KSME

Springer

Journal of Mechanical Science and Technology 34 (11) 2020

Original Article
DOI 10.1007/s12206-020-1003-9

Keywords:
· Deep feature transfer
· Fault diagnosis
· Rolling bearing
· Varying working condition

Correspondence to:
Yujing Wang
mirrorwyj@163.com

Citation:
Kang, S., Qiao, C., Wang, Y., Wang, Q., Hu, M., Mikulovich, V. I. (2020). Fault diagnosis method of rolling bearings under varying working conditions based on deep feature transfer. Journal of Mechanical Science and Technology 34 (11) (2020) 4383~4391.
http://doi.org/10.1007/s12206-020-1003-9

Received December 16th, 2019
Revised March 7th, 2020
Accepted August 17th, 2020

† Recommended by Editor
No-cheol Park

Fault diagnosis method of rolling bearings under varying working conditions based on deep feature transfer

Shouqiang Kang[1], Chunyang Qiao[1], Yujing Wang[1], Qingyan Wang[1], Mingwu Hu[1] and V. I. Mikulovich[2]

[1]School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin 150080, China, [2]Belarusian State University, Minsk 220030, Belarus

Abstract Rolling bearing vibration data and their labels are difficult or impossible to obtain under varying working conditions. Thus, multistate identification of different fault positions and degradation degrees has relatively low accuracy. This paper proposes a fault diagnosis method based on deep feature transfer. Sparse denoising autoencoder extracts deep features of the frequency-domain amplitude sequences of rolling bearing vibration signals, and the features are used to compose feature sample sets of the source and target domains. Joint geometrical and statistical alignment adaptively processes feature samples of the source and target domains, and this way reduces the distribution divergence and subspace transform shift of the inter-domain feature samples. The classification is achieved using softmax. Experimental results show that the feature visualization effect by visualization algorithm t-SNE using the proposed method is better than those of other methods in this paper. A higher accuracy can also be achieved for rolling bearing fault diagnosis under varying working conditions.

1. Introduction

Normal operation of rolling bearings, which are an important component of rotary machine, is important in ensuring the performance of production equipment [1]. In practice, the working conditions of rolling bearings change constantly, and this condition directly affects the bearing vibration characteristics [2]. The traditional fault diagnosis methods under constant working conditions is prone to misdiagnosis or missed diagnosis under varying working conditions [3, 4] Therefore, accurately identifying the running state of rolling bearings under varying working conditions is important for the healthy operation of mechanical equipment.

In recent years, rolling bearing fault identification under different working conditions and operational state identification have been the subject of considerable research. For example, Cocconcelli et al. [5] applied envelope analysis and fault frequency identification to the detection of ball bearing faults, while Jiang et al. [6] used variational mode decomposition and the envelope order spectrum to diagnose bearing faults under varying working conditions. The abovementioned fault diagnosis methods for rolling bearings under varying working conditions are largely dependent on the experience of experts. The strong data processing ability of deep learning has considerable potential in building a fault diagnosis model with more universality, timeliness, and applicability; the applications of deep learning have been continuously explored in fault diagnosis [7]. Zhang et al. [8] described a subset-based deep autoencoder model to realize fault diagnosis of bearings, and Shao et al. [9] used a novel ensemble deep autoencoder method to diagnose faults. Shao et al. [10] utilized a tracking deep wavelet autoencoder (TDWAE) method to diagnose the fault of electric locomotive bearings. Their experimental results showed that TDWAE is more efficient than the previous automatic encoder methods. Lei et al. [11] used the frequency-domain amplitude spectrum of rolling bearings as the input to an autoencoder to identify faults in rolling bearings in different states. The abovementioned methods can diagnose bearing vibration signals under constant working conditions. However,

bearing vibration signals are more complex under varying working conditions, and obtaining the vibration characteristics that represent the running state of the bearings is difficult. In this study, a sparse denoising autoencoder (SDAE) is used to extract deep features of rolling bearing vibration signals under varying working conditions. On the basis of the autoencoder network, this method adds a sparsity limitation to its network structure to improve the learning ability of its own network. The method also obtains a sparse expression of each network layer, which is conducive to obtaining useful features in deep networks. In addition, noise-processing is conducted on the input layer, and this way can enhance the model robustness and prevent overfitting. This formulation alleviates the influence of different vibration features and solves the problem of extracting multistate deep features from rolling bearings under varying working conditions.

Deep learning is good at extracting the deep features of vibration signals from rolling bearings. However, obtaining a sufficient amount of effective vibration data and corresponding labels is difficult because of the uncertain and varying working conditions during the operation of the bearings. Some divergence exists between the distributions of data in different working conditions, and this condition leads to the domain shift phenomenon between data of the source and target domains [12, 13]. Transfer learning is a new approach that uses source domain knowledge to solve the problems in target domain. Information is acquired through related fields, and knowledge is transferred to improve the learning performance in another field. In this way, hidden meanings can be identified among different fields; this capability reduces the data divergence among the fields [14]. Xie et al. [15] mapped data of different domains to potential spaces by transfer component analysis (TCA) and found that this method obtains better results of fault diagnosis. Kang et al. [16] proposed an improved multi-kernel semi-supervised TCA (MKSSTCA) to diagnose bearing fault. Shi and Sha [17] proposed an information theoretical learning (ITL) method to measure the correlation among different data samples and improve the effect of cross-domain transfer. Zhang et al. [18] proposed a maximum independence domain adaptation (MIDA) method to reduce the discrepancy among different samples and solve the problem of drift correction. However, the abovementioned transfer learning methods consider the data divergence between domains only from the perspective of sample distributions or subspaces of different fields. The joint geometrical and statistical alignment (JGSA) algorithm can be used to reduce the shift between samples in different domains statistically and geometrically; this way avoids the limitation of single data-centric or subspace-centric domain adaptation [19].

In this study, a fault diagnosis method based on deep feature transfer that combines SDAE and JGSA is presented to address the problem of the large divergence between vibration signal data obtained from rolling bearings under varying working conditions. A deep neural network is constructed layer-by-layer through unsupervised learning. SDAE is used to abstractly and adaptively extract features of the complex frequency-domain vibration data, and the features are used to compose deep feature sample sets of the source and target domains. The transfer learning JGSA algorithm is then introduced to adjust the conditional distribution and subspace mapping transformation of the deep feature samples. This way also reduces the divergence between the vibration signal feature samples of rolling bearings under varying working conditions. Finally, multistate fault diagnosis of rolling bearings of different fault positions and degradation degrees under varying working conditions is realized by softmax classifier.

2. Principle of SDAE

2.1 Sparse autoencoder

$h_j(x)$ is the activation value of a hidden neuron when the input to the sparse autoencoder (SAE) is x. The average activation value of hidden neuron j is expressed as

$$\hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}\left[h_j(x_i)\right], \tag{1}$$

where x_i denotes the eigenvector corresponding to the ith sample and m is the number of training samples.

When the average activation degree of neurons in the hidden layer is very small, this conditon can be interpreted as the sparsity limit, which is expressed as $\hat{\rho}_j = \rho$, where ρ is the sparsity parameter (generally very near 0). To satisfy the sparsity requirement, a sparsity restriction is added to the cost function as an additional penalty factor. The specific expression is

$$\sum_{j=1}^{s_l} KL\left(\rho \| \hat{\rho}_j\right) = \sum_{j=1}^{s_l}\left(\rho \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}\right), \tag{2}$$

where s_l is the number of hidden neurons and $KL(\rho\|\hat{\rho}_j)$ is the Kullback–Leibler divergence. When $\hat{\rho}_j = \rho$, the KL divergence is 0. The KL divergence increases when $\hat{\rho}_j$ diverges from ρ. Therefore, minimizing the penalty factor ensures that $\hat{\rho}_j$ is close to ρ. The sparse penalty term is defined as

$$J_{sparse}(\theta) = \beta\sum_{j=1}^{s_l} KL\left(\rho \| \hat{\rho}_j\right), \tag{3}$$

where β is the weight coefficient of $J_{sparse}(\theta)$.

The cost function of the autoencoder network structure is defined as

$$\begin{aligned} J_{cost}(\theta) &= J_{MSE}(\theta) + J_{weight}(\theta) \\ &= \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|\tilde{x}_i - x_i\right\|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{s_l}\sum_{j=1}^{m}\left(w_{ji}^{(l)}\right)^2 \end{aligned} \tag{4}$$

where $J_{MSE}(\theta)$ is the reconstruction error function of the average reconstruction error, $J_{weight}(\theta)$ is the weight decay term, $\tilde{x}_i$ is the output vector, λ is the weight parameter of $J_{weight}(\theta)$, l is

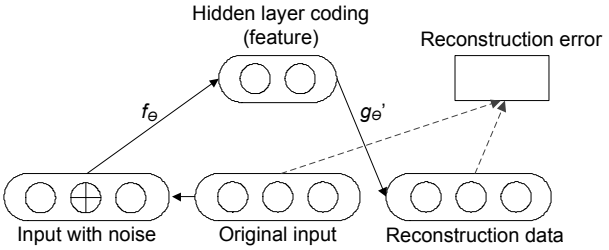


Fig. 1. Schematic of denoising autoencoder.

the number of layers of the network, and w_{ji} is the value of element *ji* in the weight matrix.

Therefore, the cost function of SAE in combination with the cost function of the autoencoder network structure can be expressed as

$$\begin{aligned} J_{cost}(\theta) &= J_{MSE}(\theta) + J_{weight}(\theta) + J_{sparse}(\theta) \\ &= \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|\tilde{x}_i - x_i\right\|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{s_l}\sum_{j=1}^{m}\left(w_{ji}^{(l)}\right)^2 + \\ &\beta\sum_{j=1}^{s_l} KL\left(\rho \,\|\, \hat{\rho}_j\right). \end{aligned} \tag{5}$$

2.2 Denoising autoencoder

An autoencoder is affected by the model complexity, the number of input data, and other factors, and is prone to overfitting when initializing the network. This phenomenon is not conducive to training a model of vibration signal data from rolling bearings under varying working conditions. To improve the robustness of the model and prevent overfitting, the input data is noised on the autoencoder. A diagram of the denoising autoencoder is illustrated in Fig. 1.

First, the denoising autoencoder sets the input data to zero randomly to obtain a noisy input signal. The specific operation is given as follows:

$$\tilde{\boldsymbol{X}} = rand(C) \times \boldsymbol{X}, \tag{6}$$

where C is the degree of added noise and $rand(\cdot)$ is a matrix with the same dimensions as the input matrix $\boldsymbol{X}$. The noisy data are input into the autoencoder for encoding and decoding according to

$$y = f_\theta(\tilde{x}) = s(\boldsymbol{W}_1\tilde{x} + \boldsymbol{b}_1) \tag{7}$$

$$z = g_{\theta'}(y) = s(\boldsymbol{W}_2 y + \boldsymbol{b}_2) \tag{8}$$

where $\boldsymbol{W}_1$ is the input layer weight matrix, $\boldsymbol{W}_2$ is the transpose of $\boldsymbol{W}_1$, $\boldsymbol{b}_1$ is the input layer offset vector, and $\boldsymbol{b}_2$ is the output layer offset vector. $\boldsymbol{z}$ is the reconstructed data given by the denoising encoder. To successfully recover the input data and minimize the cost function, the cost function of the denoising autoencoder is defined as

$$J_{Denoising}(\theta) = \frac{\sum_{i=1}^{m}\left(\frac{\left\|\tilde{x}_i - z_i\right\|^2}{2}\right)}{m} + \frac{\lambda\sum_{l=1}^{2}\sum_{i=1}^{s_l}\sum_{j=1}^{m}\left(w_{ji}^{(l)}\right)^2}{2}. \tag{9}$$

In summary, according to the principles of SAE and the denoising autoencoder and Eqs. (5) and (9), the overall cost function of SDAE can be defined as

$$\begin{aligned} J_{cost}(\theta) &= \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|\tilde{x}_i - z_i\right\|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{2}\sum_{i=1}^{s_l}\sum_{j=1}^{m}\left(w_{ji}^{(l)}\right)^2 \\ &+ \beta\sum_{j=1}^{s_l} KL\left(\rho \,\|\, \hat{\rho}_j\right). \end{aligned} \tag{10}$$

3. Principles of JGSA

The JGSA method is introduced to further reduce the divergence of different domains. JGSA simultaneously adjusts the statistical distribution of the data and applies a geometric space transformation. This way reduces the divergence among data distributions and the geometric space shift in different domains.

3.1 Minimization of the distribution divergence

The maximum mean discrepancies criterion is used to calculate the distance of the edge distribution between the source and target domain samples.

$$\min_{\boldsymbol{A},\boldsymbol{B}}\left\|\frac{1}{n_s}\sum_{x_i\in\boldsymbol{X}_s}\boldsymbol{A}^{\mathrm{T}}x_i - \frac{1}{n_t}\sum_{x_j\in\boldsymbol{X}_t}\boldsymbol{B}^{\mathrm{T}}x_j\right\|_F^2, \tag{11}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the coupled projections of the source and target domains, respectively.

The pseudo-labels for the target domain samples can be predicted using the source domain samples to train the classifier. The pseudo-labels are continuously refined through an iterative process, and the divergence of the conditional distribution between the two domains is further reduced. The calculation formula is

$$\min_{\boldsymbol{A},\boldsymbol{B}}\sum_{c=1}^{C}\left\|\frac{1}{n_s^{(c)}}\sum_{x_i\in\boldsymbol{X}_s^{(c)}}\boldsymbol{A}^{\mathrm{T}}x_i - \frac{1}{n_t^{(c)}}\sum_{x_j\in\boldsymbol{X}_t^{(c)}}\boldsymbol{B}^{\mathrm{T}}x_j\right\|_F^2, \tag{12}$$

where c denotes the category of the source and target domain samples. $c\in\{1, \ldots, C\}$, $X_s^{(c)}$ and $X_t^{(c)}$ are the sample sets of category c in the source and target domain samples, respectively. By combining the marginal and conditional distributions of data in different domains, the matrix expression for minimizing the divergence of the data distribution in different domains is

$$\min_{\boldsymbol{A},\boldsymbol{B}} Tr\left([\boldsymbol{A}^T\ \boldsymbol{B}^T]\begin{bmatrix}\boldsymbol{M}_s & \boldsymbol{M}_{st}\\ \boldsymbol{M}_{ts} & \boldsymbol{M}_t\end{bmatrix}\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix}\right), \tag{13}$$

where *Tr* represents the trace of a matrix. $\boldsymbol{M}_s$, $\boldsymbol{M}_t$, $\boldsymbol{M}_{st}$, and $\boldsymbol{M}_{ts}$ are given by Eqs. (14)-(17), respectively.

$$\boldsymbol{M}_s = \boldsymbol{X}_s\left(\boldsymbol{L}_s + \sum_{c=1}^{C}\boldsymbol{L}_s^{(c)}\right)\boldsymbol{X}_s^{\mathrm{T}}\ ,\quad \boldsymbol{L}_s = \frac{1}{n_s^2}\mathbf{1}_s\mathbf{1}_s^{\mathrm{T}}\ ,$$
$$\left(\boldsymbol{L}_s^{(c)}\right)_{i,j} = \begin{cases}\dfrac{1}{\left(n_s^{(c)}\right)^2}\ , & x_i, x_j \in \boldsymbol{X}_s^{(c)} \\ 0 & otherwise\end{cases} \tag{14}$$

$$\boldsymbol{M}_t = \boldsymbol{X}_t\left(\boldsymbol{L}_t + \sum_{c=1}^{C}\boldsymbol{L}_t^{(c)}\right)\boldsymbol{X}_t^{\mathrm{T}}\ ,\quad \boldsymbol{L}_t = \frac{1}{n_t^2}\mathbf{1}_t\mathbf{1}_t^{\mathrm{T}},$$
$$\left(\boldsymbol{L}_t^{(c)}\right)_{i,j} = \begin{cases}\dfrac{1}{\left(n_t^{(c)}\right)^2}\ , & x_i, x_j \in \boldsymbol{X}_t^{(c)} \\ 0 & otherwise\end{cases} \tag{15}$$

$$\boldsymbol{M}_{st} = \boldsymbol{X}_s\left(\boldsymbol{L}_{st} + \sum_{c=1}^{C}\boldsymbol{L}_{st}^{(c)}\right)\boldsymbol{X}_t^{\mathrm{T}}\ ,\quad \boldsymbol{L}_{st} = -\frac{1}{n_s n_t}\mathbf{1}_s\mathbf{1}_t^{\mathrm{T}},$$
$$\left(\boldsymbol{L}_{st}^{(c)}\right)_{ij} = \begin{cases}-\dfrac{1}{n_s^{(c)}n_t^{(c)}}\ , & x_i \in \boldsymbol{X}_s^{(c)}, x_j \in \boldsymbol{X}_t^{(c)} \\ 0 & otherwise\end{cases} \tag{16}$$

$$\boldsymbol{M}_{ts} = \boldsymbol{X}_t\left(\boldsymbol{L}_{ts} + \sum_{c=1}^{C}\boldsymbol{L}_{ts}^{(c)}\right)\boldsymbol{X}_s^{\mathrm{T}}\ ,\quad \boldsymbol{L}_{ts} = -\frac{1}{n_s n_t}\mathbf{1}_t\mathbf{1}_s^{\mathrm{T}},$$
$$\left(\boldsymbol{L}_{ts}^{(c)}\right)_{ij} = \begin{cases}-\dfrac{1}{n_s^{(c)}n_t^{(c)}}\ , & x_j \in \boldsymbol{X}_s^{(c)}, x_i \in \boldsymbol{X}_t^{(c)} \\ 0 & otherwise\end{cases}. \tag{17}$$

3.2 Minimization of subspace shift

The category information of the samples in the source domain and the internal feature attributes of the samples in the target domain are preserved to further decrease the inter-domain distribution shift. Moreover, $\boldsymbol{A}$ and $\boldsymbol{B}$ are optimized to ensure high similarity of the mapped domain spaces. The distance between the two coupled projections is

$$\min_{\mathbf{A},\mathbf{B}}\left\|\mathbf{A}-\mathbf{B}\right\|_F^2 . \tag{18}$$

In the subspace mapped from the target domain, the following expression is used to retain the internal feature attributes of the target domain samples while maximizing the sample variance of the mapped target domain:

$$\max_{\boldsymbol{B}} Tr\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{B}\right), \tag{19}$$

where $\boldsymbol{S}_t = \boldsymbol{X}_t\boldsymbol{H}_t\boldsymbol{X}_t^{\mathrm{T}}$ is the divergence scatter matrix of the target domain, $\boldsymbol{H}_t = \boldsymbol{I}_t - \left(1/n_t\right)\mathbf{1}_t\mathbf{1}_t^{\mathrm{T}}$ is the center matrix, and $\mathbf{1}_t \in \mathbb{R}^{nt}$ is the unit column vector.

After the mapping, the discriminant information of the source domain samples is effectively transmitted. This condition is conducive to maintaining the same category information in the mapped subspace. The expression is

$$\max_{\boldsymbol{A}} Tr\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{A}\right) \tag{20}$$
$$\min_{\boldsymbol{A}} Tr\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{A}\right) \tag{21}$$

where $\boldsymbol{S}_w$ is the intra-class scatter matrix expressed in Eq. (22) and $\boldsymbol{S}_b$ is the inter-class scatter matrix on the source domain samples defined in Eq. (23).

$$\boldsymbol{S}_w = \sum_{c=1}^{C}\boldsymbol{X}_s^{(c)}\boldsymbol{H}_s^{(c)}\left(\boldsymbol{X}_s^{(c)}\right)^{\mathrm{T}}, \tag{22}$$

$$\boldsymbol{S}_b = \sum_{c=1}^{C}n_s^{(c)}\left(m_s^{(c)} - \bar{m}_s\right)\left(m_s^{(c)} - \bar{m}_s\right)^{\mathrm{T}}, \tag{23}$$

where $\boldsymbol{X}_s^{(c)} \in \mathbb{R}^{D\times n_s^{(c)}}$ is the category set of source domain samples, $m_s^{(c)} = \left(1/n_s^{(c)}\right)\sum_{i=1}^{n_s^{(c)}}x_i^{(c)}$, $\bar{m}_s = \left(1/n_s\right)\sum_{i=1}^{n_s}x_i$, $\boldsymbol{H}_s^{(c)} = \boldsymbol{I}_s^{(c)} - \left(1/n_s^{(c)}\right)\mathbf{1}_s^{(c)}\left(\mathbf{1}_s^{(c)}\right)^{\mathrm{T}}$ is the center matrix in the category samples, $\boldsymbol{I}_s^{(c)} \in \mathbb{R}^{n_s^{(c)}\times n_s^{(c)}}$ is the identity matrix, $\boldsymbol{I}_s \in \mathbb{R}^{n_s^{(c)}}$ is the unit column vector, and $n_s^{(c)}$ is the source domain samples number.

Combining Eqs. (13) and (18)-(21) yields the objective function of the JGSA method. This function simultaneously reduces the domain shift between different domains in terms of both statistical distribution and geometric space and is written as

$$\max_{\boldsymbol{A},\boldsymbol{B}}\frac{Tr\left(\begin{bmatrix}\boldsymbol{A}^T & \boldsymbol{B}^T\end{bmatrix}\begin{bmatrix}\beta\boldsymbol{S}_b & \boldsymbol{0}\\ \boldsymbol{0} & \mu\boldsymbol{S}_t\end{bmatrix}\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix}\right)}{Tr\left(\begin{bmatrix}\boldsymbol{A}^T & \boldsymbol{B}^T\end{bmatrix}\begin{bmatrix}\boldsymbol{M}_s + \lambda\boldsymbol{I} + \beta\boldsymbol{S}_w & \boldsymbol{M}_{st} - \lambda\boldsymbol{I}\\ \boldsymbol{M}_{ts} - \lambda\boldsymbol{I} & \boldsymbol{M}_t + (\lambda+\mu)\boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix}\right)}, \tag{24}$$

where $\boldsymbol{I} \in \mathbb{R}^{d\times d}$ is the identity matrix, μ is the variance scatter matrix coefficient of the target domain, β is the coefficient of $\boldsymbol{S}_b$, and λ is the coefficient of the subspace shift in Eq. (20) after mapping the source and target domains. The objective function is maximized, and a larger variance of the target domain data distribution indicates better preservation of the internal attributes of the target domain data. For diverse domains, a smaller divergence in the statistical distribution and domain shift in geometric space are more useful for improving the similarity of data between different domains.

4. Fault diagnosis model of rolling bearings under varying working conditions

A combination of SDAE and JGSA is proposed to realize multistate fault identification when working condition is variable for further exploring the deep features of rolling bearing vibration signals and achieving their adaptive extraction. The proposed method is intended to cover the normal state, inner raceway faults, outer raceway faults, rolling ball faults, and different degradation degrees. The specific multistate fault

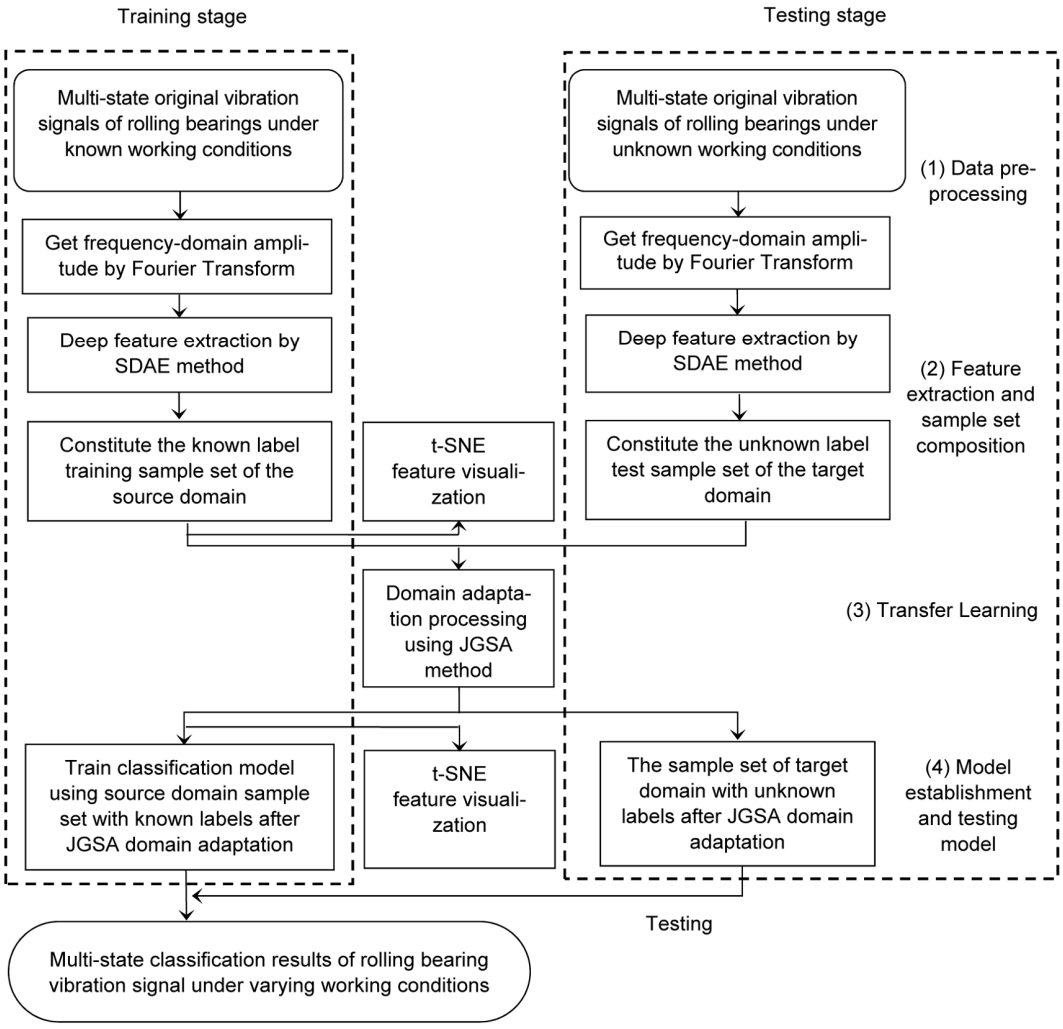


Fig. 2. Flowchart of multistate fault diagnosis method.

identification process is illustrated in Fig. 2.

The deep transfer learning method illustrated in Fig. 2 can achieve deep feature extraction of vibration signals, decrease difference among multistate vibration feature samples, and complete multistate fault diagnosis. The specific process is described as follows:

1) Data preprocessing. The multistate time domain vibration signals of rolling bearings under known and unknown working conditions are transformed into the frequency domain using a Fourier transform, and the corresponding frequency-domain amplitude sequences are obtained.

2) Feature extraction and sample set composition. Deep feature extraction of the frequency-domain amplitude sequences of vibration signals is conducted by SDAE, and the acquired features are then used to construct the training sample set of the source domain and the testing sample set of the target domain.

3) Domain adaptation processing of transfer learning. JGSA is used to process the source domain training samples and the target domain testing samples. This way reduces the difference of the conditional distribution and the shift of the subspace transformation between different domains. The method also improves the similarity of sample distribution between different domains.

4) Model establishment and testing. The softmax classifier is trained using the source domain known label sample set that processed by the JGSA. The target domain unknown label sample set is used to test the model. The multistate target domain samples of rolling bearing vibration signals under varying working conditions are classified through a continuous iterative process.

5. Application and analysis

5.1 Experimental data

The experimental data were provided by the Electrical Engineering Laboratory of Case Western Reserve University, USA. The test stand structure is shown in Fig. 3 [20]. The data were obtained using an SKF6205 deep-groove ball bearing for the drive end and an SKF6203 deep-groove ball bearing at the fan

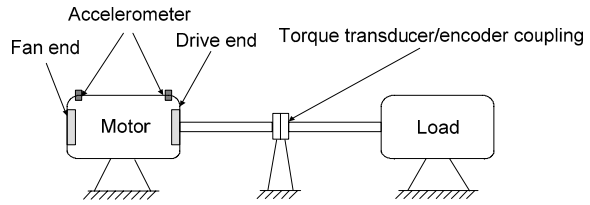


Fig. 3. Structure diagram of the test stand.

end. The 16-channel data recorder is used to collect the bearing vibration signals, and the sampling frequency includes 12 and 48 kHz.

In the experiment, vibration signals were obtained from rolling bearings under different working conditions, which included the normal state, inner raceway faults, outer raceway faults, rolling ball faults, and different performance degradation. The driving end sampling frequency was 48 kHz. Inner raceway faults with damage diameters of 7, 14, and 21 mil are denoted by IR07, IR14, and IR21, respectively. Similarly, rolling ball faults with different degrees of performance degradation are denoted by B07, B14, and B21, respectively; outer raceway faults with different degrees of performance degradation are denoted by OR07, OR14, and OR21, respectively. The normal state is denoted by N. Each sample consisted of 2048 data points, and 200 groups of data were randomly intercepted for each fault state. The sample number of each state under loads of 0, 1, 2, and 3 hp (constituting working conditions A-D, respectively) is listed in Table 1 (hp denotes British horsepower, 1 hp = 0.75 kw).

When working condition is variable, the sample set of time-domain data of rolling bearing vibration signals was transformed into a frequency-domain data sample set using Fourier transform. At this time, the dimension of the frequency-domain amplitude sequences was only half that of the original time-domain data, namely 1024 dimensional. Moreover, the frequency-domain amplitude sequences were taken as the input to the SDAE.

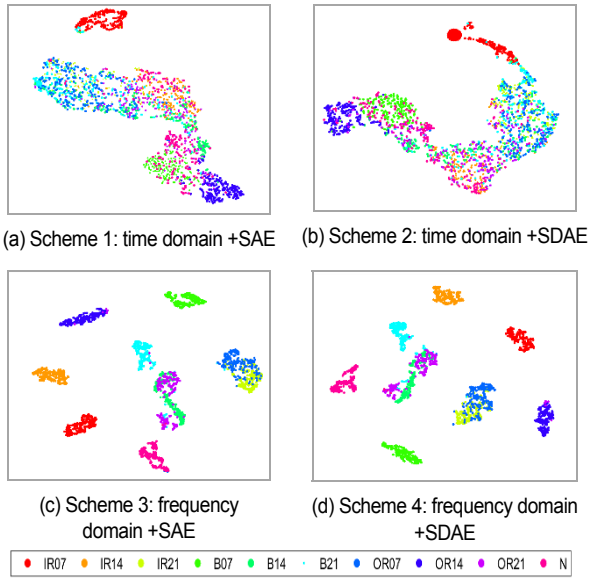
5.2 Feature extraction using SDAE

SAE and SDAE were used to extract the features of the vibration signal under varying working conditions for demonstrating the effectiveness of the deep learning method proposed in this paper. In the experiment, the input data samples were either time-domain vibration signals or frequency-domain amplitude sequences. The composition of the sample set of rolling bearing vibration signals under varying working conditions is given in Table 1. The deep features were extracted using working conditions B/C as the vibration signal data sample sets. The high-dimensional feature visualization method t-SNE was used to examine the feature extraction effects of different networks and different input data forms [21], where t-SNE achieved 1000 iterations. The visualization effects of different schemes are shown in Fig. 4.

As shownin Fig. 4, an obvious divergence exists between the

Table 1. Data sample composition of rolling bearing vibration signals under varying working conditions.

Fault location and damage degree	Working condition A	Working condition B	Working condition C	Working condition D
	Number of the samples (0 hp-1797 r/min)	Number of the samples (1 hp-1772 r/min)	Number of the samples (2 hp-1750 r/min)	Number of the samples (3 hp-1730 r/min)
IR07	200	200	200	200
IR14	200	200	200	200
IR21	200	200	200	200
B07	200	200	200	200
B14	200	200	200	200
B21	200	200	200	200
OR07	200	200	200	200
OR14	200	200	200	200
OR21	200	200	200	200
N	200	200	200	200



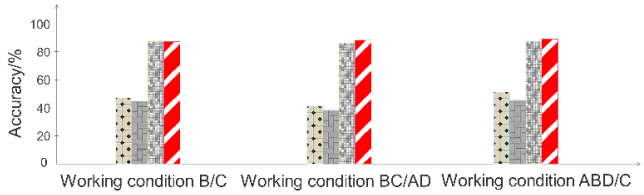
(a) Scheme 1: time domain +SAE (b) Scheme 2: time domain +SDAE

(c) Scheme 3: frequency domain +SAE (d) Scheme 4: frequency domain +SDAE

Fig. 4. Visualization effects of features obtained by different schemes.

features extracted by different schemes. Schemes 1 and 2 (Figs. 4(a) and (b), respectively) used time-domain vibration signals as input into the deep network and obtained features using SAE and SDAE, respectively. However, their effects are not ideal. This result may be due to that the input data cannot fully represent the fault types and degradation characteristics of multistate vibration signals. In schemes 3 and 4 (Figs. 4(c) and (d), respectively), frequency-domain amplitude data of the vibration signals were input into the deep network. The visualization effect is better than that of schemes 1 and 2. This finding indicates that the frequency-domain feature form of rolling bearing vibration signals provides a better reflection of the vibration signal characteristics of different states.

The fault identification accuracy is shown in Fig. 5 for further



Different schemes	Accuracy/%			Average accuracy/%
	B/C	BC/AD	ABD/C	
Scheme 1：Time domain+**SAE(softmax)**	**47.30**	**41.20**	**51.42**	**42.64**
Scheme 2：Time domain+**SDAE(softmax)**	**44.85**	**38.35**	**45.53**	**42.91**
Scheme 3：Frequency domain+**SAE(softmax)**	**87.35**	**86.50**	**87.66**	**87.17**
Scheme 4：Frequency domain+**SDAE(softmax)**	**87.45**	**88.35**	**89.20**	**88.33**

Fig. 5. Fault diagnosis accuracy of rolling bearings under varying working conditions using different schemes.

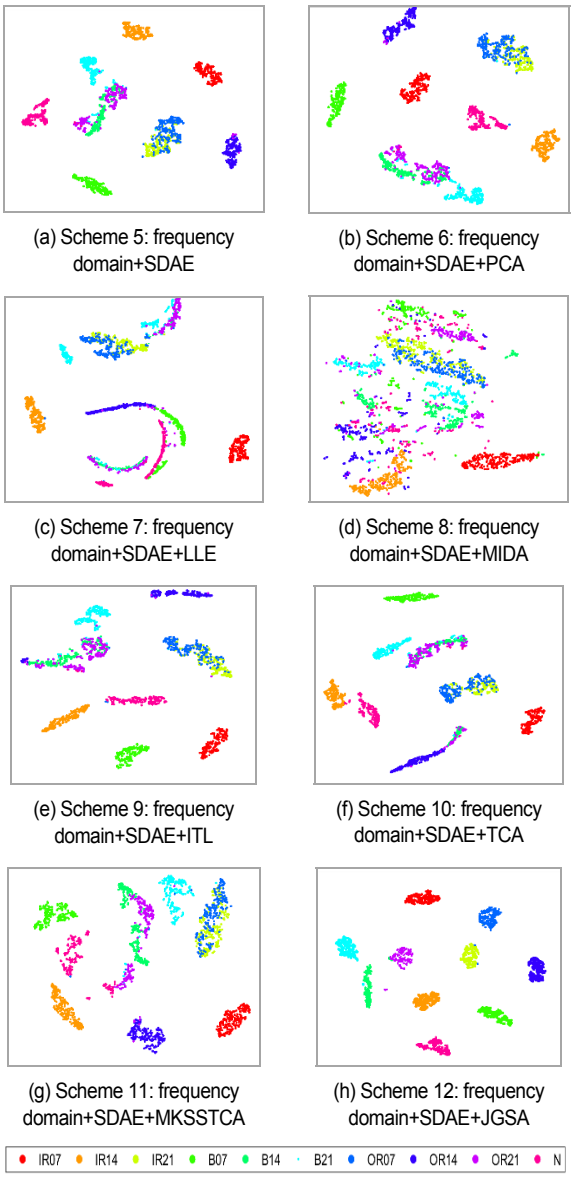
reflecting the fault diagnosis effect of multistate vibration signals using the different schemes.

Fig. 5 presents the results of using working conditions B, BC, and ABD as training feature sample sets and working conditions C and AD as testing feature sample sets. A softmax classifier was used to diagnose faults in the rolling bearings under varying working conditions. The accuracy of scheme 4 is the highest. This result is due to that the frequency-domain features better reflect the fault degradation contained in different vibration signals of rolling bearings. In addition, SDAE adds network denoising, whereby the input data are purposefully "destroyed" and the features of the vibration signal data are selectively extracted. The binding force of the neurons in the network coding and decoding is enhanced, the intra-class differentiation of deep features is increased, and the robustness of the deep model is improved.

5.3 Analysis and comparison of experimental results of JGSA domain adaptation

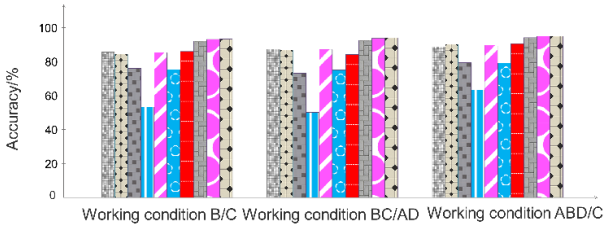
After SDAE was applied to the frequency-domain amplitude sequences of rolling bearing vibration signals, JGSA was introduced to process the resulting deep feature samples. In this case, JGSA should increase the intra-class compactness and inter-class differentiation between deep feature samples. The performance of JGSA was compared against that of PCA, locally linear embedding (LLE), MIDA [18], ITL [17], TCA [15], and MKSSTCA [16] to demonstrate its advantages. SDAE was applied with a network structure of [1024, 512, 128, 10], sparsity parameter of 0.01, sparsity penalty coefficient of 0.15, 1 % noise ratio, and a ReLU activation function. In JGSA, the coefficient μ of the variance scatter matrix in the target domain was set to 0.5, the coefficient β of the inter-class scatter matrix was set to 0.1, and the divergence coefficient λ of the two subspaces mapped by the source domain and target domain was set to 0.3. The feature visualization effects obtained using various methods are shown in Fig. 6.

As shownin Fig. 6, the visualization effects of deep features obtained by SDAE using diverse methods are quite different.



(a) Scheme 5: frequency domain+SDAE
(b) Scheme 6: frequency domain+SDAE+PCA
(c) Scheme 7: frequency domain+SDAE+LLE
(d) Scheme 8: frequency domain+SDAE+MIDA
(e) Scheme 9: frequency domain+SDAE+ITL
(f) Scheme 10: frequency domain+SDAE+TCA
(g) Scheme 11: frequency domain+SDAE+MKSSTCA
(h) Scheme 12: frequency domain+SDAE+JGSA

Fig. 6. Features visualization effects obtained by different schemes.



Different schemes	Accuracy/%			Average accuracy/%
	B/C	BC/AD	ABD/C	
Scheme 5: Frequency domain+**SDAE(KNN)**	**85.75**	**87.25**	**88.83**	**87.27**
Scheme 6: Frequency domain+**SDAE+PCA(KNN)**	**84.50**	**86.95**	**90.30**	**87.25**
Scheme 7: Frequency domain+**SDAE+LLE(KNN)**	**76.10**	**73.23**	**79.43**	**76.27**
Scheme 8: Frequency domain+**SDAE+MIDA(KNN)**	**53.40**	**50.23**	**63.48**	**55.70**
Scheme 9: Frequency domain+**SDAE+ITL(KNN)**	**85.50**	**87.26**	**89.66**	**87.47**
Scheme 10: Frequency domain+**SDAE+TCA(KNN)**	**75.35**	**75.20**	**79.25**	**77.26**
Scheme 11: Frequency domain+**SDAE+MKSSTCA(KNN)**	**86.10**	**84.15**	**90.60**	**86.95**
Scheme 12: Frequency domain+**SDAE+JGSA(KNN)**	**92.10**	**92.53**	**94.21**	**92.43**
Scheme 13: Frequency domain+**SDAE+JGSA(SVM)**	**93.20**	**93.75**	**95.00**	**93.98**
Scheme 14: Frequency domain+**SDAE+JGSA(softmax)**	**93.35**	**94.10**	**95.12**	**94.19**

Fig. 7. Fault identification accuracy of rolling bearings under varying working conditions using different schemes.

The reason is that, although schemes 6 and 7 reduce the dimension of high-dimensional features, they only change the dimension of the feature samples without substantially improving their internal attributes. As a result, a poor transformation effect of the feature samples is observed. In schemes 8-11, the deep features of the vibration signals were processed using the transfer learning domain adaptation method. Although the four schemes change the distribution properties of the feature samples to some extent, only MKSSTCA improves the discrimination and aggregation of the feature samples. The other domain adaptation methods produced a poor processing effect. In scheme 12, the JGSA method was adopted to process deep features. This scheme produces the best feature visualization effect. The reason is that JGSA can effectively improve the intra-class compactness and inter-class differentiation of the feature samples. This method can also reduce the inter-domain shift of the multistate feature sample sets of vibration signals.

Fig. 7 shows the fault identification accuracy of the different schemes for further describing the fault diagnosis effect of the deep transfer learning method on multistate vibration signals under varying working conditions.

As shown in Fig. 7, scheme 14 has the highest fault diagnosis accuracy. The reason is that, although the deep features obtained by SDAE are the same, the other domain adaptation methods only modify the edge distribution or subspace without considering their relevance. However, the JGSA method reduces the domain shift from the perspective of the conditional distribution of the feature samples. In addition, the feature samples between different domains are feature mapped to obtain the respective subspaces. The geometric domain shift is reduced in the subspace. Therefore, the proposed deep feature transfer method achieves the highest accuracy of multistate fault diagnosis of rolling bearings under varying working conditions.

6. Conclusions

1) This study investigated different forms of vibration signals from rolling bearings under varying operating conditions. The time-domain and frequency-domain amplitude signals were used as the input to a deep learning network. The experimental results show that the amplitude signal in the frequency domain can better represent the vibration characteristics of bearings than the time-domain vibration signals. The proposed method can improve the multistate fault diagnosis of rolling bearings under varying working conditions.

2) The network structure of the SDAE method was analyzed, and noise-adding process was performed on the input layer of SAE to further enhance the generalization ability of the model. For complex rolling bearing vibration signal data samples, SDAE successively learns abstract features from the lowest level to the highest level. Thus, it enables the extraction of deep features for rolling bearings under varying working conditions. SDAE has strong self-adaptive feature extraction ability, providing the foundation for further studies into rolling bearing fault diagnosis under varying working conditions.

3) The JGSA transfer learning domain adaptation algorithm was studied. The joint adjustment of the conditional distribution, edge distribution, and subspace mapping transformation of vibration feature samples was considered. The shift in the distribution domains of feature samples between different domains was reduced. As a result, the divergence between the feature samples of vibration signals was effectively reduced, and the difficulty of obtaining data and labels of rolling bearings under varying working conditions was addressed. Experiments verify that the proposed method based on deep feature transfer achieves higher fault diagnosis accuracy than previous methods.

The next steps are to study the theories of deep learning and

transfer learning and improve the multistate diagnosis accuracy of rolling bearings under different working conditions.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51805120), the Natural Science Foundation of Heilongjiang Province (LH2019E058), the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2017091), and the Fundamental Research Foundation for Universities of Heilongjiang Province (LGYC2018JC022).

References

[1] S. T. Wan and B. Peng, The FERgram: a rolling bearing compound fault diagnosis based on maximal overlap discrete wavelet packet transform and fault energy ratio, *Journal of Mechanical Science and Technology*, 33 (1) (2019) 157-172.

[2] H. M. Zhao, H. L. Liu, J. J. Xu, C. Guo and W. Deng, Research on a fault diagnosis method of rolling bearings using variation mode decomposition and deep belief network, *Journal of Mechanical Science and Technology*, 33 (9) (2019) 4165-4172.

[3] W. Zhang, C. H. Li and G. L. Peng, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment different working load, *Mechanical Systems and Signal Processing*, 100 (2018) 439-453.

[4] J. Park, M. Hamadache and J. M. Ha, A positive energy residual (PER) based planetary gear fault detection method under variable speed conditions, *Mechanical Systems and Signal Processing*, 117 (2019) 347-360.

[5] M. Cocconcelli, L. Bassi and C. Secchi, An algorithm to diagnose ball bearing faults in servomotors running arbitrary motion profiles, *Mechanical Systems and Signal Processing*, 27 (1) (2012) 667-682.

[6] Z. W. Jiang, J. D. Zheng and H. Y. Pan, Fault diagnosis under variable conditions based on parameter optimized variational mode decomposition and envelope order spectrum, *Journal of Vibration, Measurement & Diagnosis*, 37 (3) (2017) 609-616+636.

[7] Y. Lecun, Y. Bengio and G. Hinton, Deep learning, *Nature*, 521 (7553) (2015) 436-444.

[8] Y. Y. Zhang, X. Y. Li and L. Gao, A new subset based deep feature learning method for intelligent fault diagnosis of bearing, *Expert Systems with Applications*, 110 (2018) 125-142.

[9] H. D. Shao, H. K. Jiang and Y. Lin, A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders, *Mechanical Systems and Signal Processing*, 102 (2018) 278-297.

[10] H. D. Shao, H. K. Jiang and K. Zhao, A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings, *Mechanical Systems and Signal Processing*, 110 (2018) 193-209.

[11] Y. Lei, F. Jia and J. Lin, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, *IEEE Transactions on Industrial Electronics*, 63 (5) (2016) 3137-3147.

[12] X. N. Chen, Z. P. Wang and Z. Zhang, A semi-supervised approach to bearing fault diagnosis under variable conditions to-wards imbalanced unlabeled data, *Sensors*, 18 (7) (2018) 2097-2114.

[13] X. Li, W. Zhang and Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, *IEEE Transactions on Industrial Electronics*, 66 (7) (2019) 5525-5534.

[14] F. Z. Zhuang, P. Luo and Q. He, Survey on transfer learning research, *Journal of Software*, 26 (1) (2015) 26-39.

[15] J. Y. Xie, L. B. Zhang and L. X. Duan, On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis, *IEEE International Conference on Prognostics and Health Management*, Chengdu, Sichuan, China (2016) 1-6.

[16] S. Q. Kang, M. W. Hu and Y. J. Wang, Fault diagnosis method of a rolling bearing under variable working condition based on feature transfer learning, *Proceedings of the CSEE*, 39 (3) (2019) 764-773.

[17] Y. Shi and F. Sha, Information-theoretical learning of discriminative clusters for unsupervised domain adaptation, *Proceedings of the 29th International Conference on International Conference on Machine Learning*, Edinburgh, Scotland (2012) 1275-1282.

[18] D. Zhang, D. Guo and K. Yan, Drift correction using maximum independence domain adaptation, *Breath Analysis for Medical Applications* (2017) 157-178.

[19] J. Zhang, W. Q. Li and P. Ogunbona, Joint geometrical and statistical alignment for visual domain adaptation, *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA (2017) 5150-5158.

[20] *Seeded Fault Test Data*, Bearing Data Center, Case Western Reserve University, http://csegroups.case.edu/bearingdatacenter/home.

[21] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research*, 9 (2008) 2579-2605.

Shouqiang Kang obtained his Ph.D. from Belarusian State University, Minsk, Belarus, in 2011. He is currently a Professor and supervisor for Ph.D. students in Harbin University of Science and Technology, Harbin, China. His research interests include non-stationary signal processing, pattern recognition, and intelligent fault diagnosis and assessment.

Yujing Wang obtained her Ph.D. from Harbin Institute of Technology, Harbin, China, in 2015. She is currently an Associate Professor and supervisor for M.Sc. students in Harbin University of Science and Technology, Harbin, China. Her research interests include non-stationary signal processing, pattern recognition, and intelligent fault diagnosis and assessment.