IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 7, 2017, accepted June 26, 2017, date of publication June 28, 2017, date of current version August 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2720965

Transfer Learning With Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions

RAN ZHANG[1], HONGYANG TAO[2], LIFENG WU[1], (Member, IEEE), AND YONG GUAN[1], (Member, IEEE)
[1]Information Engineering, Capital Normal University, Beijing 100048, China
[2]School of Literature, Capital Normal University, Beijing 100089, China

Corresponding author: Lifeng Wu (wulifeng@cnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61202027, in part by the Beijing Natural Science Foundation of China under Grant 4122015, and in part by the Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges under Beijing Municipality under Grant IDHT20150507.

ABSTRACT Traditional machine learning algorithms have made great achievements in data-driven fault diagnosis. However, they assume that all the data must be in the same working condition and have the same distribution and feature space. They are not applicable for real-world working conditions, which often change with time, so the data are hard to obtain. In order to utilize data in different working conditions to improve the performance, this paper presents a transfer learning approach for fault diagnosis with neural networks. First, it learns characteristics from massive source data and adjusts the parameters of neural networks accordingly. Second, the structure of neural networks alters for the change of data distribution. In the same time, some parameters are transferred from source task to target task. Finally, the new model is trained by a small amount of target data in another working condition. The Case Western Reserve University bearing data set is used to validate the performance of the proposed transfer learning approach. Experimental results show that the proposed transfer learning approach can improve the classification accuracy and reduce the training time comparing with the conventional neural network method when there are only a small amount of target data.

INDEX TERMS Fault diagnosis, transfer learning, neural networks, machine learning.

I. INTRODUCTION

Prognostics and system health management is crucial to machinery of which bearings are the key components. Therefore, many bearing fault diagnosis methods are proposed. These fault diagnosis methods can be divided into model driven methods [1], [2] and data driven methods. The former methods have to know the physical structure and it is complex to set up the physical model. But the data driven methods can directly recognize faults from the bearing vibration signals which are collected by sensors and there is no need to know the inner structure of bearings. The feature selection of vibration signals is of great importance to the fault diagnosis [3]. Machine learning algorithms and classifiers with signal processing methods are commonly used in traditional data driven fault diagnosis models [4]–[6]. Wu *et al.* [7] adopted manifold learning and wavelet neural networks to make diagnosis. These methods can classify faults successfully. Cerrada *et al.* [8] proposed a fault diagnosis model based on genetic algorithm and random forest. Deep neural network based fault diagnosis methods are also implemented. Jia *et al.* [9] used deep neural networks to recognize faults by training the model with frequency spectra extracted from vibration signals. Lu *et al.* [10] investigated the application of stacked denoising autoencoder in health state indentification. Guo *et al.* [11] proposed a model based on adaptive convolutional neural network on the bearing diagnosis. Ren *et al.* [12] used a deep neural network based model to predict remaining useful life of rolling bearings. Chen *et al.* [13] employed three different deep neural network, including Deep Boltzmann Machines, Deep Belief Networks and Stacked Auto-Encoders to identify faults of rolling bearings. Li *et al.* [14] set up a Gaussian-Bernoulli deep Boltzmann machine to learn statistical features from signals and make diagnosis. Zhang *et al.* [15] proposed a fault diagnosis model based on deep neural networks to learn directly from time series data without signal preprocessing. However, these traditional data driven methods with machine learning algorithms strictly require that training and testing data must be in the same working condition and have the same distribution and feature space. They cannot utilize all kind of data in the real world changing working conditions. For these methods, enough training data are firstly used to train the fault diagnosis model. Then, testing data in the same working condition are used to test the performance of model. But the working conditions of machinery, especially bearings, cannot be unchanged in the real world. The fault diameter gets larger and larger with the working of bearings and the radical load cannot be same all the time. Traditional data driven methods with machine learning algorithms do not apply to the working condition which is changing with time. In recent years, transfer learning, a new machine learning method, is adopted in situation where source data and target data are in different feature spaces or have different distributions. It has made great progresses in image, audio and text recognition [16]–[18]. Do and Ng [19] applied transfer learning in text classification and achieved better performance than the previous methods without transfer. Oquab *et al.* [20] combined convolutional neural networks with transfer learning to learn and transfer mid-level image representations. The classification accuracy improved 8% comparing with traditional neural networks. Raina *et al.* [21] designed self-taught learning to learn from unlabeled data and it improves classification accuracies of given image, audio and text classification tasks, respectively. Dai *et al.* [22] proposed a transfer learning framework called TrAdaBoost to select most useful and different samples as additional training data to improve the performance. Yao and Doretto [23] extend the boosting based transfer learning framework to transfer knowledge from multiple sources. Above transfer learning methods learn properties from source task and transfer parameters to target task. The source task and the target task can be different but they are more or less related. For example, it is much easier to transfer from one image recognition task to another image recognition task than to a speech recognition task.

Based on transfer learning with neural networks, this paper presents a fault diagnosis model to make full use of data in different working conditions. Massive source data are used to train the original neural network model until it get the expected performance. Then the original neural network model is transferred to a new target task which is in another working condition and the structure of the network is altered according to the data distribution. These task are different but their data are similar. In real world working conditions, target data are usually much harder to get than the source data. Therefore, only a small amount of target data are available to train the transferred new model. Conventional machine learning approaches cannot take advantage of massive source data and it learns characteristic only from a small quantity of target data. Transfer learning can make use of both source data and target data, which results in better performance. The contribution of this proposed transfer learning model is to reduce the training time and improve the classification accuracy by utilizing data in different working conditions when there is only a small amount of target data.

The remainder of this paper is organized as follows. Section II describes the proposed transfer learning approach in fault diagnosis. Section III shows the experiments and results. Section IV makes discussions and section V draws the conclusion.

II. METHODS

A. TRANSFER LEARNING

Transfer learning is to improve the performance of current task by exploiting source data which are different from but similar to target data. Target task is the current task which is to be solved. Source task is a task which is similar to target task. For instance, transfer learning is a process that a man learns French much faster in condition that he has learnt well in English; or to learn surfing much easier after he has learnt skating; or to learn physics much better when he has learnt well in mathematics. Achieving remarkable improvement of performance by learning from similar tasks is the key of transfer learning. The aim of transfer is to take advantage of experience learnt in one task and improve the performance in a similar but different task [24]. In this research, transfer learning is utilized to learn some useful characteristics from a great amount of source data which are already collected in one working condition and transfer parameters to target task in another working condition where there is only a small amount of target data.

$$D_s = \{X_s, T_s\} \tag{1}$$

$$D_t = \{X_t, T_t\} \tag{2}$$

Where D_s represents the source data and D_t represents the target data. X_s, X_t are the source and target samples, respectively and T_s, T_t are the corresponding labels. The source task and target task can be denoted as follows.

$$Y_s = f_s\,(X_s, \theta_s) \tag{3}$$

$$Y_t = f_t\,(X_t, \theta_t) \tag{4}$$

Here, f_s means mapping from X_s to Y_s and f_t means mapping from X_t to Y_t. Y_s, Y_t are the real outputs, while T_s and T_t are the expected outputs of source and target task, respectively. θ_s, θ_t are parameters in source and target task, respectively. $\{X_s, T_s\}$, $\{X_t, T_t\}$ are data in different distributions or have different feature spaces but they are in related or similar domains. In this research, source data and target data are normal and faults vibration signals in different working conditions such as different working loads and different fault diameters. Transfer learning is to find some related properties in source data $\{X_s, T_s\}$ and get the mapping f_s in source task. Then the mapping in target task f_t is learnt from target data $\{X_t, T_t\}$ after transferring mapping f_s to target task. Details can be denoted as follows.

$$\theta_s = \theta_0 + \theta_1 \tag{5}$$

$$\theta_t = \theta_0 + \theta_2 \tag{6}$$

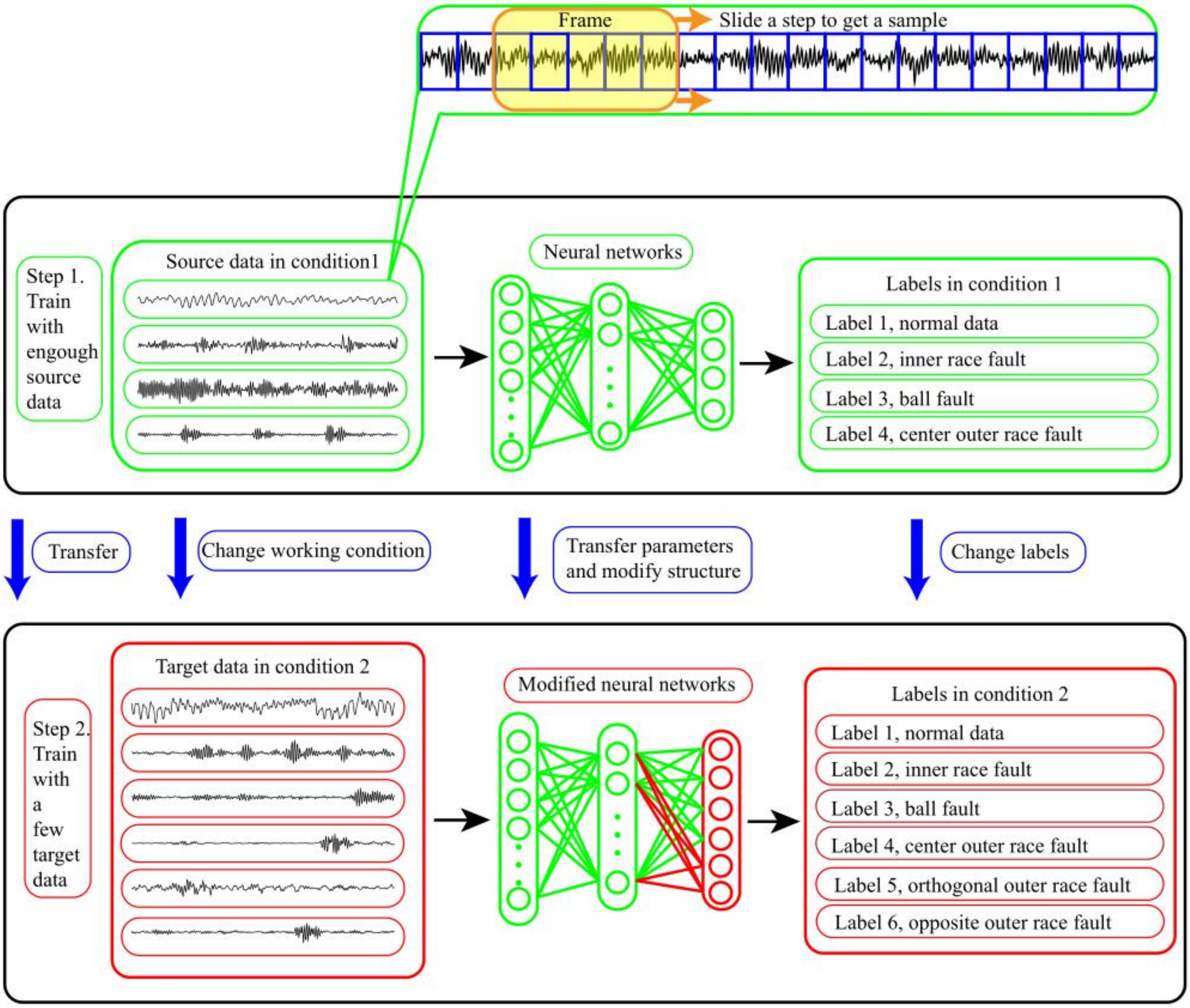

FIGURE 1. **Transfer learning based fault diagnosis model.**

$$\theta_s' = \arg min \sum\nolimits_{i=1}^{n_s} C(x_{s_i}, \theta_s, t_s) \tag{7}$$

$$\theta_t' = \arg min \sum\nolimits_{i=1}^{n_t} C(x_{t_i}, \theta_t, t_t) \tag{8}$$

In (7) and (8), $C(.)$ is cost function and the training process is to minimize the cost function. n_s, n_t are the number of samples in source and target task. θ_s, θ_t are parameters in source and target task, respectively. They have some common parts θ_0 and some different parts. In (7) and (8), θ_1 represent parameters in source task which cannot be transferred to the target task while θ_2 represent the unit parts of parameters in target. Transfer learning is try to utilize the common parts of these parameters when training models in target task. The aim of training in target task and source task are to get the optical parameters θ_s' and θ_t', respectively, as shown in (7) and (8).

B. SLIDING FRAME

The aim of sliding frame is to make useful training samples as many as possible after the normal and fault vibration data are obtained. Massive data are necessary for the training of the model. The vibration data which are collected by sensors are one dimensional and they can be very long. Therefore, a frame is adopted to take samples by sliding small steps as shown in Fig. 1. For every little step, there is a sample. The length of frame is τ, which can be set according to the experimental requirements. It means that the number of data points in every sample is τ. The step size is denoted as ε and it is set to be smaller than τ so that more training and testing samples can be taken. If the total length of the vibration signal is ζ, the number of samples n can be figured out as follow.

$$n = \left\lfloor \frac{\zeta - \tau}{\varepsilon} \right\rfloor + 1 \tag{9}$$

The length of vibration signal ζ is a fixed value. The step size ε and frame size τ should be set appropriately so that as many as effective samples can be made. If the frame size τ is set to be too large, the size of input layer will be enlarged accordingly, which will lead to dramatically increase of computing time of neural networks. However, too small value of frame size will not cover the length of enough characteristics of vibration signals, resulting that samples are confusing and the classification accuracies decrease. In the same way, the selection of step size ε is also of great importance. Large steps will get less samples, which is an adverse effect in training process. Small steps will get too many similar samples. Therefore, these sizes should be set according to the experiment and the collected data.

C. NEURAL NETWORKS

The details of the structure of neural networks in transfer learning model can be seen in Fig. 1. The neural networks are directly used for the automatic feature selection and fault classification without any other signal processing methods. It has been proved that this approach can work effectively [11].

The structure of neural networks is set according to the dimensionalities of samples and labels. The number of neurons in inputs layer of neural networks is equal to the dimensionality of samples which are made by the sliding window and the number of neurons in outputs layer is the types of labels. However, there is no strict rule for the selection of hidden layers. In this research, the structure of the hidden layers are set according to the experimental requirements. The number of neurons in next layer is simply set to be larger than that in the former layer.

In the training process of source task, the original neural networks are initialized randomly. And in the training process of target task, the neural networks are modified so the initial parameters are kept same as them in the neural networks which are trained by enough source data. However, there are some extra structure in the modified neural networks comparing with the original structure because target data and source data may have different distributions or be in different feature spaces. And these additional parameters are randomly initialized. Then the whole modified neural networks are trained with a small amount of target data.

More formally, let $\mathbf{x}$, $\mathbf{y}$ be the input vector and output vector of neural networks, respectively. The proposed neural network based fault diagnosis model has three layers, an input layer, an output layer and a hidden layer. The hidden vector is denoted as $\mathbf{h}$. The feed forward process is as follows.

$$\mathbf{h} = \sigma(\mathbf{w_1 x}+\mathbf{b_1}) \tag{10}$$

$$\mathbf{y} = \sigma(\mathbf{w_2 h}+\mathbf{b_2}) \tag{11}$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{12}$$

Where $\sigma(\cdot)$ is the sigmoid activation function. $\mathbf{w_1}$ is the weight matrix between input layer and hidden layer and $\mathbf{w_2}$ is the weight matrix between hidden layer and output layer. $\mathbf{b_1}$ and $\mathbf{b_2}$ are bias vectors of hidden layer and output layer, respectively. The loss function L is as follow.

$$\mathrm{L} = -\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \|\mathbf{t_i}\cdot\ln\mathbf{y_i} + (1+\mathbf{t_i})\cdot\ln(1-\mathbf{y_i}))\|_1 \tag{13}$$

Where $\mathbf{t_i}$ is one of the target vector and n is the number of training samples. The aim of training neural network is to minimize the loss function L by back propagation and gradient descent. The updating rule of parameters are as follows.

$$w = \mathrm{w} + \eta\frac{\partial \mathrm{L}}{\partial \mathrm{w}} \tag{14}$$

$$\mathrm{b} = \mathrm{b} + \eta\frac{\partial \mathrm{L}}{\partial \mathrm{b}} \tag{15}$$

In equation (14) and (15), η is the learning rate which can be set appropriately. W and b are parameters of the neural network. By fine-tuning these parameters for hundreds or thousands of times, L gets closer and closer to the minimal value.

D. DETAILED PROCEDURE OF FAULT DIAGNOSIS

The proposed transfer learning based fault diagnosis model is shown in Figure 1. The model can be divided into two main training blocks. The first one is the training of neural network with data in working condition 1 and second one is the training of transferred model with data in working condition 2.

The detailed steps of first training process are as follows:

a) Enough source data in working condition 1 are collected.
b) Samples are made by sliding the frame.
c) For every kind of sample, corresponding label is added.
d) The original neural networks are constructed according to the dimensionalities of samples and labels. The parameters of neural networks are randomly initialized.
e) The neural networks are trained with massive source data.

The details of transferring are as follows:

a) The working condition is changed. So the data may have different features and labels from original data.
b) The structure of neural networks are modified according to dimensionalities of new data and labels. The input size should be same as the size of samples and output size should be changed to the kind of new labels.
c) The parameters of original neural networks are kept unchanged. New parameters including weights and parameters which are added to the structure are randomly initialized.
d) The classification results of neural networks are different from original because new types of labels are added.

The detailed steps of second training process are as follows:

a) A small amount of data in working condition 2 are obtained for the training.
b) New samples are made by sliding the frame on new data.
c) For every kind of sample, corresponding label is added.
d) The modified neural networks are trained with new data.

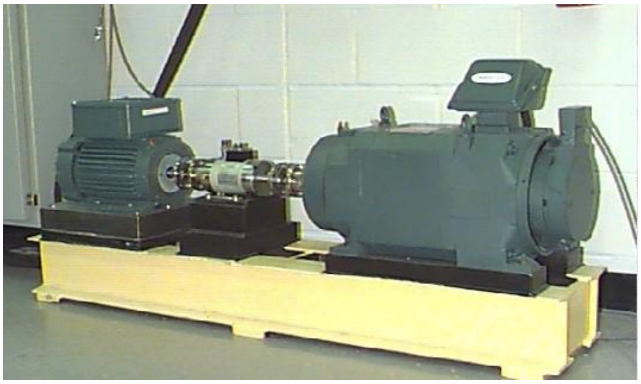

FIGURE 2. Experiment equipment.

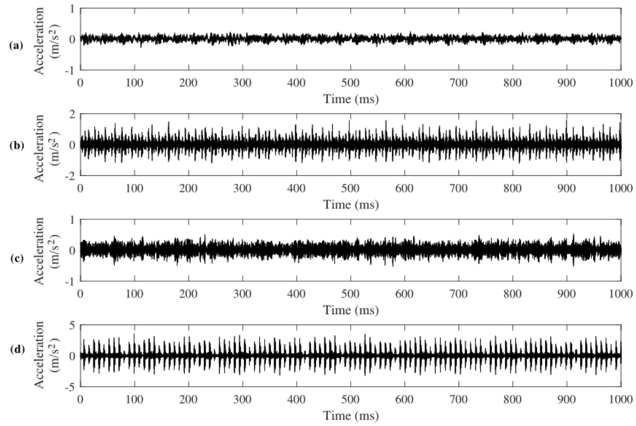

FIGURE 3. Collected signals in working condition 1. The horizontal axis represents the time and the horizontal axis are the acceleration data which were collected by the accelerometers. The normal vibration signals are shown in (a). Three kinds of fault vibration signals including inner race fault, ball fault and center outer race fault vibrations signals are shown in (b), (c) and (d), respectively. The rotation speed was kept at 1797 round per minute and there is no load added to the motor. The diameters of these fault are 0.007 of an inch.

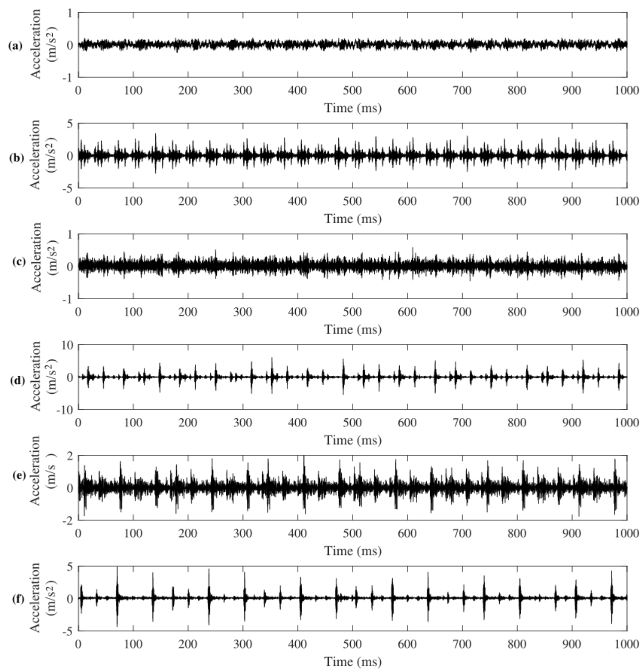

FIGURE 4. Collected signals in working condition 2. The x axis is the time and y axis are the acceleration data. Normal vibration data are shown in (a). Five kinds of fault vibration signals including inner race fault, ball fault, center outer race fault, orthogonal outer race fault and oppositely outer race fault vibrations signals are shown in (b), (c), (d), (e) and (f), respectively. The diameter of these faults are 0.021 of an inch.

III. EXPERIMENTS AND RESULTS

A. EXPERIMENT EQUIPMENT AND PROCEDURE

Experiment equipments are shown in Fig. 2. This experiment was conducted by Bearing Data Center of Case Western Reserve University [25]. The motor was added on the shaft which was supported by some bearings in this experiment. And the SKF (Sevenska Kullagerfabriken AB, Gothenburgh, Sweden) bearings were used in the experiment. Single point fault were introduced to the test bearings using electro-discharge machining with different fault diameters of 7 mils, 14 mils and 21 mils (1 mil equals to 0,001 inches). The motor rotation speed was kept instant when the motor loads are the same. Accelerometers, which can be seen as sensors, were placed to the bearing housing with magnetic bases. Some accelerometers sensors were also placed to the motor housing. Digital data was collected by sampling with the frequency of 12 kHz in the normal vibration and fault vibration conditions, that is to say, 12,000 data points were collected in one second. In this way, vibration signals were collected by these accelerometers when theses bearings were working. On the left of the test stand, there was a motor of which the maximum load was 3 hp. A torque transducer was placed in the center of the stand and a dynamometer was on the right. The experiment had a set of control electronics but it is not shown in the photo. A few kinds of fault vibration signal, including inner race fault, ball fault and outer race fault vibration signals, were recorded in different working conditions in the experiment. Outer race faults were stationary faults therefore the placement and the load zone on bearings of this kind of fault had a direct influence on system's vibration response. In some working conditions, experiments of out race fault in different places of bearings were conducted. Vibration signals in different working conditions when the motor load was different or the bearings had different faults were also recorded.

B. TRANSFERRING TO WORKING CONDITION WITH DIFFERENT FAULT DIAMETER

In order to improve the performance of model in working condition 2 when there is less training samples, experiment which transfers from working condition 1 to working condition 2 with different fault diameter and different fault types is conducted. The vibration signals in different working conditions are shown in Fig.3 and Fig.4.

Some differences of these two working conditions are shown in Tab. 1. There are enough training samples in source data comparing with target data. Neural network is transferred from working condition 1 to working condition 2 with larger fault diameter. What's more, there are more fault types

TABLE 1. Details of working conditions in transfer learning.

	Working condition 1	Working condition 2
Fault diameter (inch)	0.007	0.021
Motor load (hp)	0	0
Motor speed (rpm)	1797	1797
Number of fault types	3	5
Samples of every type	4832	1208
Frame size	400	400
Step size of frame	20	20

TABLE 2. Labels and samples in working condition 1. These data are source data of transfer learning method.

Label	Data type	Number of training samples
1	Normal	4832
2	Inner race fault	4832
3	Ball fault	4832
4	Outer race fault at center@ 6:00	4832

TABLE 3. Labels and samples in working condition 2. These data are target data of transfer learning methods. Also, they are the training data of traditional neural network without transfer in this experiment.

Label	Data type	Number of training samples
1	Normal	1208
2	Inner race fault	1208
3	Ball fault	1208
4	Outer race fault at center@ 6:00	1208
5	Outer race fault at orthogonal @ 3:00	1208
6	Outer race fault at oppositely @ 12:00	1208

in target task so the number of fault types changed from 3 to 5. And the data in both working conditions contain the normal data. Therefore, the number of neurons in last layer of neural network which is equal to the number of data types is increased from 4 to 6 after transferring as shown in Figure 1. Their motor speed is kept as same at 1797 rpm (round per minute) and there is no motor load in both working conditions. The size of sliding frame is 400, which means that every sample contains 400 collected data points. The step size of sliding frame is set to be 20 to make samples from vibration signals. According to equation (9), the number of samples can be calculated. 4832 samples of every kind of data in source task are used to train the original neural network while only 1208 samples of every kind data in target task are obtained to train the transferred neural networks.

To make a comparison, traditional machine learning method, a neural network without transfer is implemented. This neural network have the same structure as the modified neural networks in transfer learning model. However, its parameters including weights and biases are randomly initialized. 1208 samples of every kind of data in working condition 2 are used to train the neural network. The traditional neural network method has the same training samples and labels as the transferred model in target task. The details of their labels and samples are shown in Tab. 2 and Tab. 3.

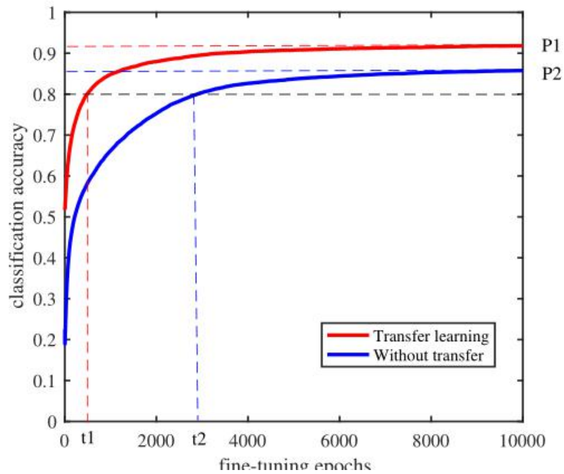

FIGURE 5. Total testing data classification accuracies of different methods. The horizontal axis represents the fine-tuning time in training process and the vertical axis is the classification accuracy.

TABLE 4. Classification accuracies of different methods.

Label	From condition 1 to condition 2		
	Without transfer	Transfer learning	Improvements
Label 1	100%	100%	0
Label 2	75.9%	86.6%	10.7%
Label 3	96.0%	97.8%	1.8%
Label 4	66.4%	85.4%	19%
Label 5	88.5%	93.5%	5%
Label 6	87.7%	87.5%	-0.2%
Total	85.8%	91.8%	6%

The testing data performance of these two methods are shown in Fig. 5 and Fig. 6. As can be seen in Fig. 5, transfer learning method can improve the total classification accuracy and reduce the training epoch comparing with traditional neural network without transfer. The performance of these two methods in the training process are of different characteristics. Firstly, transfer learning based model gets better parameters and performance in the beginning. Secondly, it learns much faster than traditional neural network. Finally, it gets better performance after training for 10000 fine-tuning epochs.

On the one hand, transfer learning method learns much faster. These two methods spend far more different time getting the same level performance. For example, transfer learning method spends t1 time achieving 80% classification accuracy but traditional method spends t2 time, which is much longer than t1. On the other hand, transfer learning improve the classification accuracy when spending same time. As can be seen in Figure 5, the total classification accuracy of transfer learning approach is 91.8% after training for 10000 epochs while traditional approach only gets 85.8% accuracy of the same training and testing dataset.

As shown in Figure 6 and Tab. 4, classification accuracies of different labels using transfer learning method are higher than or almost same as that using traditional method.

Label 1 represents the normal data and they are easy to recognize. Although the classification accuracies of label 1

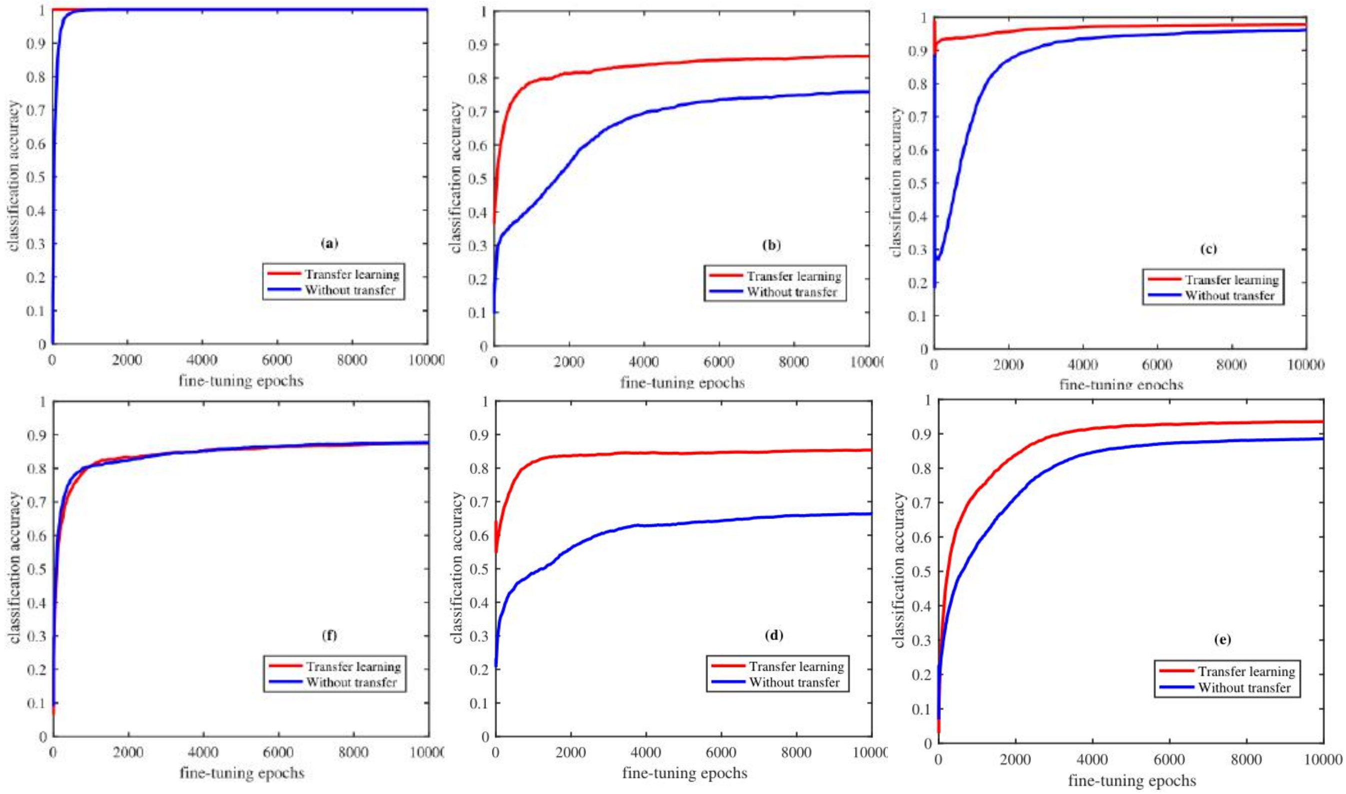

FIGURE 6. Classification accuracies of different labels. (a) - (f) represent classification accuracies of label 1 - label 6, i.e., normal data and five kinds of faults data, respectively.

using these two methods are both 100% after training for 10000 epochs, it is obviously that transfer learning method learns much faster. Label 2 and label 4 are data of inner race fault and outer race fault at center. They are not easy to recognize using traditional method without transfer. However, transfer learning method improve the performance significantly as shown in Figure 6 (b) and 6 (d). Label 3 is the ball defect data. The performances of both methods start with high level values but drop sharply in the beginning. The reason is that models are trained to get better performance for all kinds of data. Therefore, the performance of some labels may drop in some periods but the total performance would increase. Label 1 $\sim$ label 4 are labels in source task of transfer learning method and they are transferred to the target task. Label 5 and label 6 are new labels in target task. As the performance of label 1 $\sim$ label 4 made great progresses, label 5, the new label, also achieved significant improvement using transfer learning method and performance of label 6 are almost the same using both methods.

C. TRANSFERRING TO WORKING CONDITION WITH MORE DIFFERENCES

To test the performance of the model which transfers to working condition with more differences, an experiment transferring from working condition 1 to working condition 3 with different fault diameter, different motor load and different motor speed is conducted. Except for the different source data, this experiment has the same procedure as the former experiment which transfers from working condition 1 to working condition 2. The collected vibration signals in working condition 3 are shown in Fig. 7.

There are much more source data in working condition 1 than target data in working condition 3. The frame size and step size are set as 400 and 20, respectively. The fault diameter in working condition 1 is 0.07 while it is 0.021 in working condition 3. There is no load on the motor in working condition 1 but 1 hp load is added in working condition 3. What's more, the motor speed in working condition 3 is slower than the speed in working condition 1.

The transfer learning method and traditional neural network method which is used to make comparison in this experiment are similar to those in former experiment which transferred from working condition 1 to working condition 2. In the same way, a traditional neural network method is implemented. The training samples and labels of source data of transfer learning methods are same with the former experiment as shown in Table 2. The target data of transfer learning method, as well as the training and testing data of traditional neural network are changed to data in working condition 3. The number of training samples and labels are same with that in working condition 2 as shown in Table 2 and Tab. 3. The detailed comparison between working condition 1 and working condition 3 are shown in Tab. 5.

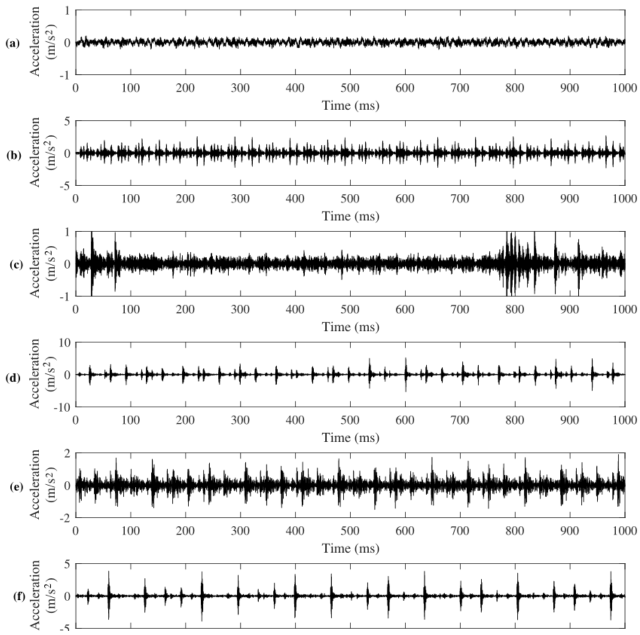

FIGURE 7. Collected vibration signals in working condition 3. The x axis represents the time and y axis are the vibration values. Normal vibration data are shown in (a). Five kinds of fault vibration signals including inner race fault, ball fault, center outer race fault, orthogonal outer race fault and oppositely outer race fault vibrations signals are shown in (b), (c), (d), (e) and (f), respectively. The diameters of these faults are 0.021 of an inch. The rotation speed is kept at 1772 rpm and there is 1 hp load added to the motor.

TABLE 5. Details of working conditions in transfer learning.

	Working condition 1	Working condition 3
Fault diameter (inch)	0.007	0.021
Motor load (hp)	0	1
Motor speed (rpm)	1797	1772
Number of fault types	3	5
Samples of every type	4832	1208
Frame size	400	400
Step size of frame	20	20

The performances of transfer learning method and traditional neural network without transfer are shown in Figure 8 and Tab. 6. The total classification accuracy are obviously improved from 79.4% to 84.2% when using transfer learning method after training for 10000 fine-tuning epochs. What's more, the transfer learning method spends less time getting to a specific classification accuracy than the traditional method. For instance, transfer learning method takes t1 time to achieve 70% classification accuracy while the traditional method takes t2 time as shown in Figure 8. Therefore, transfer learning method can improve the classification accuracy and reduce the training time when there is only a small amount of target data.

IV. DISCUSSION

A. SIMILARITY OF WORKING CONDITIONS AFFECTS THE PERFORMANCE

As can be seen in Tab. 4 and 6, transferring to different target task results in different performance. That is to say, the similarities of source task and target task affect the performance of transfer learning. Fault diameters and fault types are different in working condition 1 and working condition 2. However, there are more differences between working condition 1 and working condition 3. Besides different fault diameters and fault types, they have different motor loads and motor speeds. Great similarities of source task and target task contribute to the improvements of performance using transfer learning.

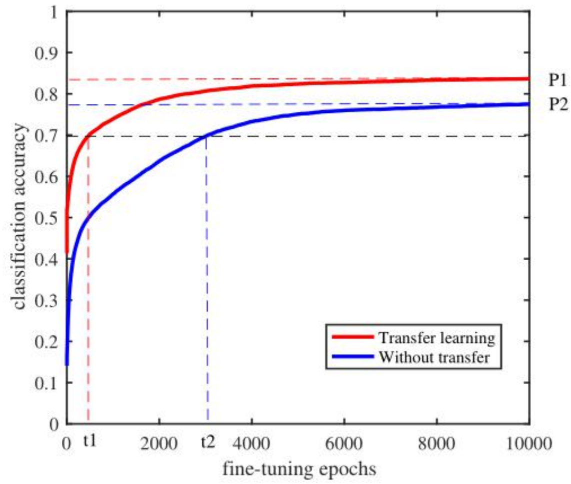

FIGURE 8. Total testing data classification accuracies of different methods. The red line represents the performance of transfer learning method which transfers from working condition 1 to working condition 3. The blue line is the performance of neural networks method without transfer learning. The horizontal axis represents the fine-tuning epochs in training process and the vertical axis is the classification accuracy.

TABLE 6. Classification accuracies of different methods.

Label	From condition 1 to condition 3		
	Without transfer	Transfer learning	Improvements
Label 1	100%	100%	0
Label 2	80.7%	91.0%	10.3%
Label 3	94.9%	95.4%	0.5%
Label 4	56.1%	70.1%	14%
Label 5	82.0%	86.7%	4.7%
Label 6	62.9%	62.2%	-0.7%
Total	79.4%	84.2%	4.8%

Data of label 1 are easy to recognize and their classification accuracies can achieve 100% no matter in which working condition and which method is used. Classification accuracy of data of label 2 can get around 95% even by traditional methods. Therefore, there is only a little improvement using transfer learning. And great improvements can be seen from performances of data of label 2, label 4 and label 5. Classification accuracies of label 4 are improve 19% and 14% in these two experiments, respectively. Also, the improvements of classification accuracy of label 2 are 10.7% and 10.3% respectively. Experimental results show that transfer learning is effective in these kinds of data.

Transfer learning between more similar tasks achieves greater improvements. The total classification accuracy is 91.8% using transfer learning from working condition 1 to working condition2, while the traditional neural network without transfer which only utilizes data in working condition 2 achieves 85.8% classification accuracy.

The improvements is 6%. However, the improvements is only 4.8% by transfer learning from condition 1 to condition 3 comparing with traditional neural network which is only trained by the data in condition 3. And how to measure the similarities of domains will be solved in the future.

B. NEGATIVE TRANSFER

Negative transfer is that the training of source task leads to worse performance of target task. How to prevent negative transfer is still an open problem. If the source domain is more different with target domain, there will be more negative transfer. Therefore, source data which are more similar with target data should be chosen in order to reduce negative transfer. As can be seen in Tab. 4 and Tab. 6, although negative transfer happens in some specific classes with a little influence, the overall performance improves significantly. There are more differences between working condition 1 and working condition 3, so the performances of transfer learning are less improved and the negative transfer is more obvious. The motor loads and rotation speeds are changed when transferring from working condition 1 to working condition 3. The sampling frame size and step size of the sliding frame are fixed values. So the change of rotation speed will result in the change of feature space of training samples. For instance, if the rotation speed in source domain is twice times as it in target domain, a single training sample in target domain would cover only half of rotation period while a single training sample in source domain covers a whole rotation period. The performances of transfer learning are badly affected by the difference of rotation speed.

Therefore, the limitation of this proposed method is that the rotation speed of source domain and target domain cannot be greatly different. A slight difference of rotation speed is acceptable. And the difference of fault diameter is also reasonable even the fault diameter in source domain is several times larger than source domain.

C. TRANSFER LEARNING SETTINGS

According to the availability of labels of source data and target data, transfer learning approaches are divided into three categories: inductive transfer learning, transductive transfer learning and unsupervised transfer learning. Target data labels are necessary in inductive transfer learning whether source data labels are available or not. In transductive learning, target data labels are available and source data labels are unavailable. In unsupervised transfer learning, both target data labels and source data labels are unavailable. There are usually enough source data and relatively small amount of target data in transfer learning process. Comparing to labeled data, unlabeled data are easy to obtain. Supervised training is usually much easier to implement and gets better performance. In the experiments of collecting vibration signals of rolling bearings, labels of data are easy to get. Therefore, it's most convenient to utilize the labeled data both in source domain and in target domain. That is to say, inductive transfer learning is the most suitable approach in this application.

There are usually four different settings in inductive transfer learning: Instance-transfer, feature-representation transfer, parameter-transfer and rational-knowledge-transfer. Instance-transfer approaches try to find and re-weight some useful data from source domain then apply them to the target domain. Feature-representation-transfer approaches aim to get the common feature representations between source domain and target domain. Parameter-transfer approaches try to find domain-across parameters or priors knowledge between source domain and target domain. Relational-knowledge-transfer approaches are to acquire the common relational knowledge or some similar mappings from inputs to outputs between both domains. In this research, labeled data in both domains are available. Firstly, neural networks are trained by labeled source data. Secondly, the structure of neural networks is changed to suit the target data and some parameters are transferred to target task. Finally, the modified neural networks are trained by labeled target data. In the whole transfer learning process, some parameters and a part of structure are transferred from source task to target task. According to the properties of data and experiment conditions, parameter-transfer is the most convenient approach in this application.

D. COMPARISON WITH OTHER METHODS

Conventional rolling bearing intelligent fault diagnosis approaches usually has two steps: feature extraction from the vibration signals and fault identification using above features. Wavelet Package Decomposition (WPD) [26] and Empirical Mode Decomposition (EMD) [27] are commonly used methods for signal preprocessing and feature extraction. Lei *et al.* [28] proposed a mechanical fault diagnosis model based on WPD and EMD. The classification accuracy on bearing dataset varies from 68.33% to 95% as different features are chosen. The performance heavily depend on the selected features. However, the fault diagnosis model proposed in this paper can adaptively learn useful features from raw signals and make identification without choosing features deliberately. The proposed transfer learning fault diagnosis method can achieve 91.8% classification accuracy. The advantage is that not only the step of how to deliberately select sensitive features is omitted, but also the classification accuracy is high. And this method can utilize the data in another condition to improve the performance when the target training data are in sufficient.

V. CONCLUSION

In this paper, a transfer learning method based on neural networks is proposed for fault diagnosis of rolling bearings. This method is validated by Case Western Reserve University bearing dataset. The traditional neural network without transfer learning is implemented to make a comparison. In real word working conditions which are changing time to time, there are usually massive data in former source task and only a small amount data in later target task. What's more, the data of source task and target task may have different distributions and working conditions. Although traditional neural networks can recognize faults accurately in the same condition where there are enough training data, they do not apply to the real world working conditions where there are less training data and conditions are always changing.

Experimental results show that transfer learning method can improve the classification accuracies and reduce the training time comparing with traditional neural network method which cannot take advantage of data in different distributions and in different working conditions.

ACKNOWLEDGMENT
Acknowledgement is made for the measurements used in this work provided by Case Reserve Western University bearing database.

REFERENCES
[1] J. Chen and R. J. Patton, "Robust model-based fault diagnosis for dynamic systems," in *The International Series on Asian Studies in Computer and Information Science*, vol. 3, K. Cai, Ed. New York, NY, USA: Springer, 1999.
[2] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electro.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.
[3] F. Cong, J. Chen, G. Dong, and M. Pecht, "Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis," *J. Sound. Vib.*, vol. 8, no. 332, pp. 2081–2097, Apr. 2013.
[4] A. Giantomassi, F. Ferracuti, S. Iarlori, G. Ippoliti, and S. Longhi, *Signal Based Fault Detection and Diagnosis for Rotating Electrical Machines: Issues and Solutions* (Studies in Fuzziness and Soft Computing), vol. 319. Cham, Switzerland: Springer, 2015, pp. 275–309.
[5] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4568–4579, Jun. 2010.
[6] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (SVMs)," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4203–4211, Sep. 2011.
[7] L. Wu, B. Yao, Z. Peng, and Y. Guan, "Fault Diagnosis of roller bearings based on a wavelet neural network and manifold learning," *Appl. Sci.*, vol. 7, no. 158, pp. 1–10, Feb. 2017.
[8] 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," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.
[9] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2015.
[10] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.
[11] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–520, Nov. 2016.
[12] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248–256, Apr. 2017.
[13] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectron. Reliab.*, to be published. [Online]. Available: https://doi.org/10.1016/j.microrel.2017.03.006
[14] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, p. 895, 2016.
[15] R. Zhang, Z. Peng, L. Wu, B. Yao, and Y. Guan, "Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence," *Sensor*, vol. 17, no. 3, p. 549, 2017.
[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[17] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, p. 9, Dec. 2016.
[18] X. Li, W. Mao, and W. Jiang, "Extreme learning machine based transfer learning for data classification," *Neurocomputing*, vol. 174, pp. 203–210, Jan. 2016.
[19] C. B. Do and A. Y. Ng, "Transfer learning for text classification," in *Proc. 18th NIPS*, 2005, pp. 1–8.
[20] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. CVPR*, Jun. 2014, pp. 1717–1724.
[21] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proc. 24th ICML*, Jun. 2007, pp. 759–766.
[22] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th ICML*, Jun. 2007, pp. 193–200.
[23] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," in *Proc. 23th CVPR*, Jun. 2010, pp. 1855–1862.
[24] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *J. Mach. Learn. Res.*, vol. 10, no. 10, pp. 1633–1685, Jul. 2009.
[25] *Bearings Data Center, Seeded Fault Test Data*, Case Western Reserve University. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file
[26] L. Eren and M. J. Devaney, "Bearing damage detection via wavelet packet decomposition of the stator current," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 2, pp. 431–436, Apr. 2004.
[27] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.
[28] Y. Lei, Z. He, and Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 9941–9948, Aug. 2009.

RAN ZHANG received the B.S. degree in information engineering from Xidian University in 2015. He is currently pursuing the M.S. degree in computer software and theory with Capital Normal University. His research interests include data-driven modeling, signal processing, fault diagnosis, power electronics, and electrical vehicles.

HONGYANG TAO received the B.S. degree in international media in English from Tianjin Foreign Studies University in 2015. She is currently pursuing the M.S. degree with Capital Normal University. Her research interests include data-driven modeling and fault diagnosis.

LIFENG WU (M'12) received the B.S. degree in applied physics from the China University of Mining and Technology in 2002, the M.S. degree in detection technology and automation device from Northeast Dianli University in 2005, and the Ph.D. degree in physical electronics from the Beijing University of Post and Telecommunication in 2010. He is currently an Associate Professor with Capital Normal University. His research interests include data-driven modeling, estimation and filtering, fault diagnosis, power electronics, and electrical vehicles.

YONG GUAN (M'12) received the Ph.D. degree in information and telecommunication engineering from the China University of Mining and Technology in 2003. He is currently a Professor with Capital Normal University. His main research interests include formal verification techniques, fault diagnosis, power electronics, and electrical vehicles.