

Research Article

Casing Vibration Fault Diagnosis Based on Variational Mode Decomposition, Local Linear Embedding, and Support Vector Machine

Yizhou Yang and Dongxiang Jiang

State Key Lab of Control and Simulation of Power Systems and Generation Equipment, Department of Thermal Engineering, Tsinghua University, Beijing 100084, China

Correspondence should be addressed to Yizhou Yang; yangyz16@mails.tsinghua.edu.cn

Received 10 October 2016; Revised 17 January 2017; Accepted 16 February 2017; Published 12 March 2017

Academic Editor: Dario Di Maio

Copyright © 2017 Yizhou Yang and Dongxiang Jiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To diagnose mechanical faults of rotor-bearing-casing system by analyzing its casing vibration signal, this paper proposes a training procedure of a fault classifier based on variational mode decomposition (VMD), local linear embedding (LLE), and support vector machine (SVM). VMD is used first to decompose the casing signal into several modes, which are subsignals usually modulated by fault frequencies. Vibrational features are extracted from both VMD subsignals and the original one. LLE is employed here to reduce the dimensionality of these extracted features and make the samples more separable. Then low-dimensional data sets are used to train the multiclass SVM whose accuracy is tested by classifying the test samples. When the parameters of LLE and SVM are well optimized, this proposed method performs well on experimental data, showing its capacity of diagnosing casing vibration faults.

1. Introduction

The fault diagnosis of gas turbine engine is a large challenge to nowadays' mechanical engineers due to the complexity of its interior structure. Displacement sensors sometimes cannot be installed inside the machine. Vibration signals will be collected by the accelerometers installed on the casing. But it is difficult to analyze casing signals because of the complex surrounding structure and the long path of transmission from the inside to the outside [1]. Frequency components of casing acceleration signals could be various and complex. Hence, the feature extraction and fault classification of its vibration faults are not the same with those designed for traditional rotating machineries.

To deal with the complex vibration signal of the rotor-bearing-casing system, signal decomposition can be used to extract time or frequency characteristics before further processing. In addition to traditional filters of IIR and FIR, the wavelet transform is another widely used technique that could decompose and reconstruct signals. Wavelet transform

can extract the time-frequency characteristics of the signal with variable resolution [2], but the selection of wavelets and their parameters is hard to decide [3]. Empirical mode decomposition (EMD) is an adaptive technique which has a high-frequency resolution [4] but still has disadvantages like end effect and modes mixing. Since 1998, the application of EMD has been well developed. Many other adaptive decomposition methods were put forward, like local mean decomposition (LMD) [5], intrinsic time-scale decomposition (ITD) [6], and variational mode decomposition (VMD) [7]. This paper is employing the VMD to do signal decomposition.

Features of signals reflect the vibrational condition of the rotor-bearing-casing system. But too many features would cause multiple dimensional difficulty during analysis. Dimensionality reduction is the way to handle multidimensional data, which could reduce the complexity of calculation [8]. Many nonlinear dimensionality reduction techniques have been proposed these years, which have advantages over the traditional linear techniques like principal components

analysis (PCA) [9]. Local linear embedding (LLE) is a nonlinear technique proposed in 2000 [10] that has been widely used in many data processing problems. LLE is employed here to do dimensionality reduction, for its high efficiency in providing local properties of data.

The neural network has shown a strong capacity of solving nonlinear problems [11]. But it still has some disadvantages like existence of many local minima solutions and the difficulty in choosing the number of hidden units. Support vector machine (SVM) is a breakthrough of neural network based on the development of statistical learning theory and structural risk minimization [12]. The original SVM algorithm was invented in 1963. And in 1992, it was improved to be a nonlinear classifier by applying the kernel trick to maximum-margin hyperplanes [13].

In this paper, faults of rotor-bearing-casing system are recognized by analyzing the casing vibration signal. Vibrational features in time and frequency domain, along with these extracted by VMD method, are used as the database of fault classification. The experimental data are obtained from the rotor-bearing-casing test system, where several kinds of mechanical faults could be simulated. Since the vectors consisting of features are multidimensional and complicated, they are dimensionally reduced through LLE to get a low-dimensional data set. Then the low-dimensional data set plays as the training data of multiclass SVM. The parameters of LLE and SVM are optimized, so that a capable classifier for faults of rotor-bearing-casing system could be produced.

2. Theoretical Background

2.1. Variational Mode Decomposition. VMD defines Intrinsic Mode Function (IMF) as an amplitude-modulated-frequency-modulated signal:

$$u_k(t) = A_k(t) \cos(\phi_k(t)), \quad (1)$$

where $A_k(t)$ is the instantaneous amplitude, $\phi_k(t)$ is the instantaneous phase, and the instantaneous frequency is $\omega_k = \phi'_k(t)$.

Three important signal processing tools are involved in the building of the VMD model: Wiener filtering, Hilbert transform, and frequency mixing.

The goal of VMD is to decompose a real valued input signal into K subsignals (modes) written as u_k , and each of them is mostly compact around a center pulsation ω_k . The constrained variational model VMD built [7] can be described as

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_k u_k(t) = f. \end{aligned} \quad (2)$$

To construct the model, first compute the associated analytic signal for each mode u_k by means of the Hilbert transform to obtain a unilateral frequency spectrum. Multiplied by $e^{-j\omega_k t}$, then shift the mode's frequency spectrum to

baseband. At last the bandwidth is estimated by means of the squared L^2 -norm of the gradient.

The augmented Lagrangian L is introduced to convert the constrained variational problem into an unconstrained variational problem:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \alpha \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 \\ &+ \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle. \end{aligned} \quad (3)$$

The solution to the original minimization problem now becomes finding the saddle point of the augmented Lagrangian L in a sequence of iterative subsignals and center pulsations. The algorithm is called alternate direction method of multipliers (ADMM) and the complete algorithm of VMD is shown as follows [14]:

- (1) Initialize $\{\hat{u}_k^1\}$, $\{\omega_k^1\}$, $\hat{\lambda}^1$, and n as 0.
- (2) $n = n + 1$; start iteration.
- (3) For $k = 1 : K$, update \hat{u}_k for all $\omega > 0$:

$$\begin{aligned} \hat{u}_k^{n+1}(\omega) &= \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i < k} \hat{u}_i^n(\omega) + \hat{\lambda}^n(\omega) / 2}{1 + 2\alpha(\omega - \omega_k^n)^2}. \end{aligned} \quad (4)$$

$\hat{u}_k^{n+1}(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}^n(\omega)$ are the Fourier transform of u_k^{n+1} , $f(t)$, and $\lambda^n(t)$.

- (4) For $k = 1 : K$, update ω_k

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega}. \quad (5)$$

- (5) Update $\hat{\lambda}$ for all $\omega > 0$:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right). \quad (6)$$

- (6) Repeat step (2) to (5) until convergence:

$$\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon. \quad (7)$$

The IMF subsignals u_k are the inverse Fourier transform of \hat{u}_k .

2.2. Locally Linear Embedding. Local linear embedding (LLE) is a nonlinear dimensionality reduction method widely used these years. It constructs a neighborhood graph representation of the data points. In LLE, the local properties of the data manifold are constructed by turning the data points into a linear combination of their nearest neighbors. And by this way the global nonlinear structure of the data is recovered. The procedure of LLE could be summarized as follows [15].

Suppose the data $X = (x_1, x_2, \dots, x_N)$, $x_i \in R^D$, $i = 1, 2, \dots, N$. For arbitrary point x_i , calculate the Euclidean distance between x_i and other points and find the k nearest points.

$$d(x_i, x_j) = \|x_i - x_j\|. \quad (8)$$

w_{ij} are linear coefficients that reconstruct each data point from its neighbor, as the weight of the j th data point to the i th reconstruction. The aim is to choose the weight to minimize the reconstruction error:

$$\varepsilon(w) = \min \left\| x_i - \sum_{j=1}^k w_{ij} x_j \right\|^2 \quad (9)$$

with the constraint

$$\sum_{j=1}^k w_{ij} = 1. \quad (10)$$

Zero filling ($w_{i1}, w_{i2}, \dots, w_{ik}$) to n -dimensional vector w_i , so all the weights combine as $n \times n$ matrix W . Then the reconstruction error can be written as

$$\varepsilon(w_i) = \min \|x_i - XW_i^T\|^2. \quad (11)$$

Suppose $M = (I - W^T)^T(I - W^T)$, where I is $n \times n$ identity matrix. Then there are

$$\begin{aligned} \varepsilon(W) &= \min \left\{ \sum_{i=1}^N \|x_i - XW_i^T\|^2 \right\} \\ &= \min \{ \text{tr}(XMX^T) \}, \end{aligned} \quad (12)$$

where tr is the trace of matrix.

To map the high-dimensional data X into low-dimensional data $Y = (y_1, y_2, \dots, y_N) \in R^{d \times N}$ representing global internal coordinates on the manifold, Y is selected to minimize reconstruction cost:

$$\begin{aligned} \varepsilon(Y) &= \min \left\{ \sum_{i=1}^N \|y_i - YW_i^T\|^2 \right\} \\ &= \min \{ \text{tr}(YMY^T) \}, \end{aligned} \quad (13)$$

where $\sum_{i=1}^N y_i = 0$ and $YY^T = I_{d \times d}$. Write Y as $Y = (z_1, z_2, \dots, z_d)^T$ and then

$$\begin{aligned} \text{tr}(YMY^T) &= \text{tr} \left(\sum_{i=1}^d z_i^T M z_i \right) = \sum_{i=1}^d \text{tr}(z_i^T M z_i) \\ &= \sum_{i=1}^d z_i^T M z_i. \end{aligned} \quad (14)$$

Because $YY^T = I_{d \times d}$, z_i is orthonormal. To minimize $\text{tr}(YMY^T)$, the bottom d nonzero eigenvalues ($\lambda_1, \lambda_2, \dots, \lambda_d$) and their corresponding eigenvectors (z_1, z_2, \dots, z_d) are chosen. So the low-dimensional data Y can be obtained as $Y = (z_1, z_2, \dots, z_d)^T$.

2.3. Support Vector Machine. The main purpose of support vector machine is to construct a hyperplane in space for the use of classification, regression, or other tasks. When used for classification, the largest distance between data points and the hyperplane means a good separation. If the samples are not linearly separated in the original space, it is proposed to map this space into a much higher-dimensional space, even an infinite-dimensional one, which may make it an easier separation [16].

Suppose the training data points $\{x_k, y_k\}_{k=1}^N$, where $x_k \in R^n$ is the input data and $y_k \in \{-1, 1\}$ is the binary class labels. Define nonlinear function $\varphi(\cdot) : R^n \rightarrow R^m$ which maps the original space into the higher-dimensional feature space. When the hyperplane is $w^T x + b = 0$ and w is its normal vector, the training data may be described as

$$\begin{aligned} w^T \varphi(x_k) + b &\geq 1, \quad \text{if } y_k = 1 \\ w^T \varphi(x_k) + b &\leq -1, \quad \text{if } y_k = -1. \end{aligned} \quad (15)$$

Thus the classifier will be the form of

$$y(x) = \text{sign}[w^T \varphi(x) + b]. \quad (16)$$

The problem of building a hyperplane then turns into an optimization problem

$$\min \left\{ \frac{1}{2} w^T w + c \sum_{k=1}^N \xi_k \right\}. \quad (17)$$

with the constraint condition

$$\begin{aligned} y_k [w^T \varphi(x_k) + b] &\geq 1 - \xi_k, \quad k = 1, \dots, N \\ \xi_k &\geq 0, \quad k = 1, \dots, N, \end{aligned} \quad (18)$$

where ξ_k are slack variables as the distance from the misclassified samples to the hyperplane, acting like the punishment. Positive constant c is the penalty coefficient and the bigger c is, the more the penalty of misclassification there will be.

It is a convex optimization problem and the Kuhn-Tucker theorem will be used to translate the problem into a dual problem

$$\max J(\alpha) = \max \left\{ \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k \alpha_l y_k y_l K(x_k, x_l) \right\} \quad (19)$$

with the constraint condition

$$\sum_{k=1}^N \alpha_k y_k = 0, \quad 0 \leq \alpha_k \leq c, \quad k = 1, \dots, N. \quad (20)$$

$\varphi(x_k)$ is not calculated in an explicit way but in the form of Mercer kernel

$$K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l). \quad (21)$$

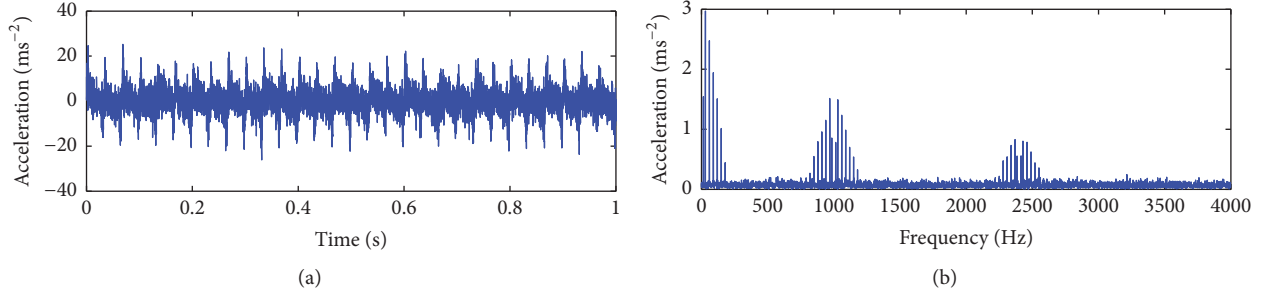


FIGURE 1: Time waveform (a) and frequency spectrum (b) of the simulation signal.

There are several kernels often used:

- (i) Linear: $K(x, x_k) = x_k^T x$.
- (ii) Polynomial: $K(x, x_k) = (x_k^T x + 1)^d$.
- (iii) RBF: $K(x, x_k) = \exp\{-\|x - x_k\|_2^2 / \sigma^2\}$.
- (iv) MLP: $K(x, x_k) = \tanh(\kappa x_k^T x + \theta)$.

So the nonlinear SVM classifier in dual space is turned into

$$y(x) = \text{sign} \left[\sum_{k=1}^N \alpha_k y_k K(x, x_k) + b \right]. \quad (22)$$

Here is the decision function of the optimized hyper-plane. The value without the “sign” function as the indicator may also be used in multiclass classification problems.

3. Simulation Signal Analysis Using VMD

In a rotor-bearing-casing structure, resonances are often excited by the harmonic frequencies of the shaft and other forms of vibration. Large amount of information including fault features is buried in the resonances. High-Frequency Resonance Technique (HFRT) is a normally used method to extract vibrational parameters from modulated signals [17]. But HFRT selects resonant frequency bands in empirical ways. It is unpractical to select frequency band for each sample when samples are too many. VMD helps to decompose the signal into a fixed number of intrinsic modes automatically.

Here is an example of a simulation signal, showing the effect of VMD decomposing modes apart. Suppose the shaft vibration x_s consists of rotational frequency and its sub-harmonic and harmonic frequencies. Two resonances of the casing are excited by the shaft vibration. Center frequencies of the resonances are set to 1000 Hz and 2400 Hz. So the casing signal x can be expressed as below, where $\text{wgn}()$ is white Gaussian noise.

$$x_s = \sum_{i=1}^n \text{amp}(i) \times \sin(2\pi \times \text{freq}(i) \times 30t); \quad (23)$$

here $\text{amp} = [1.5, 3, 2.5, 2, 1.5, 1, 0.5]$, $\text{freq} = [0.5, 1, 2, 3, 4, 5, 6]$, and

$$\begin{aligned} x &= x_s \\ &\times [1 + \sin(2\pi \times 1000t) + 0.6 \times \sin(2\pi \times 2400t)] \\ &+ \text{wgn}() \end{aligned} \quad (24)$$

Figure 1 is the time waveform of x and its frequency spectrum where components of shaft vibration and structural resonances are clear to see. Thus there are 3 so-called modes in the signal. The purpose of VMD here is to separate these 3 modes apart. Figure 2 is the output IMF subsignals and their frequency spectrums, showing that main frequency components remained in the 3 IMFs and noise not close to these 3 modes is filtered out. For comparison, the signal is also decomposed by empirical mode decomposition (EMD). Figure 3 is time waveforms of the EMD subsignals (IMFs) and frequency spectrums of the first 7 IMFs. The stop criterion is set to recommend value by Rilling et al. [18], and 13 modes are decomposed out, most of which are not needed. Other stop criterions are tried but the mode number will not be decreased to 3. Figure 4 shows the center frequencies changing during VMD iteration. After convergence, these 3 center frequencies are 0.0076, 0.1248, and 0.3002 relative to the sampling frequency of 8000 Hz (60.8 Hz, 998.4 Hz, and 2401.6 Hz), very close to the center frequencies set before the simulation. If there is not white Gaussian noise in the simulated signal, the center frequencies from VMD will be 56 Hz, 1000 Hz, and 2400 Hz. This example can be a proof of VMD's remarkable ability to deal with casing vibration signals.

4. Experimental System and Data

The rotor-bearing-casing coupled vibration experiment system is designed to simulate vibration faults and to study transmission characteristics of rotor-bearing-casing system in a gas turbine engine. Figure 5 shows the structural drawing and the photo of the test rig. Although the structure of the test section is relatively simple, it has a complete transmission path from rotor to casing. Therefore, the test rig cannot fully represent a gas turbine engine but has many similar

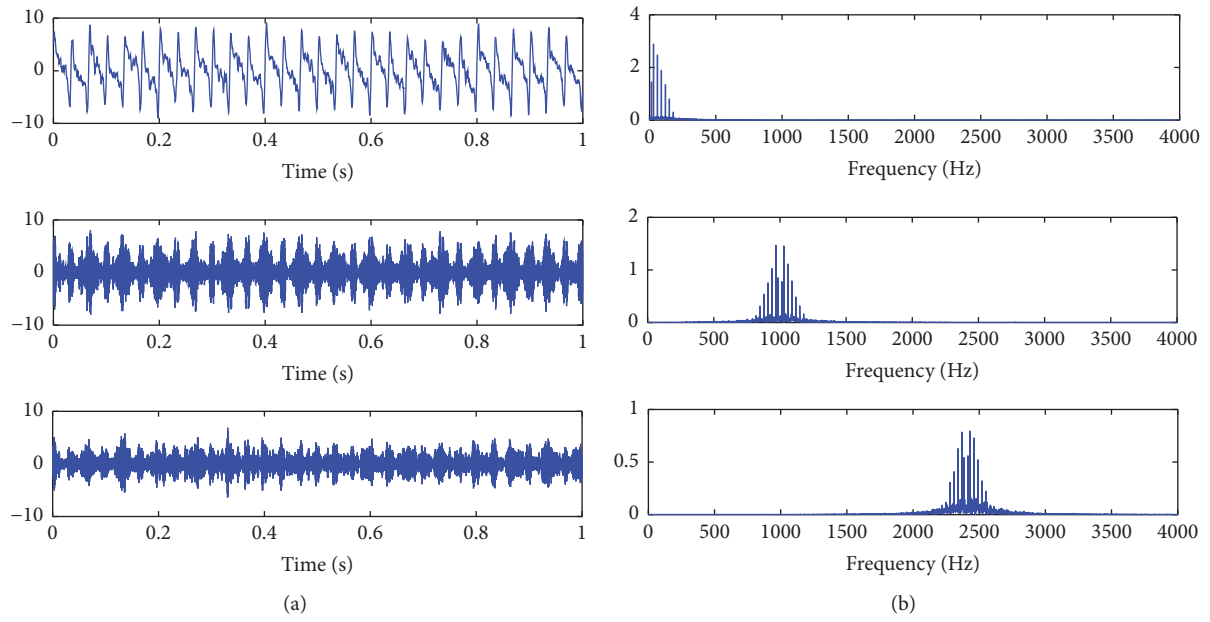


FIGURE 2: VMD subsignals (a) and their frequency spectrums (b) of the simulation signal.

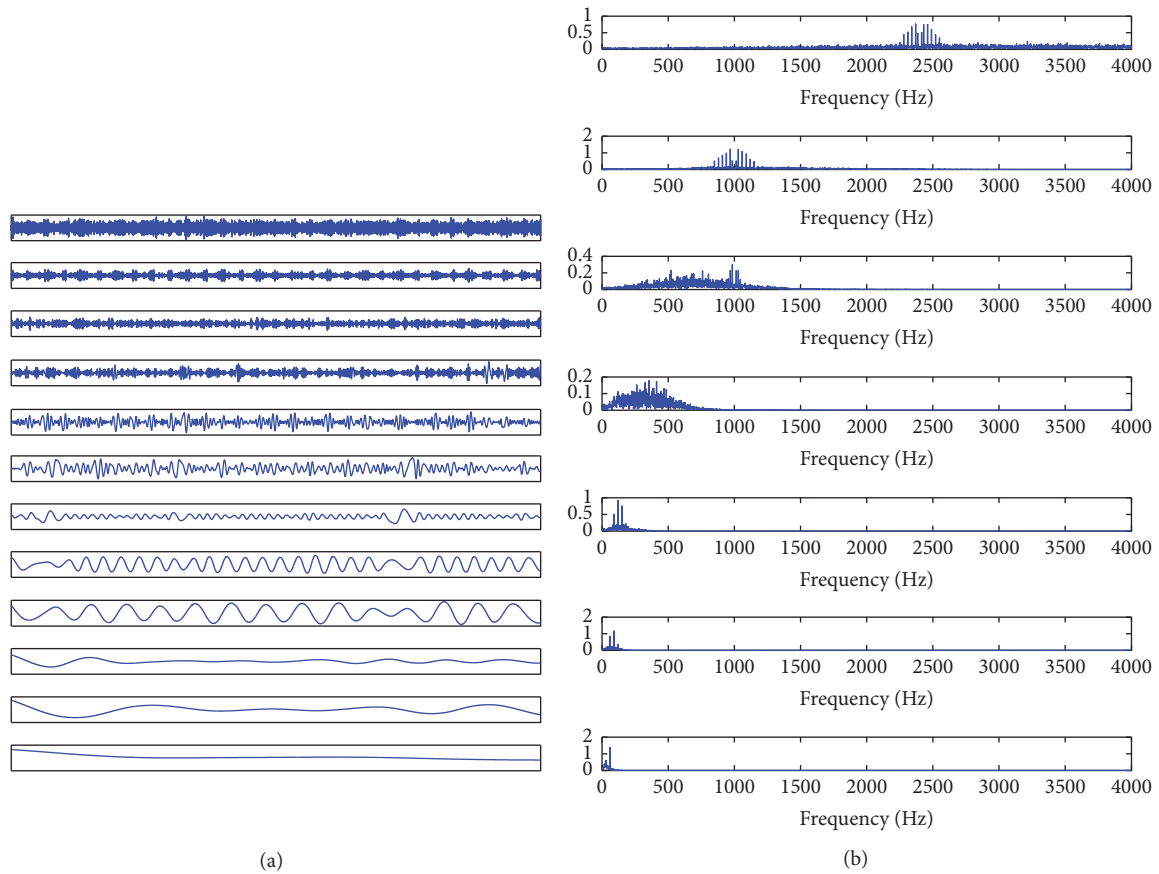


FIGURE 3: EMD subsignals (IMFs) (a) and frequency spectrums (b) of the first 7 IMFs of the simulation signal.

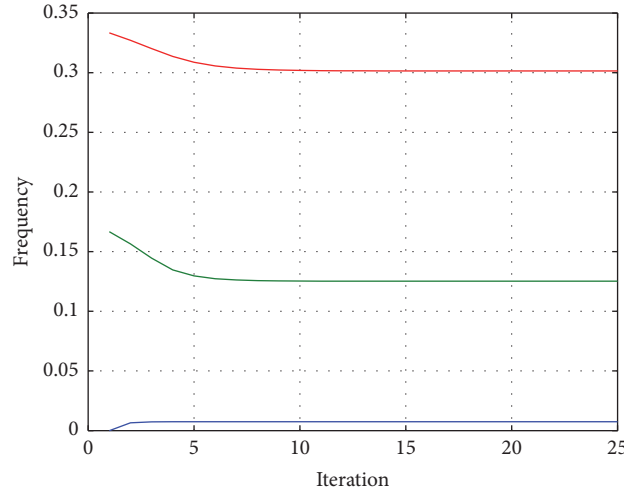


FIGURE 4: Center frequencies changing during VMD iteration.

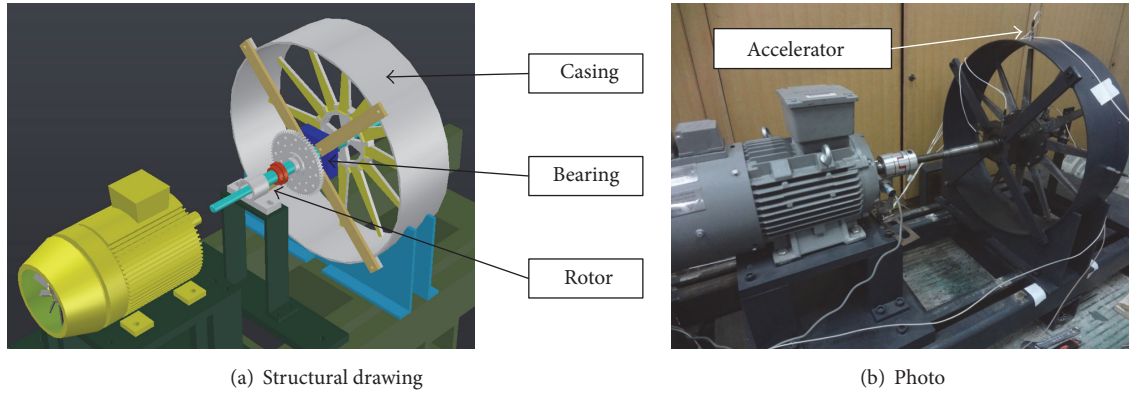


FIGURE 5: Rotor-bearing-casing test rig.

vibrational characteristics, which is useful for casing fault diagnosis. Accelerators are installed on both the bearing pedestal and the casing. The top rotating speed of the test rig can reach 3000 rpm without any fault. And the top sampling frequency of the accelerators is 15 kHz. In this paper, the vibration signals collected from the sensors on the casing are used to realize the fault diagnosis. Applying casing signals to diagnosis is a work more challenging than using bearing signals due to the complexity of signal components and the strong noise interference [19].

Five different types of mechanical faults are simulated on this test rig including unbalance, misalignment, supporting looseness, bearing looseness, and rub-impact. So plus the normal state there are 6 running conditions in total. The acceleration signals are obtained at the sampling frequency of 8000 Hz when the shaft at its rotating speed is near 1500 rpm and 1800 rpm. The signals are divided into time slices at the length of 0.5 s as samples to be further analyzed. Figure 6 shows the time waveform and frequency spectrum of one signal sample for each running condition at the rotating speed of 1800 rpm.

For now, there are 56 samples for each condition, 28 for 1500 rpm, and 28 for 1800 rpm. So, there are 336 samples in total. Some of them are chosen to be the training data and the rest are the testing data.

5. Vibrational Feature Extraction of Experimental Data

As the purpose of fault diagnosis is to distinguish vibration conditions from each other, more significant features extracted from the vibration signal often mean more possibility of diagnosing the faults correctly. Different types of features could be extracted from a signal sample.

The most common features of a signal sample are parameters in time domain and frequency domain. Features in time domain include time waveform and amplitude parameters. In this case, 12 different amplitude parameters are calculated, whose formulas are shown in Table 1. They reflect the time domain characteristics of the waveform in many ways. Figure 7 gives the distribution of 4 amplitude parameters out of all the 12 parameters, showing that mechanical faults do

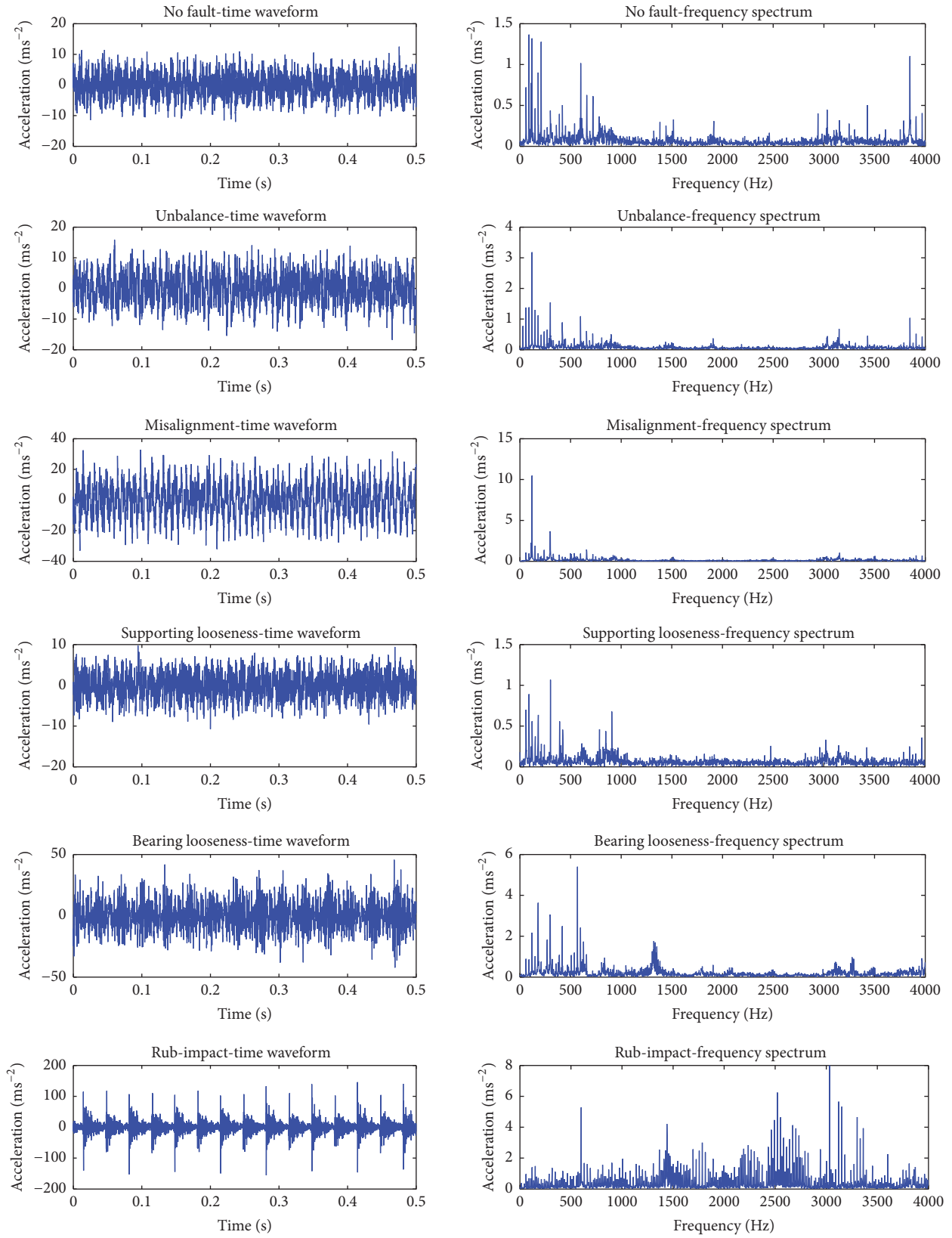


FIGURE 6: Time waveforms and frequency spectra for no fault, unbalance, misalignment, supporting looseness, bearing looseness, and rub-impact.

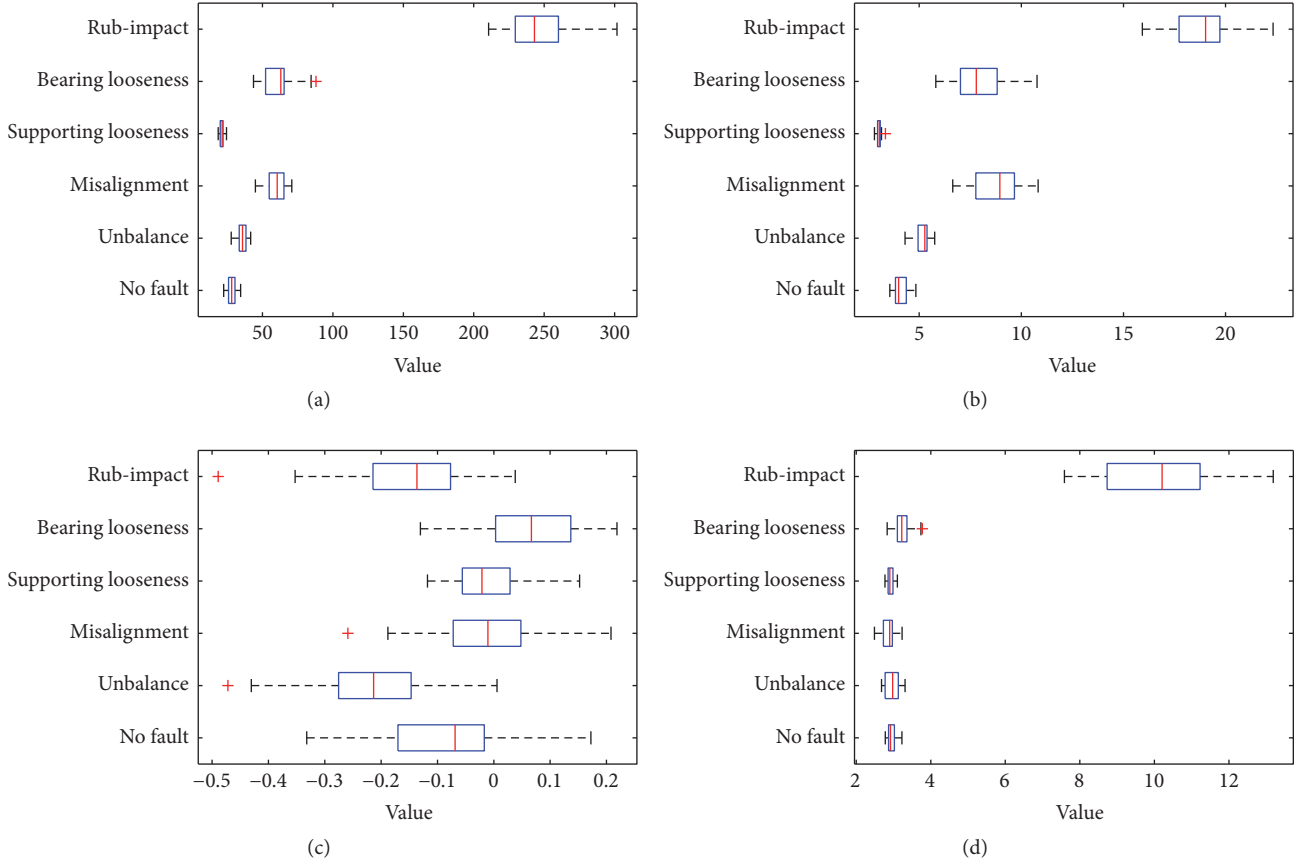


FIGURE 7: Boxplot of peak-to-peak value (a), root-mean-square (b), skewness (c), and kurtosis (d).

make changes to the distribution of these parameters, but not enough changes to completely distinguish the faults from each other.

Frequency spectrum displays the frequency domain components of a signal. Mechanical faults are often accompanied by changes in frequency components of the vibration signal. Common rotor faults, like unbalance and misalignment, may bring significant changes in low-frequency part of rotor's vibration signal, and mostly in rotational frequency and its harmonic frequencies. These changes in low-frequency part may be transferred from rotor to casing. Thus, amplitudes of the first 12 harmonic frequencies are chosen as features in frequency domain. The higher-order parts are left to VMD method to deal with.

Further feature extraction is based on VMD outputs [19, 20]. Here in Figure 8 are time waveforms, frequency spectrums, and envelope spectrums of IMFs from one signal sample. Envelope spectrum amplitude values at the rotational frequency and its 2nd to 10th harmonic frequencies of each subsignal are chosen to be features of a sample. When the harmonic order is larger than 10, envelope spectrum components are not quite clear, so they are not used.

Different faults will excite the resonances of different parts of the whole structure; then center pulsations ω_k of one condition are different from others. The values of center pulsations may also be one type of feature that could be

used in casing fault diagnosis. For example, Figure 8(d) is the plot of center pulsations for all the training data related to unbalance and rub-impact. The difference between these two groups is quite clear. The x-axis of the plot corresponds to 6 IMFs of the signal sample. The center frequencies of the first 3 modes of unbalance seem close to rub-impact, but the last 3 modes are clearly different.

6. Fault Classification Using SVM

Traditional SVM can only do binary classification. There are several methods for multiclass classification based on binary SVM, like one-versus-rest (1-v-r), one-versus-one (1-v-1), hierarchical support vector machines (H-SVMs), and Directed Acyclic Graph (DAG-SVMs) [21]. In this paper, methods of one-versus-rest and H-SVMs are used and compared.

(i) *One-versus-Rest*. Every time of training choosing one category as a class and all other categories as another class, then K binary SVMs are built. When doing classification for a sample, each SVM has a result of indicator function. This sample is assigned to the class where the SVM has the biggest indicator.

(ii) *H-SVMs*. By building binary SVMs, all the categories are divided into two subclasses and each subclass is divided again

TABLE 1: Amplitude parameter formulas.

Peak-to-peak value	$X_{p-p} = \max(x_i) - \min(x_i)$
Variance	$D_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$
Standard deviation	$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2}$
Root-mean-square	$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Square-mean-root	$X_r = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right]^2$
Mean amplitude	$ \bar{X} = \frac{1}{N} \sum_{i=1}^N x_i $
Skewness	$\alpha = \frac{1}{N} \sum_{i=1}^N x_i^3$
Kurtosis	$K_v = \frac{(1/N) \sum_{i=1}^N x_i^4}{X_{rms}^4}$
Waveform factor	$S_f = \frac{X_{rms}}{ \bar{X} }$
Crest factor	$C_f = \frac{X_{max}}{X_{rms}}$
Impulse factor	$I_f = \frac{X_{max}}{ \bar{X} }$
Margin factor	$CL_f = \frac{X_{max}}{X_r}$

until every subclass has only one category. In this process, $K-1$ binary SVMs are built in total.

Figure 9 is the procedure of training and testing a SVM classifier. There are 6 classes, whose labels are listed in Table 2. Table 3 is an example showing feature parameters of a sample. Vibrational features in amplitude domain and frequency domain are included, as well as envelope spectrum parameters of VMD subsignals and center pulsations of VMD modes. Only the first 2 IMFs are used to obtain VMD-Envelope parameters since the high-frequency subsignals contain too much noise which will cause interference with the process of fault classification. All these parameters could be part of an input vector used in the SVM training. Because the vibration data are under different rotating speeds, the value of rotational frequency of the main shaft is also included in the input vector and all other input vectors in the training.

Different types of features are used as the training data of the multiclass classifier. Each type is used individually at first. Both one-versus-rest and H-SVMs methods give a classification result. Polynomial function kernel is chosen as the kernel function of SVM because its performance is not worse than other kernel functions and it only has one discrete kernel parameter which is easy to optimize. Polynomial

function orders 1 to 7 are tested and the one giving the best classification result will be the kernel parameter.

Half of all the samples are selected as training data and the rest are testing data (168 out of 336). Training data are randomly selected for 10 times, and each time the training and testing will yield classification accuracy. The mean accuracy of these 10 times is used as the accuracy of the classifier with this specific input type. The closer the accuracy is to 100%, the better the classification will be. The results below are all average accuracy of 10 classifications.

Then all these features are combined into one single input vector. The classification results of different training data are listed in Table 4. SVMs trained by single type of feature do not yield satisfying results. But when combined features are used, the results will be improved. And in most cases, the one-versus-rest method performs better than the H-SVMs method. To compare with the performance of SVM, two kinds of artificial neural networks, including radial basis function neural network (RBF) [22] and probabilistic neural network (PNN) [23], are also tried as the classification algorithm. The results of RBF and PNN are in Table 4 as well, which shows that in most cases SVM has better results than PNN, and PNN has better results than RBF.

7. Dimensionality Reduction Using LLE

There are too many elements in the feature vector and some of them have little variance for different categories. The dimensionality reduction (DR) methods can reduce the dimensionality and simplified the input vectors. The first thought of dimensionality reduction is Principal Component Analysis (PCA), which is a linear dimensionality reduction method of simple theory but high efficiency. It is widely used to decrease the size of feature vectors. Figure 10 is the flow chart of how PCA and SVM are combined to realize the training and testing [24]. The loadings of PCA gained from the training data could also map the testing data into the dual space, where the first several components keep most of the information. During the training of PCA there are no many parameters to be selected, for only the number of dimensions after PCA needs to be decided. Classification results with PCA are shown in Table 4.

PCA is sensitive to the relative scaling of the original variables. Doing PCA before SVM training means smaller feature vector size but may not lead to much improvement of SVM's classification accuracy, as seen from the results. The performance without PCA is equivalent to, sometimes even better than, the one with PCA. One explanation is that, as the training of SVM has involved an optimization procedure, PCA, the linear dimensionality reduction method may not bring in any improvement when the kernel parameters and other parameters are already well selected.

To improve the performance of the classifier, PCA used above can be replaced with nonlinear dimensionality reduction approaches, like local linear embedding (LLE) that has been introduced in this paper. So, the flow chart is updated to Figure 11. Although LLE is a nonlinear dimensionality reduction method, it does not have many parameters. In addition

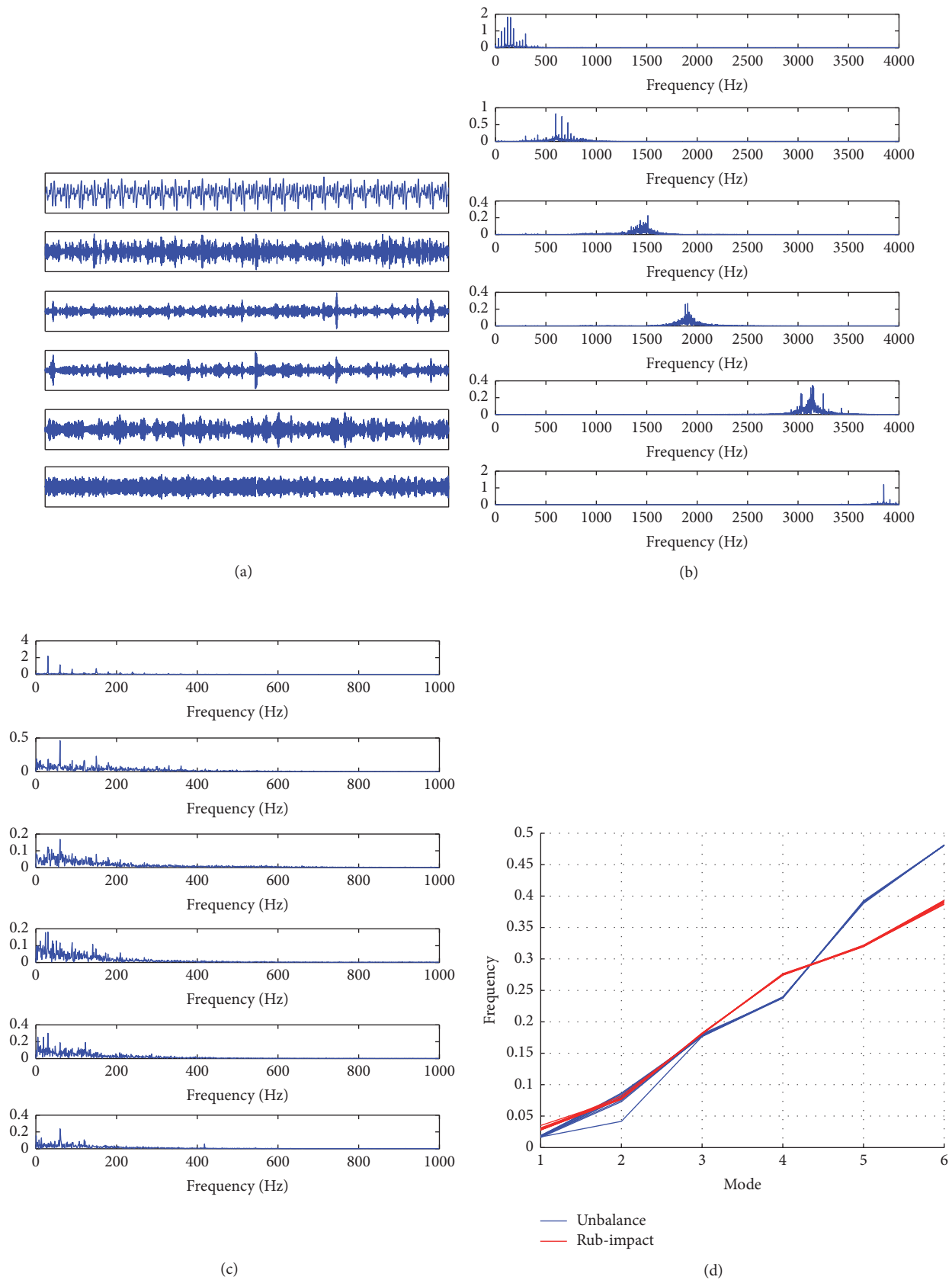


FIGURE 8: VMD subsignals (a), their frequency spectra (b), and envelope spectra (c). Center pulsations for all the training data concerning unbalance and rub-impact (d).

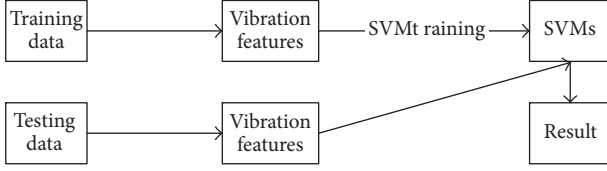


FIGURE 9: Flow chart of fault diagnosis using SVM.

to the number of dimensions, the number of nearest points, k , is another parameter to select during the optimization.

Combined features are used as input vectors here. The accuracy results in Table 5 show that LLE not only reduce the dimension of the input data but do improve the classification performance of SVM. Using two types of multiclass SVM, the classification accuracies both reach 97%. It seems strange that after dimensionality reduction the kernel function of SVM will map the data into a higher-dimensional space again. But, like nonlinear kernel function which makes the separation of different categories much easier, this nonlinear DR method could make the datasets more separable if the parameters are well optimized. These two methods seem to be contradictory but work well in fault classification problems, like the casing fault diagnosis problem in this paper.

Another important thing is that classifiers here are trained by vibration data under two rotating speeds. The results are still good, which means that the procedure is not very sensitive to rotating speed.

Some other nonlinear DR methods may not have the same performance in this problem. For example, Isomap is similar to LLE, but more likely to provide global properties of data, which makes it more sensitive to short-circuiting [25]. The results show that choosing Isomap as the DR process, the performance of the classifier declines a little. Some other DR methods are tried and the results are also listed below, like Laplacian Eigenmaps, Locality Preserving Projection (LPP), and Neighborhood Preserving Embedding (NPE) [26, 27]. Other DR methods do not suit this problem well like LLE.

The selected number of training samples has a certain influence on the classification results. Previously half of the samples are used as training data. The number of training samples is then decreased to 120 and increased to 216, and the classification results are in Table 6. More training samples will lead to higher classification accuracy. Here when the selected proportion of all the samples changes from 5/14 to 7/14 to 9/14, the accuracy will increase for no more than 2% each time. Although the influence of sample number is relatively small when the classification method is well chosen, more training data are preferred since the goal of classification is always 100%.

8. Discussion

The fault diagnosis process based on vibration signal could be simply described as 3 steps: feature extraction, feature optimization, and fault classification. The main techniques employed in this paper, including variational mode decomposition, local linear embedding, and support vector

TABLE 2: Class labels.

Class	Label
Normal	1
Unbalance	2
Misalignment	3
Supporting looseness	4
Bearing looseness	5
Rub-impact	6

machine, are dealing with the 3 steps above. The analysis above proves that, at least for the experimental data, variational mode decomposition is better than empirical mode decomposition, local linear embedding is better than other dimensionality reduction methods, and support vector machine is better than radial basis function neural network and probabilistic neural network. So the application of these methods makes the diagnosis procedure practical and effective. The diagnosis procedure proposed in this paper contains not only the application of these techniques, but a signal analysis and feature extraction process for casing signals. Vibrational features can be extracted from many aspects, including statistical properties, low-frequency characteristics, and high-frequency resonance characteristics. Figure 12 shows the flow chart of the whole procedure.

The time waveforms and frequency spectrums have shown the complexity of the experimental signals, even though the mechanical structure of the test rig is relatively simple. The vibration signal of a real rotor-bearing-casing system must be more complex, but the fault diagnosis procedure is similar. Vibrational features could be extracted from time domain, frequency domain, and VMD output too. Parameters like VMD mode number and harmonic order should be adjusted according to the real structure and its characteristics. However, when some special faults happen, like bearing faults and gearbox faults, additional analysis is necessary for more detailed features, otherwise the faults are hard to be detected. More signal preprocessing is required because of more influence from the inner structure. And for real systems the data amount will be much larger, so the classifier training progress will be quite complicated since the diagnosis procedure involves optimizations for several important parameters.

9. Conclusion

This paper presents a procedure of training a mechanical fault classifier of rotor-bearing-casing system, using casing vibration signals. The database of signals used here was obtained from the casing of the rotor-bearing-casing test rig. The test section of the rig is considered to be a rotor-bearing-casing system simplified from part of a gas turbine engine. So this procedure can be a reference of mechanical fault diagnosis for gas turbine engine and other rotating machineries which have similar structure.

Instead of High-Frequency Resonance Technique, here VMD decomposes the casing signal into several modes,

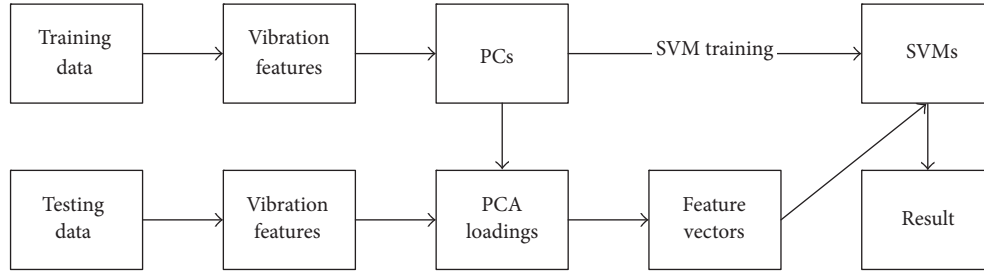


FIGURE 10: Flow chart of fault diagnosis using PCA and SVM.

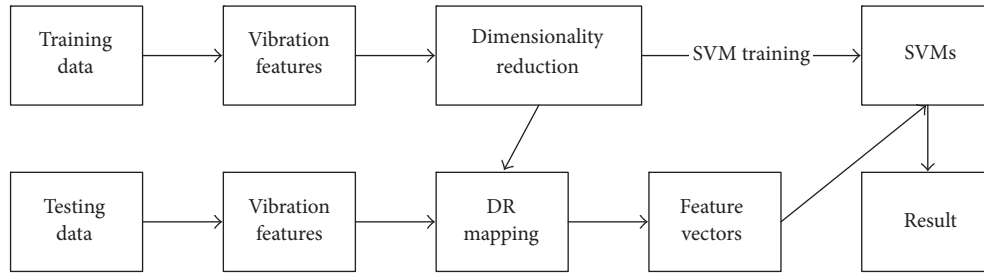


FIGURE 11: Flow chart of fault diagnosis using DR and SVM.

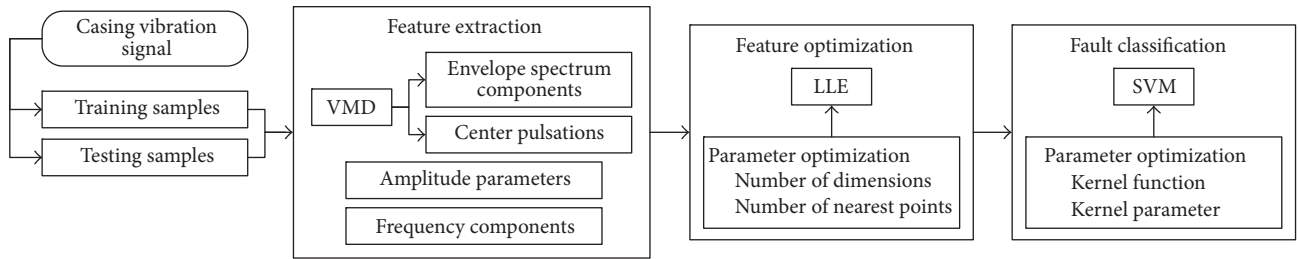


FIGURE 12: Procedure of casing vibration fault diagnosis in this paper.

TABLE 3: An example of feature parameters.

	Number											
	1	2	3	4	5	6	7	8	9	10	11	12
Amplitude	38.2316	25.3643	5.0363	5.0360	3.3729	3.9924	-0.2661	3.1359	1.2614	7.5917	9.5761	11.3350
Frequency	0.9411	1.3504	1.3869	1.8874	1.8037	1.2030	0.4113	0.5810	0.8591	1.9295	0.1392	0.1262
VMD-Envelope 1	2.1396	1.1169	0.6736	0.1940	0.8120	0.3441	0.2624	0.3452	0.2311	0.0976		
VMD-Envelope 2	0.2212	0.4462	0.1413	0.1871	0.2477	0.1253	0.1061	0.0608	0.0855	0.0855		
Omega	0.0189	0.0832	0.1794	0.2382	0.3906	0.4807						

TABLE 4: Classification results with different training data.

Training data	RBF	PNN	1-v-r	H-SVMs	1-v-r with PCA	H-SVMs with PCA
Amplitude	67.3%	82.1%	81.7%	75.6%	79.8%	74.5%
Frequency	88.2%	92.5%	94.2%	90.4%	90.1%	86.1%
Amplitude + Frequency	83.8%	87.3%	96.3%	93.6%	96.6%	94.0%
VMD-Envelope	81.6%	85.8%	85.8%	80.4%	81.3%	76.0%
VMD-Envelope + Omega	84.2%	86.3%	91.8%	87.8%	81.4%	76.1%
Amplitude + Frequency + VMD-Envelope	87.2%	88.3%	96.4%	94.8%	95.6%	94.5%
Amplitude + Frequency + VMD-Envelope + Omega	86.5%	88.3%	97.1%	96.3%	95.6%	94.5%

TABLE 5: Classification results with different DR methods.

DR method		No DR	PCA	LLE	Isomap	Landmark Isomap	Laplacian Eigenmaps	LPP	NPE
Amp + Freq + VMD	1-v-r	96.4%	95.6%	98.3%	76.1%	80.7%	82.8%	40.8%	97.3%
	H-SVMs	94.8%	94.5%	97.4%	66.7%	76.7%	81.0%	35.5%	95.7%
Amp + Freq + VMD + Omega	1-v-r	97.1%	95.6%	98.2%	76.1%	80.0%	82.7%	37.5%	97.4%
	H-SVMs	96.3%	94.5%	97.1%	66.6%	75.2%	81.4%	33.4%	95.5%

TABLE 6: Classification results with different selected numbers of training samples.

Training samples/testing samples	Training data	RBF	PNN	1-v-r	H-SVMs	1-v-r with PCA	H-SVMs with PCA	1-v-r with LLE	H-SVMs with LLE
120/216	Amp + Freq + VMD	85.0%	87.4%	95.9%	93.4%	94.3%	91.3%	97.1%	96.0%
	Amp + Freq + VMD + Omega	84.8%	87.4%	96.6%	95.0%	94.3%	91.3%	97.3%	95.9%
	Amp + Freq + VMD	87.2%	88.3%	96.4%	94.8%	95.6%	94.5%	98.3%	97.4%
168/168	Amp + Freq + VMD + Omega	86.5%	88.3%	97.1%	96.3%	95.6%	94.5%	98.2%	97.1%
	Amp + Freq + VMD	91.8%	90.6%	97.9%	96.3%	97.3%	95.4%	98.9%	98.3%
216/120	Amp + Freq + VMD + Omega	91.8%	90.6%	98.4%	97.2%	97.4%	95.3%	98.9%	98.3%

which are usually subsignals modulated by fault frequencies. Amplitude values at rotational frequency and harmonic frequencies of the modes on their envelope spectrums are chosen as part of the features, as well as center frequencies of the modes calculated by VMD, and feature parameters in time and frequency domain. LLE is used to reduce their dimensionality of feature vectors. Then the multiclass SVM is trained by the low-dimensional data. Comparing to PCA the linear DR method and other nonlinear DR methods, LLE is capable of not only reducing the dimensionality, but also turning the data samples into more separable ones if its parameters were well chosen. The classification can reach accuracy of 98% in dealing with the experimental data, in which the features of amplitude, frequency, and VMD are combined together, using LLE for DR, and then training SVM. Considering the complexity and the randomness inside the experimental data, this result shows the procedure proposed here is a promising alternative of fault classifiers.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The research work is supported by the National Natural Science Foundations of China (no. 11572167). The authors

would like to express thanks to laboratory associate Xiangyi Dang who built the rotor-bearing-casing test rig.

References

- [1] C. Guo, "Study on the recognition of aero-engine blade-casing rubbing fault based on the casing vibration acceleration," *Measurement: Journal of the International Measurement Confederation*, vol. 65, pp. 71–80, 2015.
- [2] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, no. 1, pp. 61–78, 1998.
- [3] Z. K. Peng, P. W. Tse, and F. L. Chu, "A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing," *Mechanical Systems & Signal Processing*, vol. 19, no. 5, pp. 974–988, 2005.
- [4] N. E. Huang, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [5] J. S. Smith, "The local mean decomposition and its application to EEG perception data," *Journal of the Royal Society Interface*, vol. 2, no. 5, pp. 443–454, 2005.
- [6] M. G. Frei and I. Osorio, "Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2078, pp. 321–342, 2007.

- [7] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [8] H. U. Jie, "Survey on feature dimension reduction for high-dimensional data," *Application Research of Computers*, vol. 25, no. 9, pp. 2601–2606, 2008.
- [9] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 1986.
- [10] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [11] A. J. Hoffman and N. T. Van Der Merwe, "The application of neural networks to vibrational diagnostics for multiple fault conditions," *Computer Standards and Interfaces*, vol. 24, no. 2, pp. 139–149, 2002.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 2nd edition, 2000.
- [13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining & Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [14] X. An and J. Yang, "Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy," *Transactions of the Institute of Measurement and Control*, vol. 38, no. 3, pp. 282–292, 2016.
- [15] Y. Liu, Z. Yu, M. Zeng, and Y. Zhang, "An improved LLE algorithm based on iterative shrinkage for machinery fault diagnosis," *Measurement: Journal of the International Measurement Confederation*, vol. 77, pp. 246–256, 2016.
- [16] J. A. K. Suykens, "Nonlinear modelling and support vector machines," in *Proceedings of the IEEE Instrumentation & Measurement Technology Conference*, vol. 1, pp. 287–294, IEEE, Budapest, Hungary, May 2001.
- [17] H. Prashad, M. Ghosh, and S. Biswas, "Diagnostic monitoring of rolling-element bearings by high-frequency resonance technique," *ASLE Transactions*, vol. 28, no. 4, pp. 439–448, 1985.
- [18] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," in *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03)*, Grado, Italy, June 2003.
- [19] G. Chen, "Simulation of casing vibration resulting from blade-casing rubbing and its verifications," *Journal of Sound and Vibration*, vol. 361, pp. 190–209, 2016.
- [20] T. Han and D. Jiang, "Rolling bearing fault diagnostic method based on VMD-AR model and random forest classifier," *Shock and Vibration*, vol. 2016, Article ID 5132046, 11 pages, 2016.
- [21] E. Mayoraz and E. Alpaydin, "Support vector machines for multi-class classification," in *Engineering Applications of Bio-Inspired Artificial Neural Networks*, vol. 1607 of *Lecture Notes in Computer Science*, pp. 833–842, Springer, Berlin, Germany, 1999.
- [22] A. J. Hoffman and N. T. Van Der Merwe, "The application of neural networks to vibrational diagnostics for multiple fault conditions," *Computer Standards & Interfaces*, vol. 24, no. 2, pp. 139–149, 2002.
- [23] S. Samarasinghe, *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*, CRC Press, 2006.
- [24] C. Jing and J. Hou, "SVM and PCA based fault classification approaches for complicated industrial process," *Neurocomputing*, vol. 167, pp. 636–642, 2015.
- [25] Z. Su, B. Tang, J. Ma, and L. Deng, "Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier," *Measurement: Journal of the International Measurement Confederation*, vol. 48, no. 1, pp. 136–148, 2014.
- [26] L. J. P. Van Der Maaten, E. O. Postma, and H. J. V. D. Herik, "Dimensionality reduction: a comparative review," *Journal of Machine Learning Research*, vol. 10, no. 1, 2007.
- [27] L. van der Maaten, *Matlab Toolbox for Dimensionality Reduction*, 2007.

