DOI: 10.3901/CJME.2014.1103.166, available online at www.springerlink.com; www.cjmenet.com; www.cjmenet.com.cn

Feature Extraction and Recognition for Rolling Element Bearing Fault Utilizing Short-Time Fourier Transform and Non-negative Matrix Factorization

GAO Huizhong[1, 4], LIANG Lin[1, 2, *], CHEN Xiaoguang[1], and XU Guanghua[1, 3]

1 *School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China*
2 *Key Laboratory for Modern Design and Rotor-Bearing System of Education Ministry, Xi'an Jiaotong University, Xi'an 710049, China*
3 *State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, China*
4 *The 705 Research Institute, China Shipbuilding Industry Corporation, Xi'an 710075, China*

Received March 13, 2014; revised October 10, 2014; accepted November 3, 2014

Abstract: Due to the non-stationary characteristics of vibration signals acquired from rolling element bearing fault, the time-frequency analysis is often applied to describe the local information of these unstable signals smartly. However, it is difficult to classify the high dimensional feature matrix directly because of too large dimensions for many classifiers. This paper combines the concepts of time-frequency distribution(TFD) with non-negative matrix factorization(NMF), and proposes a novel TFD matrix factorization method to enhance representation and identification of bearing fault. Throughout this method, the TFD of a vibration signal is firstly accomplished to describe the localized faults with short-time Fourier transform(STFT). Then, the supervised NMF mapping is adopted to extract the fault features from TFD. Meanwhile, the fault samples can be clustered and recognized automatically by using the clustering property of NMF. The proposed method takes advantages of the NMF in the parts-based representation and the adaptive clustering. The localized fault features of interest can be extracted as well. To evaluate the performance of the proposed method, the 9 kinds of the bearing fault on a test bench is performed. The proposed method can effectively identify the fault severity and different fault types. Moreover, in comparison with the artificial neural network(ANN), NMF yields 99.3% mean accuracy which is much superior to ANN. This research presents a simple and practical resolution for the fault diagnosis problem of rolling element bearing in high dimensional feature space.

Keywords: time-frequency distribution, non-negative matrix factorization, rolling element bearing, feature extraction

1 Introduction

Rolling element bearings are one of the most important and common components in rotary machines. Their performances are also related to the condition of rotating machinery. Therefore, it is essential to detect the occurring fault as early as possible to avoid fatal breakdowns. Vibration-based monitoring is the most widely applied technique as the traditional method. As the fault vibration signals of bearing are non-stationary, the traditional diagnosis techniques perform from the waveforms of the fault vibration signals in the time or frequency domain[1–4], and then construct the criterion functions to identify the working condition of rolling element bearing. However, because the non-linear factors, such as loads, friction and so on, have distinct influence on the vibration signals due to the complexity of the construct and working condition, it is very difficult to make an accurate evaluation on the working condition of rolling element bearing only through the analysis in time or frequency domain.

The time-frequency distribution(TFD) can well demonstrate the periodic transient component of a vibration signal by combining the time and frequency information in a two-dimensional representation. Typical TFD methods include the short-time Fourier transform(STFT)[5], the wavelet analysis[6] and the Wigner-Ville distribution[7]. Due to the capability of energy distribution in the time-frequency domain, the TFD is beneficial to non-stationary signal analysis in machinery fault diagnosis.

However, comparing with the time or frequency domain, time-frequency distribution is usually a high dimension matrix, and will thus increase the difficulty to identify faults through traditional classifier. Therefore, it is obvious that dimension reduction to a much lower dimension is appropriate. Principle component analysis(PCA)[8], independent component analysis(ICA)[9], and singular value decomposition(SVD)[10] are applied to obtain low dimension features[11–12]. While, due to the holistic nature of PCA, the resulting components are global interpretations

* Corresponding author. E-mail: lianglin@mail.xjtu.edu.cn
Supported by Shaanxi Provincial Overall Innovation Project of Science and Technology, China(Grant No. 2013KTCQ01-06)

and lack intuitive meaning.

In order to solve this problem, LEE and SEUNG demonstrated that non-negative matrix factorization(NMF) is able to learn localized features with obvious interpretation[13]. It is a new theory for factorizing a matrix as the product of two matrices, whose elements are all non-negative. Previous extraction methods usually contain negative elements, which are physically meaningless. Major physical signals, such as pixel intensity, amplitude spectra and weight, are non-negative. Because of its non-negative constraint, NMF has a good effect on part-based representation. Therefore, NMF has a widespread adoption, particularly in the feature extraction field. LIU creatively used NMF to extract features of objects and then realize recognition[14]. PU proposed a fisher NMF algorithm by inputting the fisher method into NMF to preprocess signals. On the basis, the following sparseness constraint made the whole algorithm can achieve even better results[15]. Except for feature extraction, NMF has been adopted to various applications such as image compression and classification[16], sound separation[17], and so on. Several researches have been done to apply NMF to fault diagnosis of engine and water tank system[18–19].

In addition, identifying effect of faults with the extracted features from NMF is associated with the selected classifier. In general, support vector machine(SVM) or artificial neural networks(ANN) are applied to achieve pattern recognition. However, either SVM or ANN needs extra training stage, which is very strong sensitivity to the parameter adjustment. Thus, much more calculation and less efficiency are inevitable. However, studies empirically show that NMF emerges as a promising tool for clustering[20]. Clustering implicitly performs an adaptive dimensionality reduction at each iteration, leading to better clustering accuracy compared to traditional clustering methods, such as k-means.

Naturally, the localized fault features of interest can be extracted and recognized efficiently by using the advantages of the NMF in parts-based representation and adaptive clustering. Therefore this paper proposes a feature extraction and recognition of rolling element bearing fault from TFD based on NMF. With STFT method, the TFD of a vibration signal is achieved to describe localized faults. Integrated with supervised NMF mapping, the high dimension matrix of TFD is factorized to select localized fault features. In addition, according to the clustering property of NMF, the proposed method can accomplish clustering and identification of pattern samples automatically.

The paper is organized as follows: Section 2 provides the fundamental knowledge about STFT. In section 3, the principles of NMF are introduced. And its property in extraction and clustering field is also put forward. Then in section 4, the fault diagnosis strategy based on NMF is proposed. In section 5, vibration signals of rolling element bearing faults are presented to evaluate the proposed method. At last, a conclusion is drawn in section 6.

2 Short-Time Fourier Transform

Vibration signals of bearings are complicated with rich information. With the bearing state varying, the vibration signals will change simultaneously. A good analysis scheme can distinctly express the changes, which makes the diagnosis much easier and more reliable. STFT is a typical time-frequency analysis method, which has been widely used in signal processing field. This approach uses a window function to multiply time series, in which the non-stationary signal can approximately be considered as locally stationary, and then transformed them into time-frequency domain. We can capture the spectral components in spectrogram as discrimination with this method. STFT is able to be described as below:

$$S(t,f)=\int x(t+\tau)w(\tau)\exp(-2j\pi f\tau)\mathrm{d}\tau, \tag{1}$$

where $x(t)$ is the signal to be considered, $w(t)$ is the sliding window function(i.e. Hanning window), t is the time, and f is the frequency.

The phase of $S(t,f)$ is usually ignored because the amplitude spectrum is quite convenient to deal with. Hence, we only take the amplitude spectrum of measured signals into consideration.

Although wavelet transform has been widely used in signal analysis field, wavelet basis function exerts great influence on corresponding results. In the industrial environment, vibration signal is usually influenced by different parts. Furthermore, with the fault developing, there are different types of localized feature for different fault severities, for example, single sided impulse component and double sided impulse component, etc. Aiming at reducing the effect of artificial factors and combining all the reasons mentioned above, the signal here is too hard to be decomposed by a very mother wavelet. That’s why STFT approach is chosen as an appropriate tool due to its easy principle and good capacity.

3 Non-negative Matrix Factorization

3.1 NMF algorithm

NMF is a matrix factorization algorithm with non-negative constraints. It has been investigated by many researchers, e.g. PAATERO and TAPPER[21]. However, it is popularized by the work of LEE and SEUNG published in Nature journal[13]. Based on the point that the negativity is meaningless in human perception, they proposed a smart algorithm to find proper non-negative representations of non-negative data or images. The basic NMF problem is stated as follows: given a matrix $\boldsymbol{V}$ with $n\times m$ non-negative values, and then factorize it into two matrices $\boldsymbol{W}_{n\times r}$ and $\boldsymbol{H}_{r\times m}$ as well as possible. The process can be described as follows:

$$\boldsymbol{V}_{n\times m} \approx \boldsymbol{W}_{n\times r}\boldsymbol{H}_{r\times m}. \tag{2}$$

Additionally, the reduced rank r is generally chosen as $(n+m)r < n\times m$, hence the compression effect is accomplished. As a result, $\boldsymbol{V}$ is able to be estimated as a linear combination of the vectors of the basis matrix $\boldsymbol{W}$ and gains matrix $\boldsymbol{H}$. As the key characteristic of NMF, non-negativity makes the representation purely additive. It is quite different from the other factorization techniques, such as PCA and ICA, whose elements may be negative. In practice, the amplitude of frequency spectrum presented by negative components can not represent any physical meanings.

In order to obtain matrix $\boldsymbol{W}$ and $\boldsymbol{H}$ in Eq. (2), a variety of cost function is used. Among them, the most often used one is based on Frobenius norm:

$$\boldsymbol{D}(\boldsymbol{V}\|\boldsymbol{WH}) = \frac{1}{2}\|\boldsymbol{V}-\boldsymbol{WH}\|^2. \tag{3}$$

So an iterative multiplicative algorithm is also given to obtain $\boldsymbol{W}$ and $\boldsymbol{H}$:

$$\boldsymbol{W} = \boldsymbol{W}\otimes\left[\left(\boldsymbol{VH}^{\mathrm{T}}\right)\Theta\left(\boldsymbol{WHH}^{\mathrm{T}}+\varepsilon\right)\right], \tag{4}$$

$$\boldsymbol{H} = \boldsymbol{H}\otimes\left[\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{V}\right)\Theta\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{WH}+\varepsilon\right)\right], \tag{5}$$

where $\otimes$ is the element-wise multiplication, Θ is the element-wise division and ε is a small constant(typically 10^{-16}) for enforcing positive entries.

3.2 Clustering based on NMF

In the recent years, the use of NMF for clustering of non-negative data has already attracted much attention. Some examples can be found in Refs. [22–24]. DING, et al, shows the equivalence among NMF, spectral clustering and k-means clustering[22]. KIM and PARK explain the principium of NMF in cluster[25]. In k-means concept, the objective function to be minimized is the sum of squared distance from each data point to its centroid. With $\boldsymbol{A}=[a_1,a_2,\cdots,a_n]\in\mathbf{R}^{m\times n}$, the objective function $\boldsymbol{J}_k$ with given integer k can be written as below:

$$\boldsymbol{J}_k = \sum_{j=1}^{k}\sum_{\boldsymbol{a}_i\in\boldsymbol{c}_j}\left\|\boldsymbol{a}_i-\boldsymbol{c}_j\right\|^2 = \|\boldsymbol{A}-\boldsymbol{CB}\|^2, \tag{6}$$

where $\boldsymbol{C}=[\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_k]\in\mathbf{R}^{m\times k}$ is the centroid matrix, $\boldsymbol{c}_j$ is the cluster centroid of the j-th cluster, and $\boldsymbol{B}\in\mathbf{R}^{k\times n}$ denote clustering assignment, i.e., $\boldsymbol{B}_{ij}=1$ if i-th observation is assigned to the j-th cluster, and $\boldsymbol{B}_{ij}=0$ otherwise.

$$\boldsymbol{D}^{-1} = \mathrm{diag}\left(\frac{1}{|\boldsymbol{c}_1|},\frac{1}{|\boldsymbol{c}_2|},\cdots,\frac{1}{|\boldsymbol{c}_k|}\right)\in\mathbf{R}^{k\times k}, \tag{7}$$

where $|\boldsymbol{c}_j|$ is the number of data points in cluster j, $\boldsymbol{C}$ can be expressed as $\boldsymbol{C}=\boldsymbol{ABD}^{-1}$. Then

$$\boldsymbol{J}_k = \left\|\boldsymbol{A}-\boldsymbol{ABD}^{-1}\boldsymbol{B}^{\mathrm{T}}\right\|^2, \tag{8}$$

and the task of k-means is to look for $\boldsymbol{B}$ that minimize the objective function $\boldsymbol{J}_k$ where $\boldsymbol{B}$ has only one in each row, with others zero. Take two diagonal matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ to satisfy $\boldsymbol{D}^{-1}=\boldsymbol{D}_1\boldsymbol{D}_2$. Set $\boldsymbol{F}=\boldsymbol{BD}_1$ and $\boldsymbol{H}=\boldsymbol{BD}_2$, the above objective function can be replaced by

$$\min\nolimits_{\boldsymbol{F},\boldsymbol{H}} \boldsymbol{J}_k = \left\|\boldsymbol{A}-\boldsymbol{AFH}^{\mathrm{T}}\right\|^2. \tag{9}$$

This function is similar to Eq. (3) if set $\boldsymbol{W}=\boldsymbol{AF}$. That is to say, NMF somehow can work as a clustering approach. Actually, the basic idea of NMF for clustering is very simple. Further explanation is elaborated in Ref. [26]. Moreover, there is a brief overview of its probabilistic interpretation in Ref. [20].

Let us consider the joint probability $p\left(\boldsymbol{t}_i,\boldsymbol{d}_j\right)$ of feature $\boldsymbol{t}_i$ and sample data $\boldsymbol{d}_j$, it is factorized as

$$p\left(\boldsymbol{t}_i,\boldsymbol{d}_j\right) = \sum_k p\left(\boldsymbol{t}_i\middle|\boldsymbol{c}_k\right)p\left(\boldsymbol{d}_i\middle|\boldsymbol{c}_k\right)p\left(\boldsymbol{c}_k\right), \tag{10}$$

where $p\left(\boldsymbol{c}_k\right)$ is the prior probability for cluster $\boldsymbol{c}_k$. Elements of $\boldsymbol{V}$ can be seen as $p\left(\boldsymbol{t}_i,\boldsymbol{d}_j\right)$. Relating Eq. (10) to the factorization Eq. (2), $\boldsymbol{W}_{ik}$ corresponds to $p\left(\boldsymbol{t}_i\middle|\boldsymbol{c}_k\right)$, which means the significance of feature $\boldsymbol{t}_i$ in cluster $\boldsymbol{c}_k$. Employ sum-to-one normalization to each column of $\boldsymbol{W}$, i.e., $\boldsymbol{WD}_{\boldsymbol{W}}^{-1}$ where $\boldsymbol{D}_{\boldsymbol{W}}=\mathrm{diag}\left(\boldsymbol{I}^{\mathrm{T}}\boldsymbol{W}\right)$ and $\boldsymbol{I}=[1,1,\cdots,1]^{\mathrm{T}}$. Then the Eq. (2) can be rewritten as follows:

$$\boldsymbol{V} = \left(\boldsymbol{WD}_{\boldsymbol{W}}^{-1}\right)\left(\boldsymbol{HD}_{\boldsymbol{W}}\right)^{\mathrm{T}}. \tag{11}$$

Comparing Eq. (11) with Eq. (10), $\left(\boldsymbol{HD}_{\boldsymbol{W}}\right)^{\mathrm{T}}$ can be treated as $p\left(\boldsymbol{d}_i\middle|\boldsymbol{c}_k\right)p\left(\boldsymbol{c}_k\right)$. What is more, in the task of clustering, the posterior probability $p\left(\boldsymbol{c}_k\middle|\boldsymbol{d}_i\right)$ is necessary. Applying on Bayes rule, the posterior probability is given by

$$p\left(\boldsymbol{c}_k\middle|\boldsymbol{d}_j\right) \propto p\left(\boldsymbol{d}_i\middle|\boldsymbol{c}_k\right)p\left(\boldsymbol{c}_k\right) = \left[\boldsymbol{D}_{\boldsymbol{W}}\boldsymbol{H}^{\mathrm{T}}\right]_{kj}. \tag{12}$$

Therefore, data $\boldsymbol{d}_j$ is assigned to cluster k if

$$k \leftarrow \arg\max\nolimits_k [\boldsymbol{HD}_{\boldsymbol{W}}]_{jk}. \tag{13}$$

In general, the columns in $\boldsymbol{V}$ are treated as data points in an m-dimensional space, columns in $\boldsymbol{W}$ are considered as basis vectors and each row in $\boldsymbol{H}$ represents the extent of each basis vector that is used to reconstruct data vector. According to the distribution of $\boldsymbol{H}$, n samples are divided into k clusters, which are presented in Fig. 1.

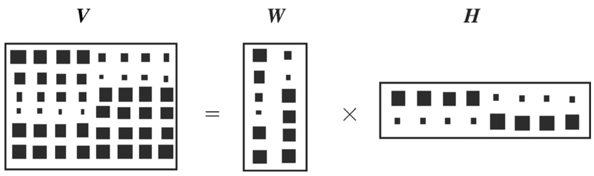


Fig. 1. Clustering with a matrix $V \in \mathbf{R}^{6\times8}$

In Fig. 1, an illustration of clustering with a matrix $V \in \mathbf{R}^{6\times8}$ is shown in the case of two clusters. The bigger square indicates bigger value in matrix. NMF produces two factors $\boldsymbol{W}$ and $\boldsymbol{H}$, where columns in $\boldsymbol{W}$ are correlated to prototypes and columns in $\boldsymbol{H}$ reflect cluster indicators.

4 Overview of the Fault Diagnosis Strategy

4.1 Fault diagnosis scheme based on NMF

The proposed fault diagnosis method mainly has two steps: training and testing. The flow chart of whole strategy is displayed in Fig. 2, and the implementation procedure is detailed as follows.

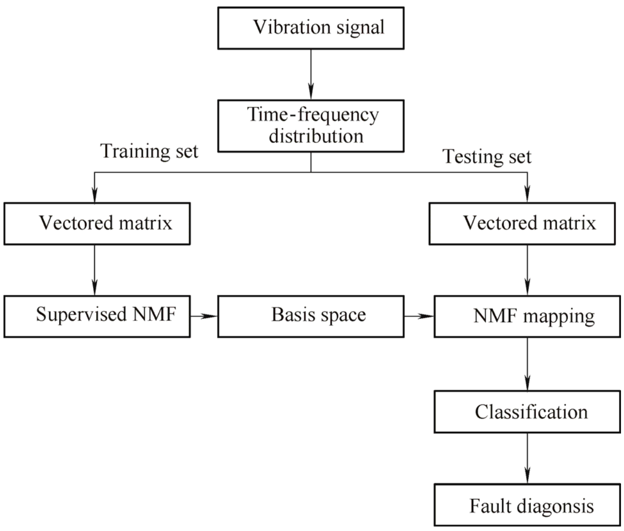


Fig. 2. Flowchart of the fault diagnosis system

(1) Transform vibration data into time-frequency distributions and get their spectrogram. In this stage, we choose STFT to realize the transformation because of less artificial influence. Then, randomly select several segments of data from each source and assemble them as training set by the vector of each spectrogram. Mostly, the dimension of a spectrogram is too big to take all of them as an input that dimension reduction is necessary.

(2) For the training, NMF is adopted to compress the dimension and obtain features respectively. As a new emerged technique, NMF has a lot of benefits in feature extraction. After compression, gather these matrices to form an over-complete basis in feature space. As a result, these bases can perfectly represent the underlying characteristics of the rest observed signals. It is believed that the parts of bases are different from different signals. That is to say, it is possible to separate various signals from a data set according to the characteristics of base vectors.

(3) In testing stage, the observed signals are mapped onto the assembled basis. The bases are kept fixed and the gains won't be updated during the iterative process in NMF way until they converge to the stationary point. After obtaining the corresponding gain matrix, we will take use of the clustering property of NMF to realize fault identification. After analyzing the basis, the biggest value in each column of the gain matrix indicates the fault that it belongs.

4.2 Feature extraction using supervised NMF

During training, a supervised NMF is applied to obtain basis space from the training data, and then use bases to recognize different faults. Supervised NMF helps us learn about the basis of each source separately from training data. In order to realize this process, training data is indispensable:

$$\boldsymbol{V}_{\text{train}}(t) = \{\boldsymbol{X}_1, \boldsymbol{X}_2 \cdots, \boldsymbol{X}_t\}\ (t = 1, 2, \cdots, T), \tag{14}$$

where $\boldsymbol{X}_t$ is the t-th amplitude spectrogram of fault's training samples whose dimension is $n \times m$. According to Eq. (2), this set can be decomposed as below:

$$\boldsymbol{V}_{\text{train}}(t) = \boldsymbol{W}_{n\times r}(t)\boldsymbol{H}_{r\times m}(t)\ (t = 1, 2, \cdots, T), \tag{15}$$

where $\boldsymbol{W}_{n\times r}(t)$ is the set of basis, and $\boldsymbol{H}_{r\times m}(t)$ is the corresponding gains of each basis. After achieving all the bases, combine them together for the over-complete basis set

$$\boldsymbol{W}_{\text{train}} = \{\boldsymbol{W}_{n\times r}(1), \boldsymbol{W}_{n\times r}(2), \cdots, \boldsymbol{W}_{n\times r}(T)\}, \tag{16}$$

where T is the type of fault.

4.3 Fault recognition

As mentioned in testing, with the basis space extracted by supervised NMF, the test data is used to be recognized. Assuming we have L transformed signals $\boldsymbol{X}_i$, assemble them as the test data

$$\boldsymbol{V}_{\text{test}} = \{\boldsymbol{X}_1, \boldsymbol{X}_2 \cdots, \boldsymbol{X}_t\}\ (t = 1, 2, \cdots, L). \tag{17}$$

Using Eqs. (3), (4) to iterate and keep the $\boldsymbol{W}$ fixed until it converges, we can achieve a resulting gains matrix $\boldsymbol{H}_{\text{test}}$, which satisfies

$$\boldsymbol{V}_{\text{test}} \approx \boldsymbol{W}_{\text{train}}\boldsymbol{H}_{\text{test}}. \tag{18}$$

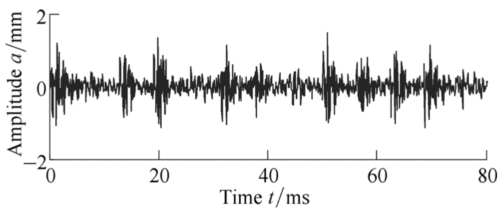
According to the contents in section 3.2, it is convenient to distinguish various types of faults by using Eq. (13). In the ideal condition, the L test signals can be correctly clustered into T training types.

5 Experiment and Discussion

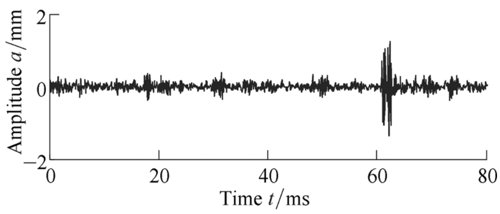
5.1 Bearing data set description

The vibration data of rolling element bearing is acquired from the bearing center of the Case Western Reserve University[27]. The type of bearings in this test was SKF 6205, deep groove ball bearing. Single point faults were introduced to the bearings with fault diameters of 0.177 8 mm, 0.355 6 mm and 0.533 4 mm. 9 kinds of bearings with various faults, i.e. inner race fault, ball fault and outer race fault with 3 diameters status. The load in the test was 1.5 kW and the motor speed was about 1257 r/min. Vibration signals were collected by accelerometers, which were attached to the driver end with magnetic bases. The sampling frequency is 12 kHz and the sample number is 120 000 points.

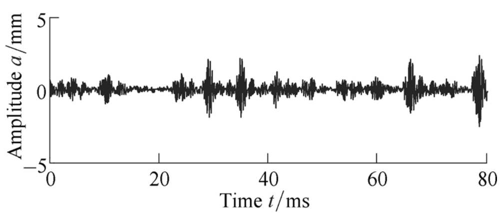
Bearing faults usually indicate as periodic transient impulses in vibration signals, so its sampling signals should cover at least two or three periods to show thus features in time-frequency domain. Therefore, take the sampling frequency and feature frequency into consideration, the selected 1024 sampling length can represent enough features. Total 180 samples containing 1024 points are randomly selected from every source. All the waveforms of various faults vibration signals with load 1.5 kW are shown in Fig. 3–Fig. 5.



(a) 0.177 8 mm

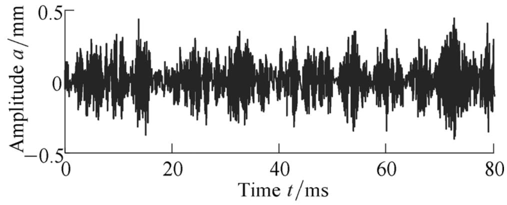


(b) 0.355 6 mm

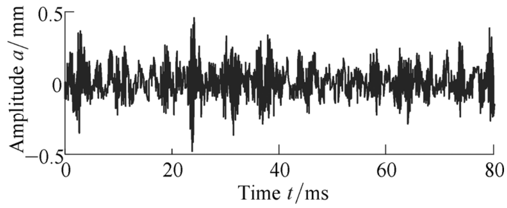


(c) 0.533 4 mm

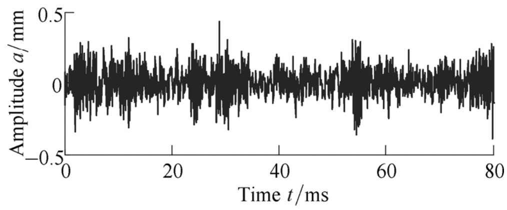
Fig. 3. Waveform of inner race fault with 3 diameters status



(a) 0.177 8 mm

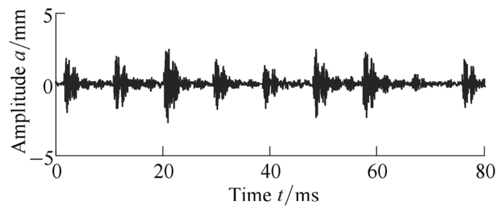


(b) 0.355 6 mm

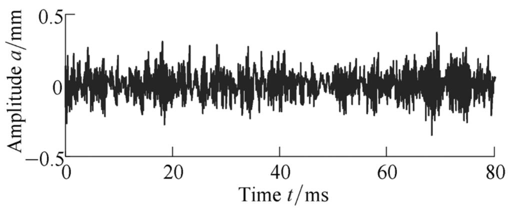


(c) 0.533 4 mm

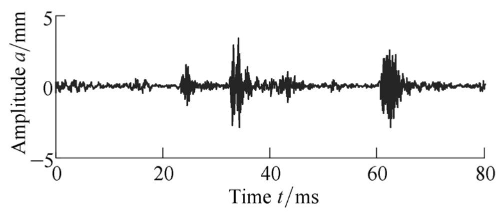
Fig. 4. Waveform of ball fault with 3 diameters status



(a) 0.177 8 mm



(b) 0.355 6 mm



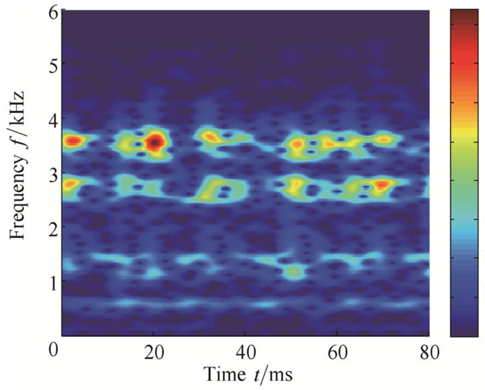
(c) 0.533 4 mm

Fig. 5. Waveform of outer race fault with 3 diameters status

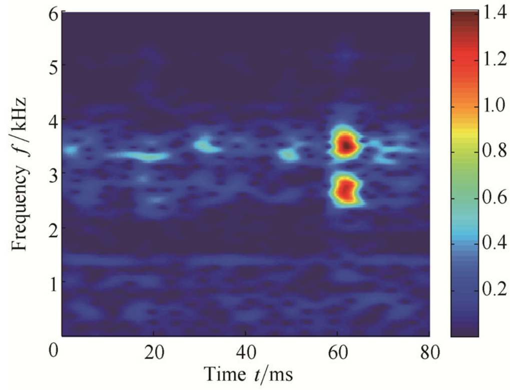
With the increasing of fault severity, the vibration signal amplitude also increases accordingly. However, under the practice operating conditions, influenced by various factors, the vibration amplitude does not strictly correspond to fault size. So, the amplitude of waveform in Fig. 5(b) is smaller than the amplitude of other fault diameters.

5.2 Feature extraction of TFD through NMF

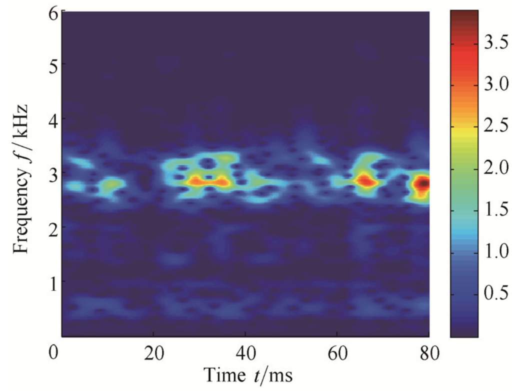
After obtaining TFD of every state, how to classify and recognize the vibration signals is a typical pattern recognition problem. For each sample of bearing, a 257×1024 time-frequency matrix can be obtained based on the STFT. However, it is impossible to take the whole matrix as input, because 257×1024 is such a huge amount for any pattern recognition system. Thus, it is necessary to reduce the data dimension to an acceptable scale, and the information is kept as much as possible. The original TFDs of various faults are shown in Fig. 6–Fig. 8.



(a) 0.177 8 mm



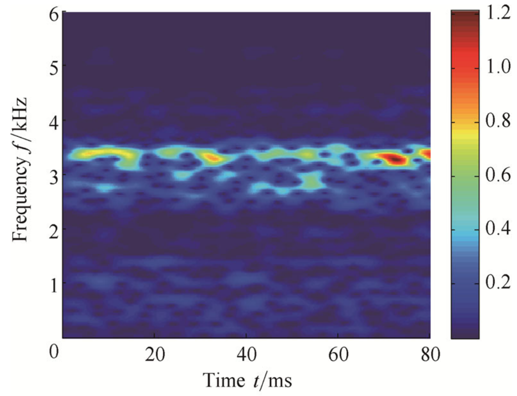
(b) 0.355 6 mm



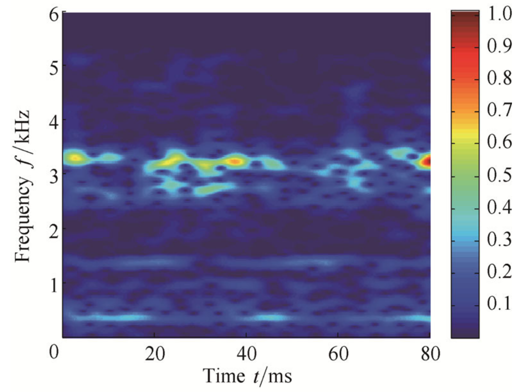
(c) 0.533 4 mm

Fig. 6. TFDs of inner race fault with different fault diameters

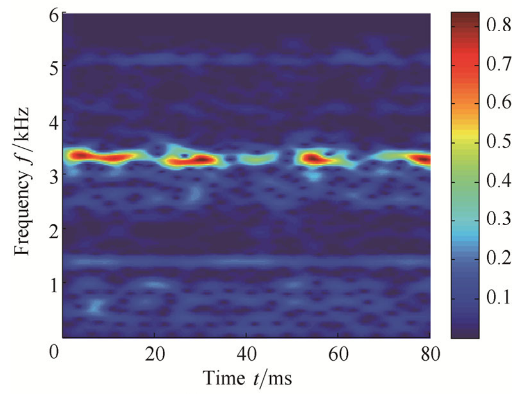
Being regionally represented for TFDs in Fig. 6–Fig. 8, also easily interpreted and understood directly, NMF is applied to compress the feature dimension of the original time-frequency matrix. Besides, before NMF is applied, these TFD matrices need to be normalized and vectored. Then randomly select 10 samples from each source, as the training set to achieve basis, and the rest samples are put together as the testing set.



(a) 0.177 8 mm



(b) 0.355 6 mm



(c) 0.533 4 mm

Fig. 7. TFDs of ball fault with different fault diameters

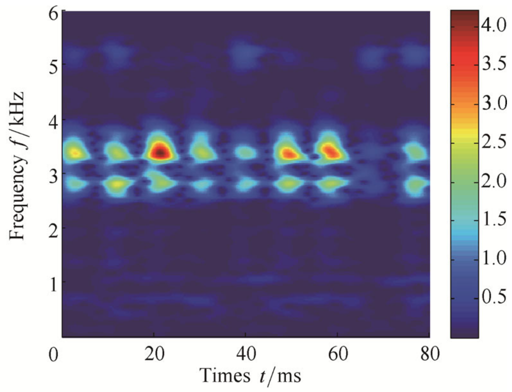
So, the training sets of 9 fault sources are transformed into 9 training matrix $\boldsymbol{V}(i)(i=1,2,\cdots,9)$. And the lower dimension r is confirmed by a heuristic method proposed by CICHOCKI[28], which was calculated to be 6 in our experiment. So, as mentioned above, the training set can be factorized as follows:

$$\boldsymbol{V}(i)_{263\ 168\times10}=\boldsymbol{W}(i)_{263\ 168\times10}\boldsymbol{H}(i)_{263\ 168\times10}(i=1,2,\cdots,9), \tag{19}$$

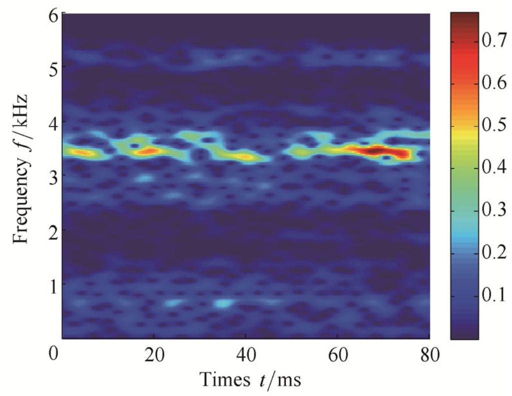
where $\boldsymbol{W}(i)$ is the basis matrix, which represents the features of each source. After factorization, combine 9 basis matrices together to form the base set $\boldsymbol{W}$ of the training samples

$$\boldsymbol{W}=[\boldsymbol{W}(1)_{263\,168\times 6},\boldsymbol{W}(2)_{263\,168\times 6},\cdots,\boldsymbol{W}(9)_{263\,168\times 6}], \tag{20}$$

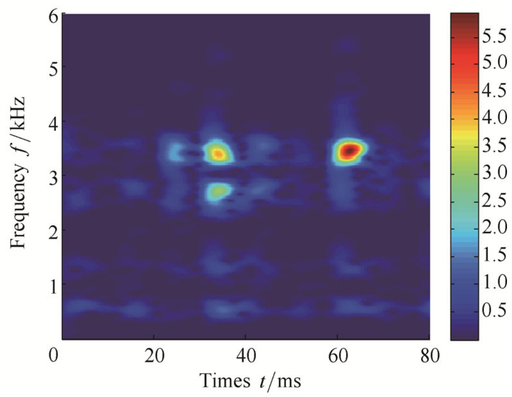
where $\boldsymbol{W}$ is $263\,168\times 54$. That is to say, all the characteristics of each source are contained in the basis set $\boldsymbol{W}$. Next, as described in Eq. (18), the $\boldsymbol{H}_{\text{test}}$ can be achieved.



(a) 0.177 8 mm



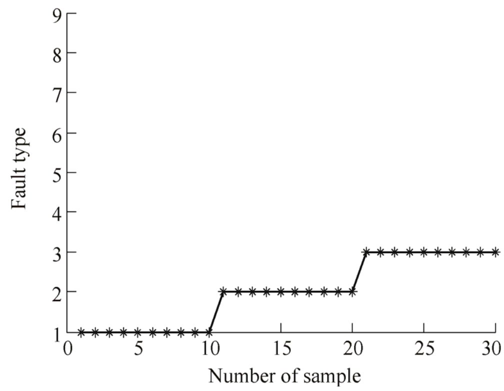
(b) 0.355 6 mm



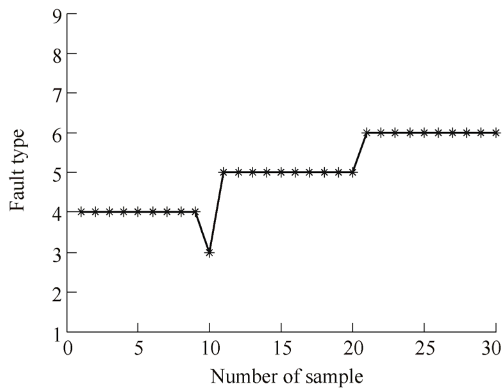
(c) 0.533 4 mm

Fig. 8. TFDs of outer race fault with different fault diameters

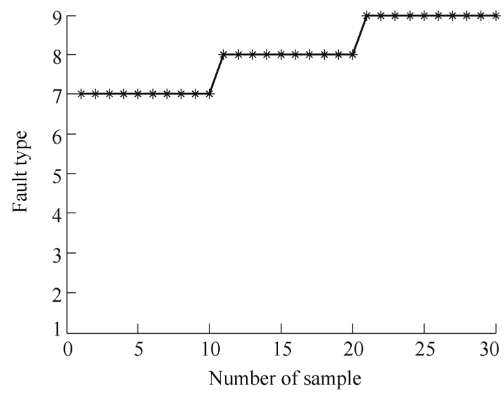
In the experiment, each of faults with different fault diameters is selected to be recognized, where fault diameter is 0.177 8 mm, 0.355 6 mm and 0.533 4 mm, respectively. The purpose is to validate the recognition performance for different fault severities. Based on the clustering property of NMF and Eq. (13), the position of the biggest value of one column would determine which cluster the sample belongs to, and the recognition result of inner race fault, ball fault and outer race fault is shown in Fig. 9.



(a) Inner race fault with 3 diameters status



(b) Ball fault with 3 diameters status



(c) Outer race fault with 3 diameters status

Fig. 9. Classification of one type with 3 diameters status

As seen from Fig. 9, y-label represents the clustered fault type. The numbers 1–9 represents inner-0.177 8 mm, inner-0.355 6 mm, inner-0.533 4 mm, ball-0.177 8 mm, ball-0.355 6 mm, ball-0.533 4 mm, outer-0.177 8 mm, outer-0.355 6 mm and outer-0.533 4 mm respectively. It is known that the difference between different fault diameters of the same sort is quite little, so the recognition performance of one sort fault is a cardinal criterion for a technique. In addition, according to the Fig. 9, especially in Fig. 9(b), a ball fault sample is wrongly recognized as inner race fault. And the main reason is that some TFDs of ball fault are really similar to inner race fault's.

Although there is a mistake during the recognition, the diagnosis result is satisfactory. In order to further examine the performance of proposed method, the inner race fault, ball fault, and outer race fault with fault diameter of 0.355 6 mm are chosen. The recognition result is displayed in the Fig. 10.

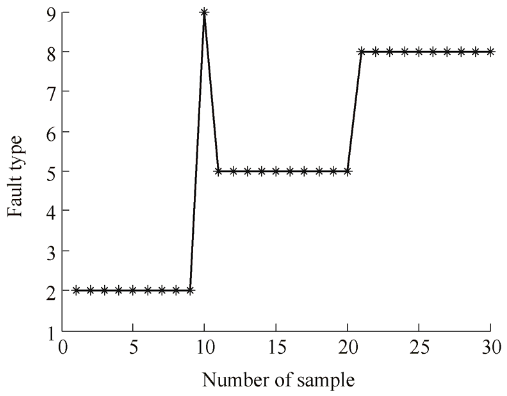


Fig. 10. Classification of 3 types with the same fault diameter

Comparing Fig. 9 with Fig. 10, we can come up with the conclusion that our strategy is capable to deal with signals of variable types or various fault severities.

Eventually, sum all above, and totally 9 distinct sorts of signals will be introduced in the following test, where different types and fault diameters are involved. The classification performance of 9 kinds of fault is shown in Fig. 11.

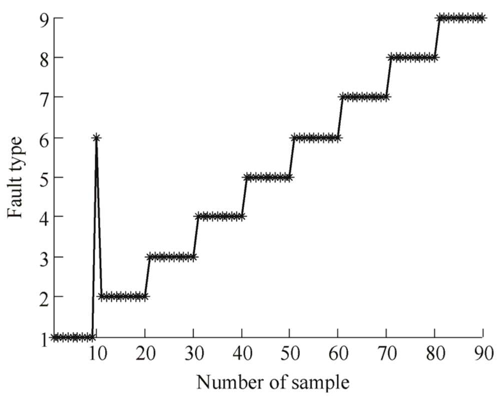


Fig. 11. Classification results of all types

As seen from Fig. 10 and Fig. 11, there are misjudgments of sample. The waveform of vibration signal has a certain similarity between different fault type and fault severities, thus the similar features are extracted form TFDs, which achieve some misjudgment. Particularly, as shown in Fig. 11, the total 90 samples are clustered into 9 classes. Although among them, a sample is mistakenly classified to another fault type, the whole performance of the test is impressive. As a consequence, we can conclude that the strategy based on NMF is available for bearing fault clustering and diagnosis.

For the sake of a robust result, the experiment is implemented for 10 times to get a more stable evaluation. In order to improve recognition performance, denoising can be conducted by combining signal transform with thresholding operation to remove the noise. On the other hand, typical fault sample is added to basis space to more accurately describe the clustering boundary.

5.3 Recognition based on artificial neural network

For comparison, ANN toolbox of Matlab is adopted in this experiment. The ANN method used is the neural network pattern recognition toolbox. It's a two-layer feed-forward network, with sigmoid hidden and output neurons, given neurons in its hidden layer. The network will be trained with scaled conjugate gradient backpropagation, where weight change parameter $\sigma=0.000\,5$ and maximum number of epochs is set to 100.

Because this toolbox is not able to cope with high dimension data, statistical parameters from the vibration signals are evaluated in advance. As a consequence, some extra criterions are needed and the calculation becomes more complicated. In this experiment, we choose 7 parameters of time domain (standard derivation, kurtosis, root mean square, mean, maximum value, shape factor and inlay) and 8 energy indices of diverse frequency bands, 15 statistical parameters in total. During the classification, randomly select 70% of the signals as training set and the rest parts are used for test data. For comparison, ANN is also repeated for 10 times. The classification accuracy of ANN and NMF is listed in Table 1.

Table 1. Effect of two methods

Statistical indices	Accuracy of ANN-based classification α_1/%	Accuracy of NMF-based classification α_2/%
Maximum	100	100
Minimum	22.2	97.78
Mean	75.8	99.3

According to the classification results, NMF not only has a relatively stable performance, but also accomplishes better mean accuracy, such as a more stable recognition and higher accuracy. However, In the 10 repeated tests, there is a large difference between maximum and minimum accuracy of ANN, which implicates that the corresponding results are not throughout creditable. Meanwhile, NMF obtains satisfactory result on both the accuracy of each test and the whole recognition performance. Although, both of them could reach maximum accuracy α=100%, NMF yields 99.3% mean accuracy which is much superior to ANN. Obviously, the result proves that it is more appealing to make use of NMF-based classification instead of other methods. Therefore, it is credible that the NMF can offer a good resolution in fault diagnosis problem of bearing.

6 Conclusions

(1) According to characteristics of vibration signal for rolling element bearing, a feature extraction and recognition method based on STFT and NMF is proposed. Experiments demonstrate that the drawbacks of FFT analysis for non-stationary signal feature representation can be solved by TFD with STFT, where the noise and impulse component can be separated effectively.

(2) Considering the high dimensional feature space, the supervised NMF mapping is adopted to select local features from TFD. Meanwhile, with the clustering property of NMF, fault samples can be recognized automatically. Therefore, the fault recognition capabilities can be improved obviously.

(3) The application of rolling element bearing faults shows that, the algorithm successfully discover the low-dimensional feature spaces and reveal interest and the separability of sample patterns. Besides, comparing with ANN, the feature extraction and recognition capability of proposed method is superior to that of ANN.

References

[1] CHEN Xiaoguang, LIANG Lin, XU Guanghua, et al. Feature extraction of kernel regress reconstruction for fault diagnosis based on self-organizing manifold learning[J]. *Chinese Journal of Mechanical Engineering*, 2013, 26(5): 1041–1049.

[2] SMAMANTA B, NATARAJ C. Application of particle swarm optimization and proximal support vector machines for fault detection[J]. *Swarm Intelligence*, 2009, 3(4): 303–325.

[3] LEI Yaguo, HE Zhengjia, ZI Yanyang, et al. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs[J]. *Mechanical Systems and Signal Processing*, 2007, 21(5): 2280–2294.

[4] SUGUMARAN V, RAMACHANDRAN K I. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing[J]. *Mechanical Systems and Signal Processing*, 2007, 21(5): 2237–2247.

[5] HE Qingbo, WANG Xiangxiang, ZHOU Qiang. Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis[J]. *Sensors*, 2014, 14(1): 382–402.

[6] LU Feng, FENG Fuzhou. Wavelet transform technology in the non-stationary signal fault diagnosis in engineering application[J]. *Advanced Materials Research*, 2012, 518(1): 1355–1358.

[7] WANG Huaqing, CHEN Peng. Fuzzy diagnosis method for rotating machinery in variable rotating speed[J]. *Sensors Journa*l, 2011, 11(1): 23–34.

[8] ZHOU Changjun, WANG Lan, ZHANG Qiang, et al. Face recognition based on PCA image reconstruction and LDA[J]. *Optik*, 2013, 124(22): 5599–5603.

[9] DU Xianfeng, LI Zhijun, BI Fengrong, et al. Source separation of diesel engine vibration based on the empirical mode decomposition and independent component analysis[J]. *Chinese Journal of Mechanical Engineering*, 2012, 25(3): 557–563.

[10] LIU Hongxing, LI Jian, ZHAO Ying, et al. Improved singular value decomposition technique for detection and extraction periodic impulse component in a vibration signal[J]. *Chinese Journal of Mechanical Engineering*, 2004, 17(3): 340–345.

[11] LI Weihua, SHI Tielin, LIAO Guanglan, et al. Feature extraction and classification of gear faults using principal component analysis[J]. *Journal of Quality in Maintenance Engineering*, 2003, 9(2): 132–143.

[12] LIANG Xingyu, WANG Yuesen, SHU Genqun, et al. Identification of axial vibration excitation source in vehicle engine crankshafts using an auto-regressive and moving average model[J]. *Chinese Journal of Mechanical Engineering*, 2011, 24(6): 1022–1027.

[13] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755): 788–791.

[14] LIU Weixiang, ZHENG Nanning. Non-negative matrix factorization based methods for object recognition[J]. *Pattern Recognition Letters*, 2004, 25(8): 893–897.

[15] PU Xiaorong, ZHANG Yi, ZHENG Zinming, et al. Face recognition using fisher non-negative matrix factorization with sparseness constraints[J]. *Lecture Notes in Computer Science*, 2005, 3497: 112–117.

[16] LIU Haifeng, WU Zhaohui, CAI Deng, et al. Constrained non-negative matrix factorization for image representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1299–1311.

[17] MEHMOOD A, DAMARLA T, SABATIER J. Separation of human and animal seismic signatures using non-negative matrix factorization[J]. *Pattern Recognition Letters*, 2012, 33(16): 2085–2093.

[18] LI Bing, ZHANG Peilin, LIU Dongsheng, et al. Feature extraction for rolling element bearing fault diagnosis utilizing generalized S transform and two-dimensional non-negative matrix factorization[J]. *Journal of Sound and Vibration*, 2011, 330(10): 2388–2399.

[19] WANG Qinghua, ZHANG Youyun, CAI Lei, et al. Fault diagnosis for diesel valve trains based on non-negative matrix factorization and neural network ensemble[J]. *Mechanical Systems and Signal Processing*, 2009, 23(5): 1683–1695.

[20] YOO J, CHOI S. Orthogonal non-negative matrix tri-factorization for co-clustering: multiplicative updates on siefel manifolds[J]. *Information Processing and Management*, 2010, 46(5): 559–570.

[21] PAATERO P, TAPPER U. Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values[J]. *Environmetrics*, 1994, 5(2): 111–126.

[22] DING C, LI Tao, PENG Wei, et al. Orthogonal nonnegative matrix tri-factorizations for clustering[C]//*Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, USA, August 20–23, 2006: 126–135.

[23] LI Tao, DING C. The relationships among various nonnegative matrix factorization methods for clustering[C]//*Proceedings of the 6th International Conference on Data Mining*, Hong Kong, China, December 18–22, 2006: 362–371.

[24] OKUN O G. Non-negative matrix factorization and classifiers: experimental study[C]//*Proceedings of the 4th IASTED International Conference on Visualization, Imaging, and Image Processing*, Marbella, Spain, September 6–8, 2004: 550–555.

[25] KIM J, PARK H. *Sparse Nonnegative Matrix Factorization for Clustering*[R]. Georgia, Atlanta, Georgia Institute of Technology, 2008.

[26] DING C, LI Tao, PENG Wei. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing[J]. *Computational Statistics & Data Analysis*, 2008, 52(8): 3913–3927.

[27] LIU Haining, LIU Chengliang, HUANG Yixiang. Adaptive feature extraction using sparse coding for machinery fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2011, 25(2): 558–574.

[28] CICHOCKI A, ZDUNEK R, PHAN A H, et al. *Nonnegative matrix and tensor factorizations*[M]. West Sussex: John Wiley & Sons Inc, 2009.

Biographical notes

GAO Huizhong, born in 1989, is currently an associate engineer at *The 705 Research Institute, China Shipbuilding Industry*

Corporation, *China*. He received his master degree from *Xi'an Jiaotong University*, *China*, in 2014. His research interests concentrate on intelligent signal processing.
E-mail: stevein8@163.com

LIANG Lin, born in 1973, is currently working at *Xi'an Jiaotong University, China*. He received his PhD degree from *Xi'an Jiaotong University, China*, in 2007. His research interests include fault diagnosis and measurement.
Tel: +86-29-82663707; E-mail: lianglin@mail.xjtu.edu.cn

CHEN Xiaoguang, born in 1988, is currently a PhD candidate at *School of Mechanical Engineering, Xi'an Jiaotong University, China*. He received his bachelor degree from *Xi'an Jiaotong University, China*, in 2010. His research interests include signal process and intelligent fault diagnosis.
Tel: +86-29-83395054; E-mail: chenxiaoguang2008@163.com

XU Guanghua, born in 1964, is currently professor at *Xi'an Jiaotong University, China*. He received his PhD degree from *Xi'an Jiaotong University, China*, in 1995. His research interests include fault diagnosis and brain-computer interface.
Tel: +86-29-82663707; E-mail: xugh@mail.xjtu.edu.cn