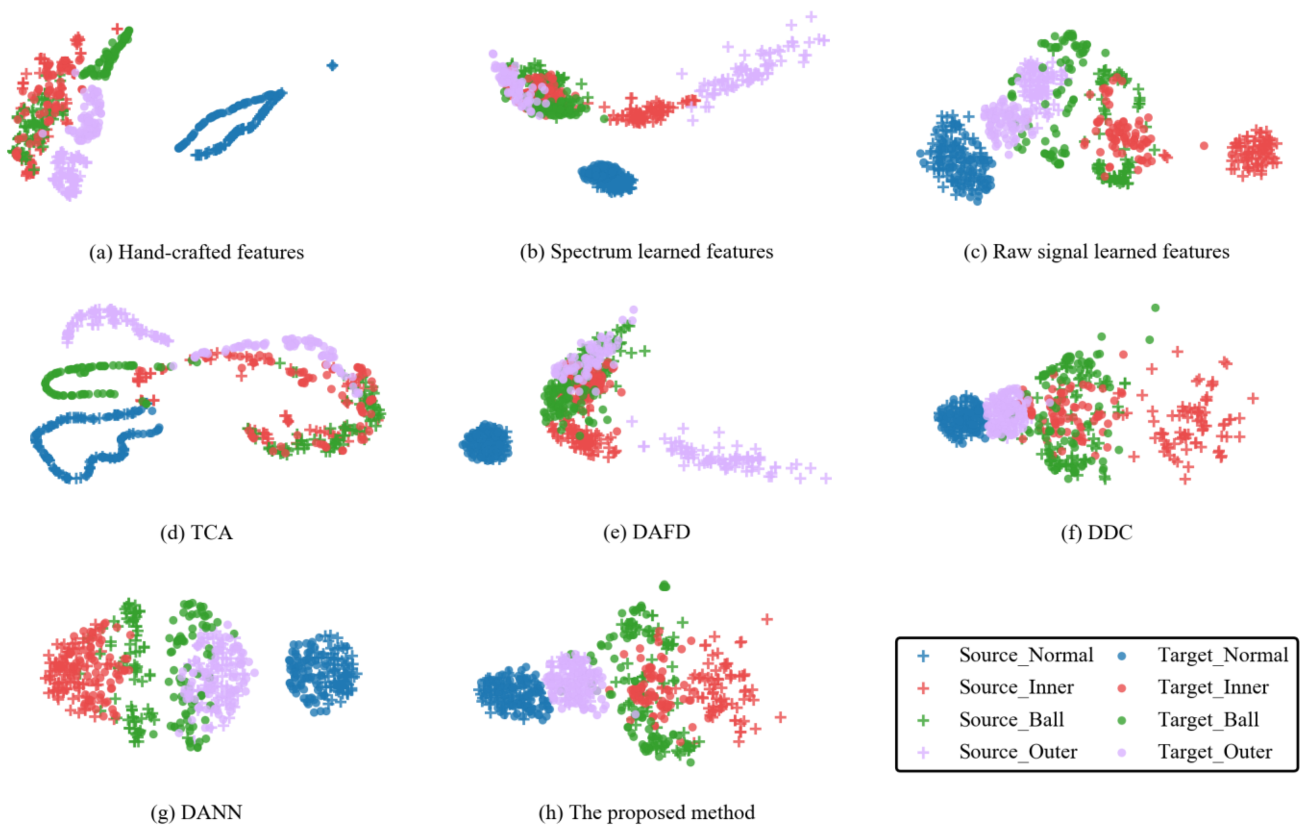


Fig. 5. The t-SNE visualization of features.

1) Comparison to the method with no transfer learning. The first type is designed to illustrate the improvement of transfer learning based methods for transfer fault diagnosis task where data from the target domain are unlabeled. The comparison method is CNN trained only by the source data, in which a CNN is trained just by the data from a source domain.

2) Comparison to handcrafted feature based transfer learning methods. The second type is designed to illustrate effects of learned features for transfer fault diagnosis tasks. Two handcrafted feature based methods, TCA and DAFD, are applied to compare with DCTLN. TCA is a conventional transfer learning method using MMD-regularized subspace learning, which is built with some handcrafted features. In the TCA method, 14 handcrafted features, i.e., mean, RMS, kurtosis, variance, crest factor, wave factor, and eight energy ratios of wavelet package transform, are used. DAFD is a deep domain adaptation fault diagnosis model, in which the input signal of the model is the frequency spectrum of vibration signals. Note that the DAFD is also able to learn features. However, it learns features from the frequency spectrum of vibration signals rather than vibration signals. Therefore, we consider it as a method based on handcrafted features.

3) Comparison to state-of-the-art transfer learning methods. The third type is designed to compare two widely used transfer learning methods, i.e., DDC and DANN, with DCTLN. DDC adds an adaptation layer and a distribution match term MMD in a CNN architecture to ensure that the learned features are domain invariant. DANN adds a domain discriminative component into a deep neural network. The method is expected to learn representation features which are predictive for the source domain data samples, but uninformative about the domain of the input.

The classification accuracies and standard errors on six transfer fault diagnosis experiments are shown in Table III. It can be seen that the proposed DCTLN outperforms all of the methods for comparison on the six transfer fault diagnosis experiments. More specifically, through the comparison results, we obtain the following three observations. 1) Transfer learning based method outperforms the classical method without transfer learning for the transfer fault diagnosis tasks where no labeled data is available in the target domain. The only difference between DCTLN and the CNN trained only by source data is that a domain adaptation module is added into DCTLN, while the results show that DCTLN obtains higher classification accuracies than the CNN trained only by source data. It means that transfer learning may be a promising tool to promote the successful application of intelligent fault diagnosis of machines with unlabeled data. 2) The learned features outperform the handcrafted features for the transfer fault diagnosis tasks. In the comparison methods, TCA uses 14 classical handcrafted features and DAFD extracts features from the frequency spectrum of vibration signals. The other methods, i.e., DDC, DANN and the proposed DCTLN, directly learn features from the vibration signals. From the results, it can be

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

seen that three learned feature based methods outperform two handcrafted feature based methods. The possible reason is that the extracted features may discard some useful condition representation information compared with the features learned from raw signals. Additionally, it also means that deep learning trained with raw signals may be able to reduce the distribution discrepancy of learned features between domains. 3) Compared with the two widely used transfer learning methods, i.e., DDC and DANN, the proposed DCTLN obtains higher recognition accuracies of health conditions in six transfer fault diagnosis experiments. This validates that DCTLN reduces the distribution discrepancy between domain data more effectively than the two widely used transfer learning methods, DDC and DANN. Additionally, the relatively high recognition accuracies confirm the practicability of the proposed DCTLN.

In order to provide visual insights into the effects of transfer learning on the distribution discrepancy of features from the source and target domains, we use the t-distributed stochastic neighbor embedding (t-SNE) [31] technique to map the high-dimensional features into a two-dimensional space. The first transfer fault diagnosis experiment A→B is taken as an example, and the results are shown in Fig. 5. Fig. 5(a) to Fig. 5(c) plot features without transfer learning processing. It can be seen that the distributions of learned features from the source and target domains are closer than the ones of the handcrafted features. This confirms the current practice that deep neural networks learn features that can reduce the domain distribution discrepancy. Fig. 5(d) to Fig. 5(h) plot features with transfer learning processing, namely transferred feature, where Fig. 5(d), Fig. 5(e) and Fig. 5(f, g, h) correspond to Fig. 5(a), Fig. 5(b) and Fig. 5(c). For example, Fig. 5(d) plots the transferred features, where the corresponding features without transfer learning processing are shown in Fig. 5(a). From the visualization of features with and without transfer learning processing, we can find that the distributions of transferred features from the source and target domains are closer than the ones of the features without transfer learning processing. It validates that transfer learning processing is able to reduce the distribution discrepancy of data acquired from different machines. Additionally, we can find that features learned by DCTLN exhibit tighter health condition class clustering while mixing the feature distribution between domains than features shown in Fig. 5(a)-Fig. 5(g).

V. CONCLUSIONS

In this paper, we introduced transfer learning into the field of intelligent fault diagnosis of machines with unlabeled data and proposed a new intelligent method, DCTLN, for transfer fault diagnosis tasks. We demonstrated the ability of our proposed method through six transfer fault diagnosis experiments where no labeled data are available in the target domain. Three conclusions can be drawn from the experiment results. (1) The transfer learning based intelligent fault diagnosis methods obtain higher recognition accuracies of bearing heath conditions compared with the traditional method without transfer learning processing. (2) The features learned from the raw signals may reduce distribution discrepancy between different domain datasets. (3) The proposed DCTLN outperforms the two widely used transfer learning methods, i.e., DDC and DANN, in the six transfer fault diagnosis experiments of bearings. Those conclusions indicate that DCTLN trained with labeled data acquired from one machine is able to effectively classify unlabeled data acquired from the other machines. Therefore, DCTLN is able to promote the successful applications of intelligent fault diagnosis of machines with unlabeled data.

REFERENCES

[1] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Butterworth-Heinemann, 2016.

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

[3] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, " A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 107, pp. 241-265, 2018.

[4] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mechanical Systems and Signal Processing*, vol. 70, pp. 87-103, 2016.

[5] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637-646, 2015.

[6] K. B. Lee, S. Cheon, and C. O. Kim, "A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135-142, 2017.

[7] Z. Su, B. Tang, Z. Liu, and Y. Qin, "Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine," *Neurocomputing*, vol. 157, pp. 208-222, 2015.

[8] R. Zhang, H. Tao, L. Wu, and Y. Guan, "Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions," *IEEE Access*, vol. 5, pp. 14347-14357, 2017.

[9] W. Qiao and D. Lu, "A survey on wind turbine condition monitoring and fault diagnosis—Part II: Signals and signal processing methods," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6546-6557, 2015.

[10] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137-3147, 2016.

[11] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[12] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799-834, 2018.

[13] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296-2305, 2017.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[15] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, 79(1-2), 151-175, 2010.

[16] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2615-2626, 2016.

[17] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International Conference on Machine Learning*, 2015, pp. 97-105.

[18] K. D. Feuz and D. J. Cook, "Transfer learning across feature-rich heterogeneous feature spaces via feature-space remapping (FSR)," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 1, p. 3, 2015.

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2018.2877090, IEEE Transactions on Industrial Electronics

[19] K. Yan and D. Zhang, "Correcting Instrumental Variation and Time-Varying Drift: A Transfer Learning Approach with Autoencoders," *IEEE Transactions on Instrumentation and Measurement,* vol. 65, no. 9, pp. 2012-2022, 2016.

[20] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717-1724.

[21] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine,* vol. 11, no. 1, pp. 20-31, 2016.

[22] C. Devin, A. Gupta, T. Darrell, P. Abbeel, and S. Levine, "Learning modular neural network policies for multi-task and multi-robot transfer," *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 2017, pp. 2169-2176: IEEE.

[23] L. Wen, L. Gao, and X. Li, "A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* 2017.

[24] J. Xie, L. Zhang, L. Duan, and J. Wang, "On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis," *Prognostics and Health Management (ICPHM), 2016 IEEE International Conference on*, 2016, pp. 1-6: IEEE.

[25] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474,* 2014.

[26] "Case Western Reserve University Bearing Data Center Website " *http://csegroups.case.edu/bearingdatacenter/home*.

[27] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of sound and vibration,* vol. 289, no. 4, pp. 1066-1090, 2006.

[28] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," Journal of Machine Learning Research, vol. 17, no. 59, pp. 1-35, 2016.

[29] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks,* vol. 22, no. 2, pp. 199-210, 2011.

[30] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model based domain adaptation for fault diagnosis," *IEEE Transactions on Industrial Electronics,* vol. 64, no. 3, pp. 2296-2305, 2017.

[31] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of machine learning research,* vol. 15, no. 1, pp. 3221-3245, 2014.

Liang Guo was born in ChangDe, Hunan Province, P. R. China. He received the B.S. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, P. R. China, in 2011 and 2016, respectively. His research interests focus on machinery intelligent fault diagnostics, and remaining useful life prediction.

He is currently an Assistant Professor at Southwest Jiaotong University. Prior to joining Southwest Jiaotong University in 2018, he was a Postdoctoral Research Fellow with the Xi'an Jiaotong University, Xi'an, P. R. China.

Yaguo Lei (M'15) received the B.S. and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, P. R. China, in 2002 and 2007, respectively. He is currently a Full Professor of mechanical engineering at Xi'an Jiaotong University. Prior to joining Xi'an Jiaotong University in 2010, he was a Postdoctoral Research Fellow with the University of Alberta, Edmonton, AB, Canada. He was also an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Duisburg, Germany. His research interests focus on machinery condition monitoring and fault diagnosis, mechanical signal processing, intelligent fault diagnostics, and remaining useful life prediction.

Dr. Lei is a member of the editorial boards of more than ten journals, including Mechanical System and Signal Processing and Neural Computing & Applications. He is also a member of ASME and a member of IEEE. He has pioneered many signal processing techniques, intelligent diagnosis methods, and remaining useful life prediction models for machinery.

Saibo Xing was born in Nantong, Jiangsu Province, P. R. China. He received the B.S. degree in material science and engineering from Xi'an Jiaotong University, in 2015. He is devoted to the research of intelligent fault diagnostics of rotating machinery.

He is currently working for the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, P. R. China.

Tao Yan is currently working toward the Ph.D. degree in mechanical engineering at the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, P. R. China. He received the B.S. degree in mechanical engineering from Central South University, P. R. China, in 2016.

His research interests are machinery condition monitoring, intelligent fault diagnostics and remaining useful life prediction of rotating machinery.

Naipeng Li is currently working toward the Ph.D. degree in mechanical engineering at the Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, P. R. China. He received the B.S. degree in mechanical engineering from Shandong Agricultural University, P. R. China, in 2012.

His research interests are machinery condition monitoring, intelligent fault diagnostics and remaining useful life prediction of rotating machinery.

0278-0046 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

View publication stats