A Novel Unsupervised Learning Method Based on Cross-Normalization for Machinery Fault Diagnosis

ZONGZHEN ZHANG\textsuperscript{1}, SHUNMING LI\textsuperscript{1}, JIANTAO LU\textsuperscript{1}, JINRUI WANG\textsuperscript{2}, AND XINGXING JIANG\textsuperscript{3}

\textsuperscript{1}College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
\textsuperscript{2}College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266000, China
\textsuperscript{3}School of Rail Transportation, Soochow University, Suzhou 215137, China

Corresponding author: Shunming Li (smli@nuaa.edu.cn)

This work was supported in part by the Major National Science and Technology Projects under Grant 2017-IV-0008-0045, in part by the National Natural Science Foundation of China under Grant 51675262 and Grant 51975276, in part by the Advance Research Field Fund Project of China under Grant 61400040304, and in part by the National Key Research and Development Program of China under Grant 2018YFB2003300.

\textbf{ABSTRACT} Sparse representation is the important principle of unsupervised learning method. In order to accurately identify the fault condition of machines, the desired feature distribution should show population sparsity and lifetime sparsity. In this paper, to improve the accuracy and robustness of the classification, a novel fault diagnosis method named Cross-sparse Filtering (Cr-SF) is proposed based on the cross $l_{1/2}$-norms of the feature matrix, which mean the population sparsity and lifetime sparsity terms. After the weights training process, a novel nonlinear activation function is used for feature extraction in the test process. Cr-SF can learn discriminative features from the raw data and accurately identify the fault condition. Rolling bearing fault and gear-box fault datasets are employed to validate the performance of the proposed method. The verification results confirm that Cr-SF is an effective tool for handling big data. The robustness and accuracy of the classification results using Cr-SF are comparable to convolutional networks with a much faster training process.

\textbf{INDEX TERMS} Cross-normalization, unsupervised learning, intelligent fault diagnosis, rotating machinery, big data.

\section{INTRODUCTION}
Rotating machinery is the important component of the modern mechanical systems. As the main components of rotating machinery, bearings and gears are prone to failure during the runtime, which may reduce the working efficiency and even cause accidents and disasters \cite{1}, \cite{2}. In order to ensure the safety and effectiveness of the machines operation and avoid the safety accidents and economic losses, intelligent fault diagnosis of the rotating machinery has attracted widely attention \cite{3}, \cite{4}. Deep learning is an effective tool for learning the discriminative information which can reduce the dependence of human labor and prior knowledge. Therefore, it has become a promise diagnosis technology for the era of big data \cite{5}, \cite{6}.

In recent years, many deep learning methods have been developed for the machinery fault diagnosis \cite{7}. In \cite{8}, stacked denoising autoencoder is firstly employed for the rotating machinery diagnosis. Jia \textit{et al.} proposed a local layer-based autoencoder with the normalized weights for the rolling bearing fault diagnosis and the weight vectors are visually interpreted and studied in time domain and frequency domain \cite{9}. Convolutional neural network (CNN), which is a popular and effective deep learning method, has attracted many researches due to its stable feature extraction performance and strong noisy adaptability \cite{10}--\cite{13}. An \textit{et al.} proposed a rolling bearings fault diagnosis method using feedback mechanism convolutional neural network-sparse representation \cite{14}. Sparse coding is used to sparsely express
the deep learning model to remove redundant information. However, the limitation of CNN-based method for real applications is that the training process is very time consuming. It should be noticed that these methods have many parameters to be weighted, which is also a time-consuming and priori-based process.

In the application of intelligent fault diagnosis, the key is to distinguish different samples and identify fault conditions accurately. Therefore, unsupervised learning method, which is a simple and efficient algorithm when it just focuses on the feature distribution but not the accurate expression of raw data, can make the diagnosis simpler and more intelligent. In [6], sparse filtering and softmax regression are employed to construct a two layers network structure for the rotating machinery fault diagnosis. Extracted features are employed to train the softmax regression and identify the health condition. Sparse representation is the core principle in unsupervised learning method, sparse filtering using the row normalization and column $l_1/2$-norm of the normalized features to achieve the sparse distribution of features [15]. Zhang et al. discussed the population and lifetime sparsity with the generalized normalization and developed the general normalized sparse filtering for fault diagnosis [16]. To address the deficiency that SF fails to consider the local structure of input samples, Zhang et al. proposed a novel method constructed by SF and a local structural regularization formulated to preserve the local structure of the input samples [17]. In addition, sparse filtering can also be applied in the weak signature detection [18]. Jia et al. proposed a blind deconvolutional method using convolutional sparse filtering and developed its general normalized 1-dimentional sparse filtering for impulsive signature enhancement [19].

Usually, the number of segments need to be set in the training which can change the dimension of training matrix and improve the feature learning ability of the algorithm. In the testing process, the weight vector has been trained, the function of segment number ($N_{s} \in \mathbb{N}$) is to average the test features of each sample in this case. However, both $N_{s1}$ and $N_{s2}$ are set to the same number in most studies.

In this paper, a novel fault diagnosis method is proposed based on the cross $l_{1}/2$-norm of feature matrix to improve the robustness and accuracy of diagnosis model. A novel nonlinear activation function is introduced for feature extraction after the training process. The proposed method can learn discriminative information from original data and accurately identify the health condition. The proposed method is employed to classify the health condition of a bearing fault dataset and a gear-box fault dataset. The contributions of this paper can be summarized as follows:

First, a novel unsupervised learning method using cross-normalization is developed for the machinery intelligent fault diagnosis, which shows higher sensitivity to features and stronger feature extraction performance. Second, a novel nonlinear activation function is introduced to improve the accuracy and robustness of the diagnosis model. Third, this paper studies the influence of $N_{s1}$ and $N_{s2}$ on the performance of the algorithm.

The rest of this paper is organized as follows. In section II, standard sparse filtering and feature distribution is briefly introduced. The proposed method is detailed in section III. The application of the proposed method and the experimental results are discussed in Section IV. Finally, conclusions are given in Section V.

II. STANDARD SPARSE FILTERING AND FEATURE DISTRIBUTION

In this section, the theory of standard sparse filtering and the relationship between feature distribution and normalization are briefly introduced.

A. STANDARD SPARSE FILTERING

As shown in Fig. 1, sparse filtering can be regarded as a two-layer network [20], [21]. The features $f_{j} \in \mathbb{R}_{\times M}$ can be obtained using $f = \text{Wx}$, where $\{x_{i}^j\}_{i=1}^M$ are the training samples, $\text{W} \in \mathbb{R}_{\times N}$ means the weight matrix, $M$ is the number of samples, $f_{j}$ denotes the $j$th feature of the $i$th sample.

First, the normalized features are obtained using $l_{2}$-normalization, as shown in (1) and (2).

$$\hat{f}_{j} = \frac{f_{j}}{\|f_{j}\|_{2}}$$

(1)

$$\tilde{f}_{i} = \frac{f_{i}}{\|f_{i}\|_{2}}$$

(2)

Then, the $l_{1}$-norm of $\tilde{f}_{i}$ is used as the objective function for sparseness optimization which can be written as

$$J_{sf}(f) = \sum_{i=1}^{m} \|\tilde{f}_{i}\|_{1}$$

(3)

B. FEATURE DISTRIBUTION

In the field of intelligent fault diagnosis, the discrimination of features between different fault conditions is important for the classification accuracy. Competition exists among all features because of the $l_{2}$-norm constraint of the features across rows. When one feature increases, the other feature will decrease to ensure that all features are lied on their $l_{2}$-norm sphere. This is actually a cross-competition, which makes the feature matrix sparse in columns and discriminative in rows. Finally, the final features of samples with similar faults are uniformly distributed by averaging the local features of segments. However, one obvious weakness of sparse filtering is that it requires a lot of samples for training. It is difficult to extract features from a single fault signal. Meanwhile, when the training samples or categories are small, the performance of the algorithm will deteriorate. In this paper, we propose another sparsity optimization method, which also achieves cross-competition between rows and columns. The robustness of feature learning performance of a single fault signal is enhanced. Meanwhile, higher accuracy can be guaranteed with fewer training samples.
III. PROPOSED METHOD

The proposed method using cross-normalization is detailed in this section, and the proposed algorithm is used to construct a diagnosis method for the rotating machinery fault diagnosis.

A. PROPOSED METHOD WITH CROSS-NORMALIZATION

There are two items in the objective function of the proposed method: $l_{1/2}$-norms of rows and $l_{1/2}$-norms of columns of feature matrix, which is different from the objective function of sparse filtering. These two terms are simultaneously optimized to realize the cross-optimization of population sparsity and lifetime sparsity.

Specifically, supposing that $N_{in}$ and $N_{out}$ are the input dimension and output dimension of the proposed method, the input signal $x \in \mathbb{R}^{N_{in}}$ is transformed into a segment matrix $S \in \mathbb{R}^{N_{in} \times M}$ using random segmentation, where $M = N_{in} \times m$, $N_{in}$ is the number of segments in the training process, $m$ means the number of training samples.

The linear expression of the training matrix and the weight, $F = WS$, is used to obtain the features, where $W \in \mathbb{R}^{N_{out} \times N_{in}}$ is the weight matrix. Then, each row and column of the feature matrix are normalized using their $l_2$-norm, respectively. As shown in (4) and (5)

$$
\mathcal{L}_1 = \frac{1}{2} \left\| F^i \right\|_{l_2}^2 \quad (4)
$$

$$
\mathcal{L}_2 = \frac{1}{2} \left\| F_j \right\|_{l_2}^2 \quad (5)
$$

In order to eliminate the influence of redundancy in optimization process, the weight vectors are constrained to unit vectors. The final objective function of the proposed method is written as

$$
\mathcal{L} = \sum_i^M \mathcal{L}_1 + \lambda \sum_j^N \mathcal{L}_2 \quad S,T \left\| W^i \right\|_{l_2} = 1 \quad (6)
$$

where $\lambda \geq 0$ control the tradeoff between the two terms. $L$ is nonsmooth and nonconvex. Therefore, $|F|$ is replaced by $\sqrt{F^2 + \varepsilon}$, which is a soft-absolute function and $\varepsilon$ equals $1 \times 10^{-8}$. Then, L-BFGS algorithm is used to optimize the objective function. The gradient function is given by

$$
\frac{\partial L}{\partial w} = \left[ \frac{O}{(\sum_{i=1}^{N_{out}} |F_i|^2)^{1/2}} - \frac{F_i \sum_{j=1}^{N_{in}} (F_j)}{(\sum_{i=1}^{N_{out}} |F_i|^2)^{1/2}} \right] \cdot \frac{f}{|f|} \cdot x' \quad (7)
$$

where $o \in \mathbb{R}^{N_{out} \times M}$ is a matrix of all ones.

It should be noted that $\lambda = 0$ is a special condition, in which the sparse feature between samples is not optimized, and the algorithm only ensure the sparse distribution of features per sample.

B. DIAGNOSIS MODEL USING THE PROPOSED METHOD

This section presents an intelligent diagnosis model based on the proposed algorithm. The flow chat of the diagnosis model is shown in Fig. 2 and the general procedures are summarized as follows.

Step 1: Training matrix construction. The original signal $\{x^i\}_{i=1}^M$ are segmented into a training set $S \in \mathbb{R}^{N_{in} \times M}$.

Step 2: Feature activation using the linear expression and normalization following Eq. 4 and 5.

Step 3: Weights training using the proposed algorithm.

Step 4: Feature extraction using the non-linear activation function $\log(1 + WS)$ and the trained weights. This activation function is different from that of the training process. The waveform of the activation function is shown in Fig 3.

Step 5: Classifier training process. In this step, the extracted features $\{f_i\}_{i=1}^m$ and the corresponding label set $\{y^i\}_{i=1}^m$ are fed into the softmax classifier for training, where $y^i \in \{1, 2, 3, \ldots, k\}$ is the label vector. and the weight parameters of the classifier are obtained. The hypothesis $h_\theta (f)$ takes the form:

$$
h_\theta (x') = \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T f)} \left[ \sum_{j=1}^{k} \exp(\theta_j^T f) \right]
$$

where $\theta_1, \theta_2, \ldots, \theta_k$ are the parameters of softmax model and $p \left( y^i = j \mid f \right)$ is the probability for each $f_i$. 

FIGURE 1. Sketch of sparse filtering.
The objective function of the softmax regression can be expressed as:

\[ J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y^j = 1\} \frac{\exp(\theta_T f^i)}{\sum_{l=1}^{k} \exp(\theta_T f^i)} + \frac{\lambda}{2} \sum_{k=1}^{K} \sum_{l=1}^{N} \theta_{kl}^2 \]  

where \( m \) is the number of training samples, \( 1\{\cdot\} \) is the indicator function, and \( k \) is the category number.

Step 6: Feature activation of test samples using the non-linear activation function introduced in step 4. The collected signal is segmented into matrix \( S_2 \in \mathbb{R}^{N_{10} \times N_{2}} \).

Step 7: Diagnostic results. The extracted test features are fed into the trained classifier and the diagnostic results are given.

IV. EXPERIMENTAL VERIFICATION

A. DATA DESCRIPTION

In this experiment, the rolling bearings fault dataset provided from the Case Western Reserve University Lab are used to validate the performance of the proposed method. The main components of the motor bearing test bench are the testing bearings, an induction electrical motor and an acceleration sensor. The vibration sensor is fixed at the drive end of the motor. Four different health cases are designed, including normal condition (NC), outer race fault condition (OF), inner race fault condition (IF) and roller fault condition (RF). For each fault case, there are three different severity levels (0.18, 0.36, and 0.53 mm) and 1200 data points in each sample with the sampling frequency of 12 kHz. Each health condition includes 100 vibration samples under one load. There are totally ten kinds of fault conditions and 1000 samples for this study, which is presented in Table 1.
TABLE 1. Bearing fault conditions description.

| Fault condition | Fault diameter (mm) | Sample number | Condition label |
|-----------------|---------------------|---------------|----------------|
| NC              | -                   | 400           | 1              |
| OF              | 0.18                | 400           | 2              |
| OF              | 0.36                | 400           | 3              |
| OF              | 0.53                | 400           | 4              |
| IF              | 0.18                | 400           | 5              |
| IF              | 0.36                | 400           | 6              |
| IF              | 0.53                | 400           | 7              |
| RF              | 0.18                | 400           | 8              |
| RF              | 0.36                | 400           | 9              |
| RF              | 0.53                | 400           | 10             |

FIGURE 4. Sparsity of the sample features using different method.

B. DISCUSSION ON THE FEATURE LEARNING PERFORMANCE

The feature learning performance of the proposed algorithm for single fault signals is firstly studied. Roller fault condition is selected for this analysis. Each sample is randomly divided into 100 segments containing 100 data points. The extracted features and the corresponding sparsity measured by $l_{1/2}$-norm are shown in Fig. 4 and Fig. 5, respectively. The sparsity of feature measured by $l_{1/2}$-norm using the proposed method is about 6. However, the feature sparsity of features using sparse filtering is about 9.4. The distribution of features extracted by the proposed method shows obvious sparsity within the sample. However, the features extracted by sparse filtering show little difference. The spectral of the trained weight vectors are shown in Fig. 6. Compared to the sparse filtering, vectors trained by the proposed method corresponding to the features with larger values show obvious frequency components. When the amplitude is small, the weights have no frequency component, which means a strong filtering performance. We can draw the conclusion that the proposed method shows stronger feature extraction performance and higher sensitivity to features.

C. CASE STUDY 1: ROLLING BEARING FAULT DIAGNOSIS

In this experiment, the fault dataset detailed in the previous section are used to validate the performance of the proposed method as displayed in Table 1. We randomly select 10% of the samples as the training data. The accuracy and robustness of diagnostic results under different value of lambda are firstly analyzed to demonstrate the performance of the proposed method. Each original data is randomly divided into 100 segments with a dimension of 100. The weight decay term of softmax is set as 1E-5. The output and input dimension are both 100. In this paper, to reduce the randomness, the diagnostic results of experiment under each parameter setting are averaged by 20 repeated trials and the standard deviations are represented using the error bars (the computation platform is a PC with an Intel I7 CPU and 8G RAM).

1) PROPERTY OF THE PROPOSED METHOD

Fig. 7 presents the influence of $W$-norm, activation function and lambda. When $\lambda > 10$, the diagnostic accuracy will be significantly reduced. When the value of $\lambda$ is small, the robustness and accuracy are better. When $\lambda = 0$, only column constraints exist in the objective function. The accuracy of the proposed method can still achieve 99.59% in this case. However, the calculation time is a little longer when $\lambda = 0$. The calculation time is 13.2s, when lambda is not equal to 0, the training time is about 12.8s. Through the comparison of three cases, we can see that $W$-norm and activation function can significantly improve the diagnostic accuracy rate of the diagnosis model. When the value of Lambda is small, $W$-norm can improve the accuracy more obviously. By the contrast, when the value of lambda is large, the effect of activation function is more obvious. Fig 7 also shows that lambda has a wide suitable range that can achieve enough accuracy, which indicates that lambda is a very easy parameter to choose.

Through the implementation of this method, we can infer that $W$-norm mainly affects the feature extraction and optimization of the algorithm. The activation function is not in the training process of the algorithm, it mainly affects the extracted features and distinguishability. Therefore, to further explore the property of the proposed algorithm, we analyze the weight vector and the corresponding spectrum.

Fig. 8 is the spectrum color map of the weight vectors obtained by different methods. The weight vectors have obvious high-frequency noise under constraints. In order to realize the sparse distribution between features, the proposed method will generate high-frequency or low-frequency based weight vectors. These vectors have little practical significance in
the process of feature extraction. Another advantage of the $W$-norm is that the frequency components of the weight is more complete, similar to sparse filtering, which can guarantee to learn more complete information. In case of without $W$-norm, the energy gap between the frequency components weights is large, and the extracted features are not comprehensive. This also explains why the performance of constrained case is better than that of unconstrained case.

2) PERFORMANCE WITH DIFFERENT SEGMENT NUMBERS
In the process of fault diagnosis, there are two steps needs to set the number of segments of the original sample. Generally, these two numbers will be studied as one in most papers. In fact, their function is quite different. The first segment number is set in the training process, which can change the dimension of training matrix and improve the feature learning performance of the algorithm. The second segment number is set in the test process. In this case, the weight vectors have been trained, and the segment number is mainly used as the average of a feature. The influence of $N_{s1}$ and $N_{s2}$ on the performance of the algorithm is studied in this experiment.

Fig. 9 displays the influence of $N_{s1}$ on the diagnosis performance and the comparison results with sparse filtering with various $N_{s2}$. When the segment number $N_{s1}$ is increasing with small value of $N_{s2}$, the accuracy is becoming higher and the corresponding standard deviations are continuously smaller. When $N_{s2} > 100$, the increase of $N_{s1}$ cannot improve the accuracy, but only increase the training time. Even when $N_{s1} = 5$, the accuracy is higher than 99% as long as $N_{s2}$ is greater than 50. On the contrary, even if $N_{s1}$ is taken as 500, the accuracy is only 98% with $N_{s2} = 20$. As can be seen from Fig 10, In any case of $N_{s1}$, the accuracy become higher with the increase of $N_{s2}$. As long as $N_{s2}$ is greater than 50, the accuracy of the proposed method can reach above 99%. We can also see that under the same parameter settings, the accuracy of the proposed method is significantly higher than that of sparse filtering. Even when the values of $N_{s1}$ and $N_{s2}$ are small, the accuracy of the proposed method can still achieve high accuracy. $N_{s1}$ can change the dimension of training matrix and affect the calculation efficiency. As shown in Fig. 11, with the increase of $N_{s1}$, the computation time shows a linear increase trend.

**FIGURE 5.** Sparsity of the sample features using different method.

**FIGURE 6.** Frequency information of the trained weights for the roller fault.
As a summary, $N_{s2}$ is a very important parameter for classification accuracy and $N_{s1}$ has a great influence on the calculation efficiency. To ensure the accuracy and calculation efficiency through the above analysis, the calculation efficiency and accuracy with $N_{s1} = 20$ and $N_{s2} = 100$ are selected as the comparison.
They are sparse filtering in [6], GNSF and PCA in [17], CNN in [13] and the proposed method. The dataset under four working conditions are used as training samples. In this comparison, the mean diagnostic accuracy and computation time are the final measure term. The comparison results are displayed in Table 2. Sparse filtering, GNSF with $p = 3$ and $q = 2$ and the proposed method have the same parameter settings.

The accuracy of the proposed method which can reach 99.85%, is better than that of sparse filtering and its improved method GNSF under the same parameter settings, and is close to that of CNN (99.63%). However, CNN needs more training samples. In addition, the training process is time-consuming. The calculation efficiency of the proposed method is significantly higher.

**D. CASE STUDY 2: PLANETARY GEARBOX FAULT DIAGNOSIS**

The gear fault datasets collected from the bearing seat at the drive end of gearbox are used to demonstrate the
performance of the proposed diagnosis method in this experiment. As depicted in Fig. 11, the test devices mainly include a driving electrical motor, a planetary gearbox, a vibration sensor and a tachometer. The rotation speed of the driving motor is maintained at 1500 rpm. There are 4 health conditions in this experiment: normal condition (NC), worn fault (WF), broken tooth (BT) and compound fault between WF and BT (CF). The samples number in each health condition is 100 with the sampling frequency of 12.8KHz. Therefore, this experiment contains 400 samples and is a four-class classification problem, as displayed in Table 3. It should be noticed that compared with the data of CWRU Lab, the fault dataset of gearbox is collected in the machining workshop which means that environmental noise exists in the collected signals.

Fig 13 and 14 shows the diagnostic results using different methods. The accuracy gradually going higher with the increase of \( N_{s1} \). When \( N_{s2} > 50 \), \( N_{s1} \) has little effect on the accuracy. However, increasing \( N_{s1} \) at this time will significantly increase training time. The standard deviation of the diagnostic result of sparse filter is larger, which shows that the robustness is poor. As can be seen from Fig 13, the accuracy of the proposed method and sparse filtering are both becoming higher with the increase of \( N_{s2} \). Sparse filtering has much lower accuracy compared with the proposed method. The proposed method can still get higher accuracy with fewer segments. For example, when
$N_{s1} = 20$ and $N_{s2} = 50$, the accuracy of the proposed method is above 98%. However, the accuracy of sparse filtering is only 87% and the highest accuracy of sparse filtering is only 97% with cases of $N_{s1} = 100$, $N_{s2} = 100$.

Table 4 presents the comparison results of various methods with $N_{s1} = 5$ and $N_{s2} = 100$. Similar to the results of bearing fault diagnosis, the proposed method can achieve 99.64% test accuracy with 1.73s training time. The calculation efficiency of sparse filtering and the proposed method is close, but the accuracy of the proposed method is significantly higher than that of standard sparse filter and its variant. CNN can achieve the accuracy of 99.13%, but it needs 95s for training, which is significantly more time-consuming than the proposed method.

V. CONCLUSION

Based on cross-normalization, a novel unsupervised learning method is proposed for machinery fault diagnosis in this paper. A novel nonlinear activation function is introduced to improve the robustness and accuracy. The proposed method is applied in the diagnosis of the rolling bearing and gear-box fault datasets.

The proposed method shows higher sensitivity to features and stronger feature learning performance. The extracted features using the proposed method show better sparsity. The
proposed feature learning algorithm and nonlinear activation function can improve the robustness and accuracy of the diagnosis model.

The research on the two segment numbers shows that larger segment number in the test process can improve the accuracy, and smaller segment number in the training process can significantly improve the calculation efficiency. Therefore, in practical application, we can choose a small number of training segments and a large number of test segments to ensure both accuracy and training efficiency.

The comparison experiment indicated that the proposed method is a promising fault diagnosis technology, which can significantly improve the accuracy and training efficiency.

REFERENCES

[1] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, “A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines,” Mech. Syst. Signal Process., vol. 116, pp. 668–692, Feb. 2019.

[2] Y. Xin, S. Li, Z. Zhang, Z. An, and J. Wang, “Adaptive reinforced empirical Morlet wavelet transform and its application in fault diagnosis of rotating machinery,” IEEE Access, vol. 7, pp. 65150–65162, 2019.

[3] X. Zhou, F. Lu, and J. Huang, “Fault diagnosis based on measurement reconstruction of HPT exit pressure for turbofan engine,” Chin. J. Aeronaut., vol. 32, no. 5, pp. 1156–1170, May 2019.

[4] Z. Wang, W. Du, J. Wang, J. Zhou, X. Han, Z. Zhang, and L. Huang, “Research and application of improved adaptive MOMEDA fault diagnosis method,” Measurement, vol. 140, pp. 63–75, Jul. 2019.

[5] S. Yin and O. Kaynak, “Big data for modern industry: Challenges and trends [point of view],” Proc. IEEE, vol. 103, no. 2, pp. 143–146, Feb. 2015.

[6] Y. Lei, F. Jia, J. Lin, S. Xing, and X. Ding, “An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, May 2016.

[7] J. Wang, S. Li, B. Han, Z. An, Y. Xin, W. Qian, and Q. Wu, “Construction of a batch-normalized autoencoder network and its application in mechanical intelligent fault diagnosis,” Meas. Sci. Technol., vol. 30, no. 1, Jan. 2019, Art. no. 015106.

[8] J. 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.

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

[10] Y. Yang, H. Zheng, Y. Li, M. Xu, and Y. Chen, “A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network,” ISA Trans., vol. 91, pp. 235–252, Aug. 2019.

[11] Z. Chen, K. Gryllias, and W. Li, “Mechanical fault diagnosis using convolutional neural networks and extreme learning machine,” Mech. Syst. Signal Process., vol. 133, Nov. 2019, Art. no. 106272.

[12] J. Yuan and Y. Tian, “A multiscale feature learning scheme based on deep learning for industrial process monitoring and fault diagnosis,” IEEE Access, vol. 7, pp. 151189–151202, 2019.

[13] L. Wen, X. Li, L. Gao, and Y. Zhang, “A new convolutional neural network based data-driven fault diagnosis method,” IEEE Trans. Ind. Electron., vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[14] F.-P. An, “Rolling bearing fault diagnosis algorithm based on FMCNN-sparse representation,” IEEE Access, vol. 7, pp. 102249–102263, 2019.

[15] J. Njiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. H. Ng, “Sparse filtering,” in Proc. Adv. Neural Inf. Process. Syst., 2011, pp. 1125–1133.

[16] Z. Zhang, S. Li, J. Wang, Y. Xin, and Z. An, “General normalized sparse filtering: A novel unsupervised learning method for rotating machinery fault diagnosis,” Mech. Syst. Signal Process., vol. 124, pp. 596–612, Jun. 2019.

[17] Z. Zhang, Q. Yang, and W. Zhou, “A feature extraction method based on sparse filtering with local structure preserved and its applications to bearing fault diagnosis,” IEEE Access, vol. 7, pp. 160559–160572, 2019.

[18] X. Jia, M. Zhao, Y. Di, P. Li, and J. Lee, “Sparse filtering with the generalized ℓpq norm and its applications to the condition monitoring of rotating machinery,” Mech. Syst. Signal Process., vol. 102, pp. 198–213, Mar. 2018.

[19] X. Jia, M. Zhao, Y. Di, C. Jin, and J. Lee, “Investigation on the kurtosis filter and the derivation of convolutional sparse filter for impulsive signature enhancement,” J. Sound Vib., vol. 386, pp. 433–448, Jan. 2017.

[20] J. Wang, S. Li, Y. Xin, and Z. An, “Gear fault intelligent diagnosis based on frequency-domain feature extraction,” J. Vib. Eng. Technol., vol. 7, no. 2, pp. 159–166, Apr. 2019.

[21] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, “Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines,” Neurocomputing, vol. 329, pp. 53–65, Feb. 2019.

ZONGZHEN ZHANG received the B.S. and M.S. degrees from the University of Jinan, Jinan, China, in 2008 and 2011. He is currently pursuing the Ph.D. degree with the College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include fault diagnosis of rotating machinery, noise and vibration analysis and control, weak signature detection and separation, signal processing, and machine learning.

SHUNMING LI received the Ph.D. degree in mechanics from Xi’an Jiaotong University, China, in 1988. He is currently a Professor with the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include noise and vibration analysis and control, signal processing, machine fault diagnosis, sensing and measurement technology, and intelligent vehicles.

JIANTAO LU received the Ph.D. degree from the School of Mechanical Engineering, Xi’an Jiaotong University, China, in 2019. He is currently a Lecturer with the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. His current research interests include mechanical fault diagnosis and underdetermined and nonlinear blind source separation.

JINRUI WANG received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2019. He is currently working as a Postdoctoral Researcher with the Shandong University of Science and Technology, Qingdao, China. His current research interest focuses on intelligent fault diagnosis of machines.

XINGXING JIANG received the B.S. and Ph.D. degrees in vehicle engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2012 and 2016, respectively. He is currently an Associate Professor with Soochow University, Suzhou, China. His research interests are rotating machinery condition monitoring, intelligent machinery fault diagnostics, and data mining.