

2825. Novel bearing fault diagnosis model integrated with dual-tree complex wavelet transform, permutation entropy and optimized FCM

Ping Ma¹, Hongli Zhang², Wenhui Fan³, Cong Wang⁴

^{1,2,4}College of Electrical Engineering, Xinjiang University, Urumqi, China

³Department of Automation, Tsinghua University, Beijing, China

²Corresponding author

E-mail: ¹694073078@qq.com, ²xj dq mp@foxmail.com, ³mapping_0000@sina.com, ⁴2896438968@qq.com

Received 23 February 2017; received in revised form 22 July 2017; accepted 29 July 2017

DOI <https://doi.org/10.21595/jve.2017.18278>



Abstract. In order to enhance the capability of feature extraction and fault classification of bearings, this study proposes a feature extraction approach based on dual-tree complex wavelet transform (DTCWT) and permutation entropy (PE), using the fuzzy c means clustering (FCM) to identify fault types. The vibration signal of bearings can be decomposed into several wavelet components with DTCWT which can describe the local characteristics of vibration signals accurately. And the PE of each wavelet component, which can describe the complexity of a time series, is calculated to be regarded as the fault features. Then forming the standard clustering centers by the FCM, we defined a standard using the Hamming approach degree to evaluate the classification results in the FCM. In order to verify the effectiveness of the proposed approach, compared with two other typical signal analysis methods: ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD), through extracting fault features, it required to identify the fault types and severities under variable operating conditions. The experimental results demonstrate that the proposed approach has a better accuracy and performance to diagnose a bearing fault under different fault severities and variable operating conditions. The proposed approach is suitable for a fault diagnosis due to its good ability to the feature extraction and fault classification.

Keywords: bearing fault diagnosis, features extraction, dual-tree complex wavelet transform, permutation entropy, optimized FCM.

1. Introduction

Bearings are the key components in the mechanical structure, and the condition monitoring and fault diagnosis are critical to ensure the normal operation of machinery and equipment. The process of bearing fault diagnosis mainly consists of the fault feature extraction and fault classification [1, 2]. In case of a bearing failure, the frequency band energy of vibration signals will change. If frequency band signal characteristics can be extracted, the bearing fault classification can be carried out. Therefore, the effective extraction of fault feature is the key for a fault diagnosis.

The main problem of the feature extraction is how to process the different bearing vibration signals [3, 4]. Sometime-frequency analysis methods have been proposed to diagnose fault of rotating machinery in past few years, such as the empirical mode decomposition (EMD) [5, 6], local mean decomposition (LMD) [7], ensemble empirical mode decomposition (EEMD) [5, 8], and variational mode decomposition (VMD) [9]. The EMD is a self-adaptive method, for analyzing nonlinear and non-stationary signals to decompose the signals into a few intrinsic mode functions (IMFs). Each IMF is regarded as a mono-component [10]. Like with EMD, LMD decomposes the signals into a few product functions (PFs), each PF is a product of envelope and frequency modulated signals, from which the instantaneous frequency can be derived [11]. Yet the EMD has some drawbacks, including end effects, mode mixing and border effect problem. So, the EEMD has been proposed to eliminate the mode mixing problem of EMD [12]. The EMD, LMD and EEMD belong to recursive mode decomposition, by contrast, the VMD transforms the

signal decomposition into non-recursive, variational modal decomposition model, and has better noise robustness [9, 13]. Thus, these time-frequency analysis methods, which combine the autoregressive model (AR), singular value decomposition (SVD) [2, 14], entropy-based method [8, 15, 16] and support vector machine (SVM) [9, 17], are widely used in the bearing fault diagnosis. However, there are some different problems in these methods. As mentioned above, the EMD and LMD also have mode combination and border effect problems, and they are all affected by the sampling frequency. The EEMD decomposition process is affected by the parameters so that it still has illusive components [12, 17]. Similarly, the VMD also is affected by the parameter determination [18].

So far, the discrete wavelet transforms (DWT) and second-generation wavelet transform (SGWT) have gained enormous applications for signal processing and bearing fault diagnosis in the literatures [19-21]. The DWT has the advantages of low entropy, multi resolution, flexible selection of substrates. While the main disadvantage of DWT consists in that it only decomposes the lowest frequency sub-band of signals and lacks the shift invariance, which may weaken the identification of fault features [22]. The SGWT is a wavelet transform based on a lifting scheme, but it also lacks the shift invariance and has severe aliasing in the process of decomposition. In order to overcome the defects of the above wavelet, the dual-tree complex wavelet transform (DTCWT) was initially proposed by Kingsbury [23], and Selesnick [24] who turned it up to the dyadic case. It was demonstrated that the performance of DTCWT was better than that of DWT and SGWT. The DTCWT retains the excellent properties of complex wavelet transform, and uses the dual-tree filter form to ensure the perfect signal reconstruction. Therefore, the dual tree complex wavelet transform is a kind of wavelet transform. It has good properties of approximate shift invariance, good directional selectivity, limited data redundancies, perfect reconstruction and high computational efficiency [24]. Because the DTCWT enjoys many attractive properties such as nearly shift-invariance and reduced aliasing which may be suitable to both the surveillance and diagnosis of rotating machine. And it has been used in the image processing [25], signal noise reduction processing [26] and engine fault diagnosis [27] etc.

In this literature, a new technique is proposed based on the DTCWT and permutation entropy (PE) to extract fault features in bearing vibration signals. The PE is an entropy-based method to analyze the signals complexity [28]. The permutation entropy is used in the statistical measuring method. It describes the local order structure and complexity of a time series via phase space reconstruction, and gets a great importance to the nonlinear behavior of the time series. It has been applied in many fields with the advantages of simplicity, robustness and invariance to nonlinear monotonous transformations [16, 29]. The wavelet components gained from the DTCWT can accurately describe the local characteristics of vibration signals, and the PEs of wavelet components can reflect more useful fault information. Therefore, in this study, the PE of a wavelet component is calculated to be regarded as the signal feature, and it represents the features of different signal types and is used to separate different faults. Meanwhile, the standard fuzzy C means clustering (FCM) is used to identify fault types, and the hamming approach degree is introduced to FCM to evaluate the result of clustering. the FCM is one of the most widely used algorithms in fuzzy clustering, which divides the samples of high similarity into the same category via iteration and optimization of the objective function [30]. And it also has been widely applied in the fault detection and diagnosis [31, 32].

The rest of this paper is organized as follows. The DTCWT, permutation entropy, optimized FCM are briefly reviewed in Section 2. Section 3 describes the proposed fault diagnosis method. Then in Section 4, the proposed approach is compared with the EEMD and VMD via extracting multiple fault signatures in the vibration signal and identifying the fault types and severities of bearings, using bearing vibration signals for validation. Finally, the conclusion is presented.

2. Related works

2.1. Dual-tree complex wavelet transform(DTCWT)

The DTCWT is a relatively recent enhancement to the DWT, which has important additional properties including nearly shift-invariant and directionally selective in two and higher dimensions [24]. The DTCWT uses two parallel DWTs (real tree and imaginary tree) with different low-pass and high-pass filters in each scale for signal decomposition and reconstruction. The decomposition and reconstruction of DTCWT are shown in Fig. 1.

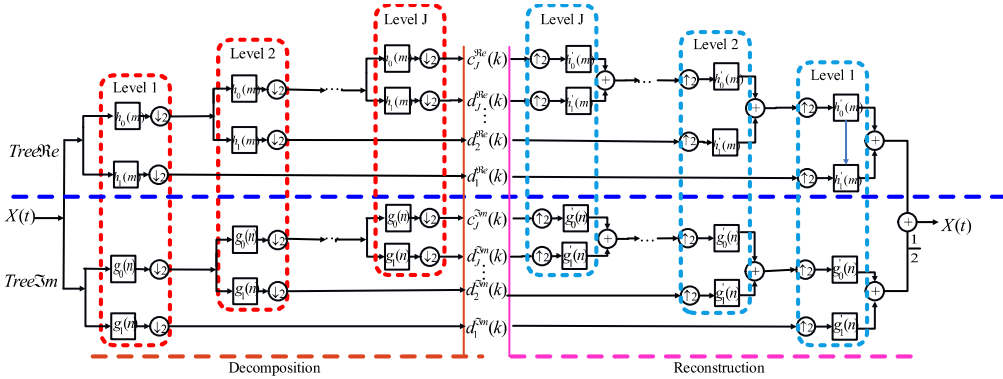


Fig. 1. DTCWT decomposition and reconstruction

These two DWTs use two different sets of filters, and they satisfy the perfect decomposition and reconstruction. Let $\psi_h(t)$ and $\psi_g(t)$ represent the used two real-valued wavelets of DTCWT, and $\varphi_h(t)$ and $\varphi_g(t)$ are the corresponding scaling functions. A complex-valued wavelet $\psi(t)$ can be denoted as [26]:

$$\psi(t) = \psi_h(t) + i\psi_g(t). \tag{1}$$

Then the DTCWT consists of two parallel wavelet transforms, therefore, according to the wavelet theory, the wavelet coefficients $d_j^{Re}(k)$ and the scaling coefficients $c_j^{Re}(k)$ of the above upper tree can be calculated by inner products as follows:

$$d_j^{Re}(k) = 2^{j/2} \int_{-\infty}^{+\infty} x(t) \psi_h(2^j t - k) dt, \quad j = 1, 2, \dots, J, \tag{2}$$

$$c_j^{Re}(k) = 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_h(2^j t - k) dt, \tag{3}$$

where j is the scale factor and J is the maximum decomposition scale. Similarly, the wavelet coefficients $d_j^{Im}(k)$ and scaling coefficients $c_j^{Im}(k)$ of the above lower tree can be computed via Eqs. (4) and (5):

$$d_j^{Im}(k) = 2^{j/2} \int_{-\infty}^{+\infty} x(t) \psi_g(2^j t - k) dt, \quad j = 1, 2, \dots, J, \tag{4}$$

$$c_j^{Im}(k) = 2^{j/2} \int_{-\infty}^{+\infty} x(t) \varphi_g(2^j t - k) dt. \tag{5}$$

Then, the wavelet coefficients $d_j^c(k)$ and the scaling coefficients $c_j^c(k)$ of the DTCWT can be obtained by combining the dual tree output as follows:

$$d_j^c(k) = d_j^{Re}(k) + id_j^{Im}(k), \quad j = 1, 2, \dots, J, \tag{6}$$

$$c_j^c(k) = c_j^{Re}(k) + ic_j^{Im}(k). \tag{7}$$

In the decomposition process, the input signal through the low-pass and high-pass filters, respectively. Since the signal frequency is divided into high-pass and low-pass, the bandwidth of the filtered output signal is only a half of the original signal. According to the sampling theorem of bandlimited signals, it is possible to reduce the sample rate by half without losing any information. Therefore, using binary sampling in the decomposition is reasonable, and the number of the obtained wavelet coefficients reduced by half with the depth of decomposition.

So, the complex wavelet coefficients and last-level scaling function coefficient are reconstructed to correspondingly obtain the detail signals at all levels and the approximation signal at the last level by the following equations:

$$d_j(t) = 2^{(j-1)/2} \left[\sum_n d_j^{Re}(k) \psi_h(2^j t - n) + \sum_m d_j^{Im}(k) \psi_g(2^j t - m) \right], \quad j = 1, \dots, J, \tag{8}$$

$$c_j(t) = 2^{(j-1)/2} \left[\sum_m c_j^{Re}(k) \varphi_h(2^j t - m) + \sum_n c_j^{Im}(k) \varphi_g(2^j t - n) \right], \tag{9}$$

where m and n are the lengths of the adopter filters.

Finally, the reconstructed signal $x'(t)$ is equal to the sum of all detail signals and approximation signal:

$$x'(t) = c_j(t) + \sum_{j=1}^J d_j(t). \tag{10}$$

In the literature [24], it was proposed to design methods for dual-tree filters based on the half-sample delay condition. In this paper, (13, 19) – tap near symmetry biorthogonal filters are used at the first level, and the 14-tap linear phase filters produced by the Q-shift solution are used at the level except the first one. For more detailed information about the design of dual-tree filters in this paper, please refer to the literature [33].

2.2. Basic principle of permutation entropy

Permutation entropy (PE) is used to analyze the data complexity [28]. The permutation entropy is a statistical measuring method that has been effectively applied in many fields. The basic principle of PE is described as follows.

For a time series $\{x(i), i = 1, 2, \dots, N\}$, a d -dimensional delay embedding matrix is expressed as:

$$\begin{aligned} X(1) &= \{x(1), x(1 + \tau), \dots, x(1 + (d - 1)\tau)\}, \\ X(2) &= \{x(2), x(2 + \tau), \dots, x(2 + (d - 1)\tau)\}, \\ X(i) &= \{x(i), x(i + \tau), \dots, x(i + (d - 1)\tau)\}, \\ &\vdots \\ X(N - (d - 1)\tau) &= \{x(N - (d - 1)\tau), x(N - (d - 2)\tau) \dots x(N)\}, \end{aligned} \tag{11}$$

where d denotes the number of the embedding dimension, τ is the time delay. For $d = 2$, there are two possible ordinal patterns of $X(i)$, namely $\pi_1 = x(i) < x(i + \tau)$ and $\pi_2 = x(i + \tau) < x(i)$. The d number of real values $X(i)$ can be arranged in an increasing order:

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \leq x(i + (j_d - 1)\tau), \tag{12}$$

where j represents the index in the column of embedding vector's element and $1 \leq j_i \leq d$. In general, there are just $d!$ possible order patterns.

If there are two or more elements having the same value, e.g. $x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau)$, we can sort the original positions for $j_1 \leq j_2$ and $x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau)$ can be recorded. Any $X(i)$ is uniquely mapped onto $\pi = (j_1, j_2, \dots, j_d)$, which is one of the $d!$ permutations of d distinct symbols. When each such permutation is considered as a symbol, then the reconstructed trajectory in the m -dimensional space is represented by a symbol sequence S_d . We can say that $X(i)$ has a permutation π , if and only if π is unique in S_d and satisfies Eq. (11).

For each permutation π , the relative frequency can be calculated as follows:

$$p(\pi) = \frac{\text{Number}\{X(i) \text{ has type } \pi | 1 \leq i \leq N - (d - 1)\tau\}}{N - (d - 1)\tau}. \tag{13}$$

According to the Shannon's entropy, the permutation entropy $H_p(d)$ of order d for the time series is defined as:

$$H_p = - \sum_{\pi} p(\pi) \ln(p(\pi)). \tag{14}$$

2.3. Optimized fuzzy C means clustering

The FCM is the most classical clustering algorithm based on the objective function. Its central message is that each data point is associated with a cluster through a membership degree. And this technique divides a collection of data points into fuzzy groups and finds a cluster center in each group such that a cost function of a dissimilarity measure is minimized [31].

For a data sample $X = \{x_1, x_2, \dots, x_N\}$, giving the fuzzy membership matrix $U = [u_{ij}]_{c \times N}$ and the clustering center $C = [c_1, c_2, \dots, c_c]^T$, the most popular clustering method for identifying optimal fuzzy c -partitions in X is associated with the generalized least-squared errors functional [31]:

$$\begin{cases} \min\{J(U, C^T)\} = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m d_{ij}^2, \\ \text{s. t. } \sum_{i=1}^c u_{ij} = 1, \quad \forall j = 1, \dots, n, \end{cases} \tag{15}$$

$$d_{ij}^2 = \|x_j - c_i\|_A^2 = (x_j - c_i)^T A (x_j - c_i),$$

where $m \in [1, +\infty)$ is weighting exponent, N and c represent the number of the data samples and clustering centers, respectively. U is fuzzy c -partitions in X . c_i is the center of cluster i . The squared distance d_{ij}^2 between x_j and c_i is computed using metric matrix A . A is positive-definite ($n \times n$) weight matrix. $\sum_{i=1}^N \sum_{i=1}^c (u_{ik})^m d_{ik}^2$ represents the overall weighted sum of generalized A -errors due to replacing X by C .

Then, the extremum problem with constraint condition converts to the problem of unconstrained condition by introducing the operator $\lambda = [\lambda_1, \dots, \lambda_n]$:

$$\bar{J}(U, C^T, \lambda) = J(U, c_1, \dots, c_c) + \sum_{j=1}^N \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) = \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right). \quad (16)$$

In the FCM algorithm, the fuzzy partitioning is obtained by an iterative optimization of the above function (16). The membership functions and cluster centers are updated by the following:

$$c_i = \frac{\left(\sum_{j=1}^N u_{ij}^m x_j \right)}{\sum_{j=1}^N u_{ij}^m}, \quad 1 \leq i \leq c, \quad (17)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}, \quad 1 \leq j \leq N, \quad 1 \leq k \leq c. \quad (18)$$

Conditions expressed in Eqs. (17) and (18) are necessary, but not sufficient. They provide means for optimizing \bar{J} via simple Picard iteration, by looping back and forth from Eq. (17) to (18) until the iterate sequence shows but small changes in successive entries of c_i or u_{ij} .

The iteration can be stopped when the following iteration termination condition is satisfied:

$$\max_{ij} \{ |u_{ij}^{k+1} - u_{ij}^k| \} < \varepsilon, \quad (19)$$

where k is the iteration number and ε is 0.00001 which satisfies a termination criterion.

The FCM algorithm consists of a series of iterations using Eqs. (17) and (18), which determine the cluster centers and membership matrix as follows:

Step 1). Definition of the number of cluster centers c . The weighting exponent $m = 2$. Choose an initial fuzzy membership matrix U , and the number of iterations k is 0.

Step 2). Calculation of the clustering center C of all samples according to Eq. (17).

Step 3). Update of the fuzzy membership matrix U by a new computed u_{ij} using Eq. (18).

Step 4). Objective function calculation. If the difference between adjacent values of the objective function is less than the iteration termination criterion ε then the iteration terminates. Otherwise, return to step 2.

In order to determine the classification accuracy of FCM, we propose an optimized FCM.

In this algorithm, we define a principle of hamming approach degree. Given the dataset $X = \{x_1, x_2, \dots, x_N\}$, $x \in R^m$, The X contain samples of c states, which has c clustering centers. $C = [c_1, c_2, \dots, c_c]^T$, $c \in R^m$. For a test dataset $T = \{t_1, t_2, \dots, t_N\}$, $t \in R^m$, define $N = [N_1, N_i, \dots, N_c]$ to be a hamming approach degree. In order to determine whether the data sample t_i belongs to the data sample category at the state i for calculating hamming approach degree N_i between the test sample t_i and the clustering center c_i under the state i :

$$N_i = 1 - \frac{1}{m} \sum_{k=1}^m |c_{ik} - t_k|, \quad i = 1, \dots, c. \quad (20)$$

For data sample t_i , find the state that has the maximum hamming approach degree, and this data sample t_i can be determined as it belongs to this state. In this paper, when the difference of the maximum hamming approach degree in 2 states is less than 0.01, that it cannot be identified or can be wrong.

3. Bearing fault diagnosis based on DTCWT-PE and optimized FCM

A method based on DTCWT, PE and optimized FCM is proposed for bearing fault diagnosis in this section. The procedure of this method is illustrated in Fig. 2 and summarized as follows:

Step 1) Perform the DTCWT for a vibration signal. In this step, each vibration signal is divided into 30 segments, and each segment contains 1024 points. According to the principle of DTCWT, each segment is decomposed into a few sub-band wavelet components. In this paper, the methods of EEMD and VMD are used for a comparison.

Step 2) For sub-band wavelet components of each segment, calculate the PE of three former wavelet sub-band components that contain more information, and it can be made up a 3×30 dimensional fault feature matrix. In this paper, DTCWT-PEs, EEMD-PEs, VMD-PEs are all calculated to make a comparison.

Step 3) Use the optimized FCM method to identify the type of fault. For vibration signals of different faults, it considers some part fault feature matrix obtained by step 2 as a data sample, using the FCM method, we calculate the clustering center of each data sample. And we regard the rest of fault feature matrix as a test data sample, calculating the hamming approach degree between cluster center of each fault to the test data. The fault is identified as a fault type when the test data and cluster center of this fault have the maximal Hamming approach degree.

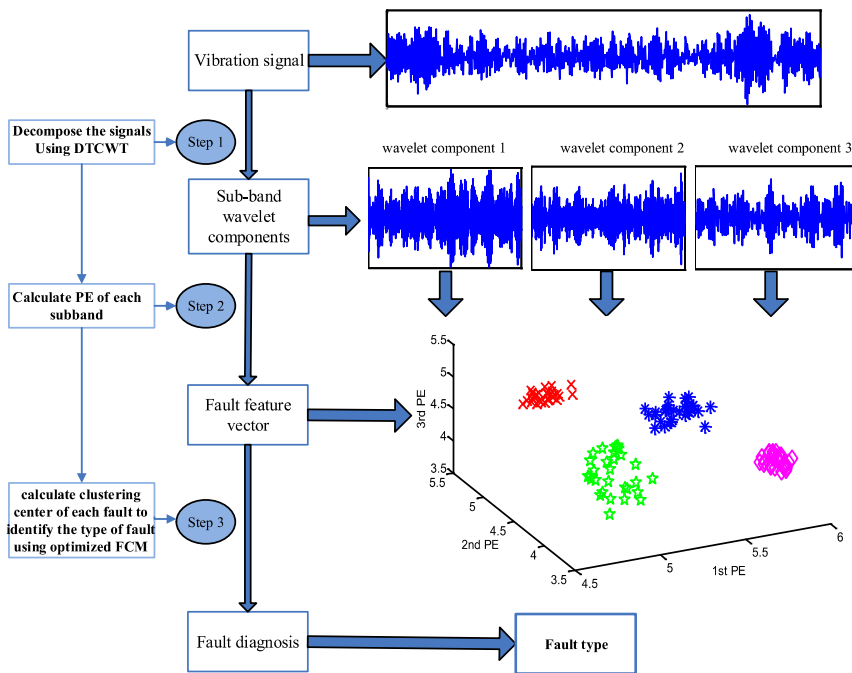


Fig. 2. Sketch of bearing fault diagnosis method

4. Experimental verification

In this study, the bearing data are obtained from the Bearing Data Center of the Case Western Reserve University [34], to verify the effectiveness of the proposed method. And the data are obtained from a fault injection test of 6205-SKF deeply grooved ball bearings. The test rig consists of a 2-horse power motor, a torque transducer, a dynamometer, and control electronics (not shown), as shown in Fig. 3(a). Vibration data is collected using accelerometers, which are attached to the housing with magnetic bases. Accelerometers were placed at the 12 o'clock position at both the drive end and fan end of the motor housing. The test bearing supports the motor shaft and motor loads of 0 to 3 horsepower (hp) (corresponding to motor speed of 1797, 1772, 1750 and 1720 rpm). Single point faults were introduced to the inner race, rolling element and outer race of the test bearing using electro-discharge machining with fault diameters of 0.1778 millimeters, 0.3556 millimeters and 0.5334 millimeters. The sampling frequency is 12 kHz for drive end

bearing experiments. Let's take 1772 rpm as an example, the time domain waveforms of vibration signals under four working states are shown in Fig. 3(b). And Fig. 3(c) gives the vibration signals of rolling element fault with different fault diameters. Because of the non-stationary and nonlinear characteristics of the vibration signals, the DTCWT method is used to decompose them into a set of wavelet components.

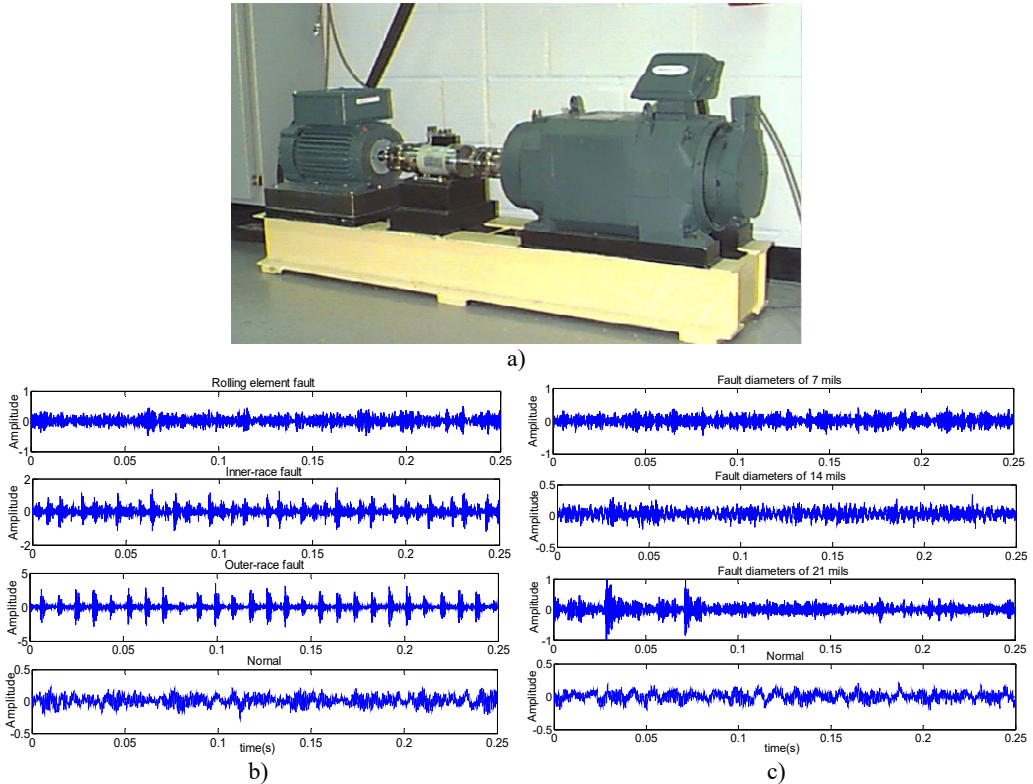


Fig. 3. a) Bearing test rig, b) vibration signals under four operating states, c) vibration signals of rolling element fault with different fault diameters

4.1. Feature extraction based on DTCWT and PE

In this section, each group of the vibration data is divided into several segments, and each segment contains 1024 points. Decompose each segment into a few wavelet components using the DTCWT. Then, calculate the PE of each signal component and regard it as the fault feature vectors.

In order to assess the advantages of the proposed approach, it is compared with several widely used feature extraction methods. In this paper, we choose DTCWT, EEMD and VMD to decompose signals into several components, and the PE groups can be calculated separately. The EEMD and VMD decompose signals into a few intrinsic mode functions (IMFs). Each IMF is regarded as a component. As the principles of DTCWT, EEMD and VMD are different, and the number of signal components is different for the same data, the number of the PEs is different similarly. But we know that there is the most dominant fault information from former mono-components, thus three former signal components can be considered and calculated. So, three PEs of each segment of vibration data can be regarded as the fault feature vectors.

Groups of feature vectors can be gained from each segment of vibration data. In this work, each vibration data is divided into 30 groups, and each group contains 1024 points. The detail of test dataset is shown in Table 1.

Table 1. Detail of test dataset

	Health state	Fault width	Motor load (operating condition)			
		Millimeters	0 hp	1 hp	2 hp	3 hp
1	Normal	–	90 grs	–	–	–
	Inner-race fault	0.1778	30 grs	30 grs	30 grs	30 grs
	Rolling element fault	0.1778	30 grs	30 grs	30 grs	30 grs
	outer-race fault	0.1778	30 grs	30 grs	30 grs	30 grs
2	Normal	–	90 grs	–	–	–
	Inner-race fault	0.3556	30 grs	30 grs	30 grs	30 grs
	Rolling element fault	0.3556	30 grs	30 grs	30 grs	30 grs
	outer-race fault	0.3556	30 grs	30 grs	30 grs	30 grs
3	Normal	–	90 grs	–	–	–
	Inner-race fault	0.5334	30 grs	30 grs	30 grs	30 grs
	Rolling element fault	0.5334	30 grs	30 grs	30 grs	30 grs
	outer-race fault	0.5334	30 grs	30 grs	30 grs	30 grs
4	Normal	–	30 grs			
	Inner-race fault	0.1778, 0.3556, 0.5334		30, 30, 30 grs		
	Rolling element fault	0.1778, 0.3556, 0.5334		30, 30, 30 grs		
	Outer-race fault	0.1778, 0.3556, 0.5334		30, 30, 30 grs		
grs: groups, hp: horsepower						

4.1.1. Fault features extraction under variable operating conditions

The fault of bearings consists of inner-race fault, rolling element fault, and outer-race fault, the motor loads changes can be made under the variable operating conditions of bearings. As the motor loads influence the vibration intensity, then they influence the extracted PEs. It may lead to a mixture of features of different fault types. So, this is necessary to confirm the performance of diagnosis under variable operating conditions.

In this part, we consider the fault features extraction of different fault types under variable operating conditions (motor loads of 0 to 3 horsepower). Datasets 1-3 from table 1 are prepared for different fault severities under different operating conditions, and each dataset contains 13 subsets. For each segment of original signal, groups of three-dimensional fault features under different operating conditions are obtained by DTCWT-PEs, EEMD-PEs and VMD-PEs. Then scatter plots of DTCWT-PEs, EEMD-PEs and VMD-PEs for different fault severities under variable operating conditions are shown in Fig. 4-6.

Fig. 4 shows the scatter plots of the fault feature extracted from dataset 1 using DTCWT-PEs, EEMD-PEs and VMD-PEs. Among the Fig. 4(a) and (b), the scatter plots of EEMD-PEs and VMD-PEs have the mixing phenomena. Features of three fault types at different operating conditions do not cluster very well, and there is a large overlap between three fault types clustering in EEMD-PEs. From the scatter plots of DTCWT-PEs in Fig. 4(c), we can see that features of different fault types at four operating conditions cluster well and are far away from the health state. So, three fault types and normal states can be separated effectively.

The scatter plots of feature vectors are obtained from dataset 2 using DTCWT-PEs, EEMD-PEs and VMD-PEs, as shown in Fig. 5. In Fig. 5(b), where the features between rolling element fault and inner-race fault obtained by VMD-PEs occur with the mixing phenomena. In Fig. 5(a) and (c), the scatter plots of the feature vectors obtained from EEMD-PEs and DTCWT-PEs are gathered well at variable operating conditions, and the separation between the fault types is clear.

Fig. 6 presents the scatter plots of feature vectors gained from dataset 3 with fault diameters equal to 0.3556 millimeters by three techniques. From Fig. 6(a), it can be seen that the fault features between rolling element fault and outer-race fault using EEMD-PEs have an overlap. And there are some mixing phenomena of feature vectors between inner-race fault and outer-race fault of VMD-PEs in Fig. 6(b). From Fig. 6(c), we can see that the features extracted by DTCWT-PEs

are obviously separable.

According to the analysis of Fig. 4-6, comparing the fault features extraction results using EEMD-PEs and VMD-PEs, we can see that the fault features extraction results of DTCWT-PEs have a great advantage. The feature vectors of different fault types at four operating conditions cluster well. At the stage of three different fault severities, different fault types and normal states are effectively separated. It indicates that the proposed fault feature extraction method of DTCWT-PEs has better performance at variable operating conditions.

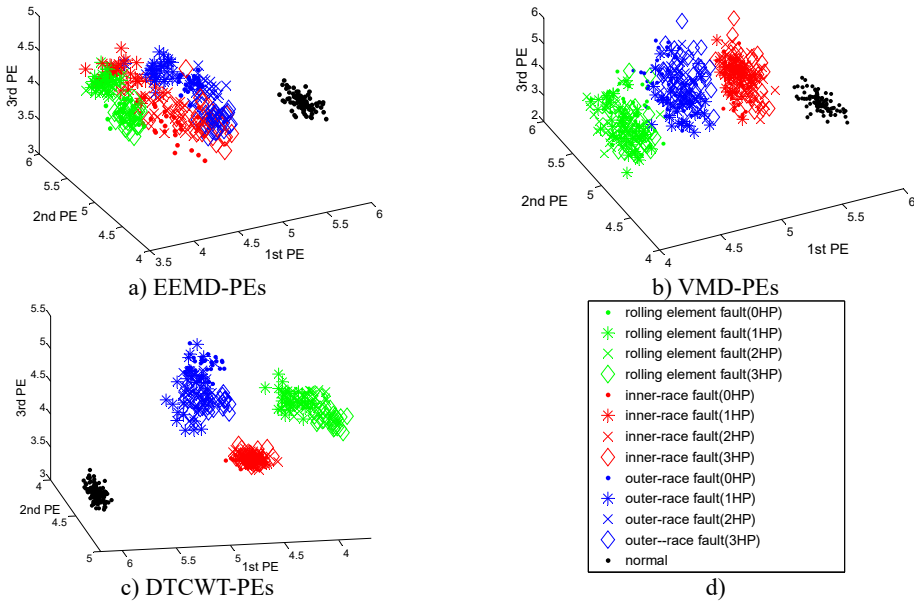


Fig. 4. Scatter plot of: a) EEMD-PEs, b) VMD-PEs and, c) DTCWT-PEs for fault feature extraction results of dataset 1 under variable operating conditions

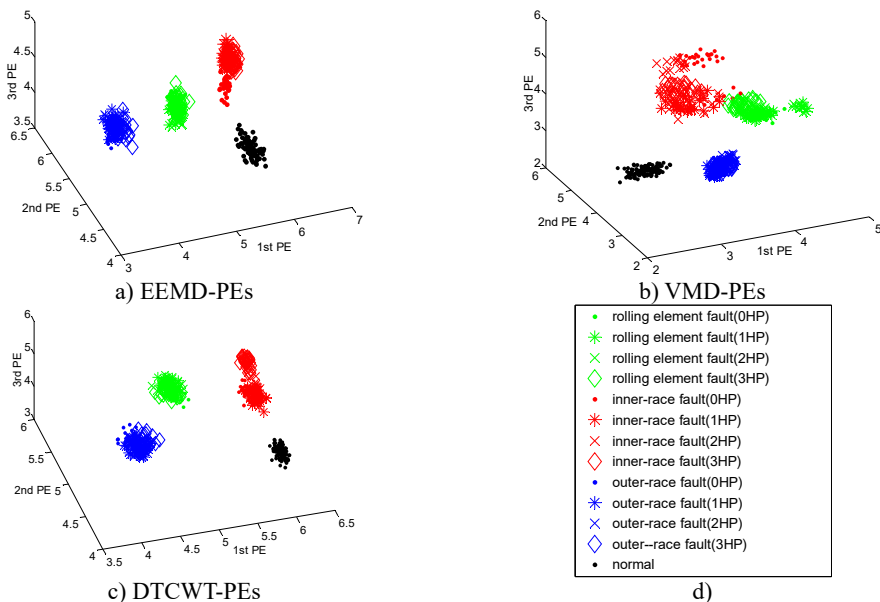


Fig. 5. Scatter plot of: a) EEMD-PEs, b) VMD-PEs and, c) DTCWT-PEs for fault feature extraction results of dataset 2 under variable operating conditions

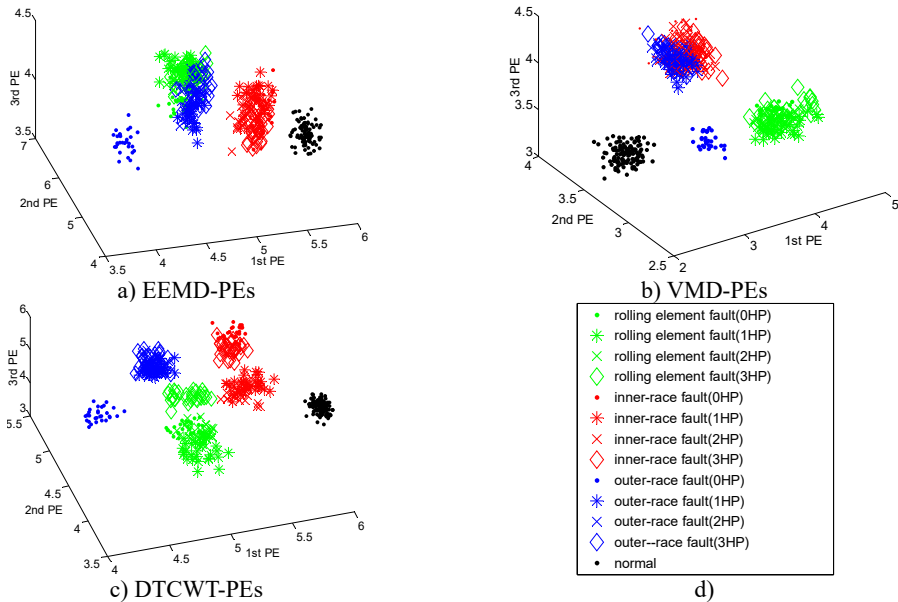


Fig. 6. Scatter plot of: a) EEMD-PEs, b) VMD-PEs and, c) DTCWT-PEs for fault feature extraction results of dataset 3 under variable operating conditions

4.1.2. Fault feature extraction for different fault severities

In order to demonstrate the feature extraction performance of DTCWT-PEs in identifying different fault severities, vibration data is analyzed at different fault severities. The detail of the extracted feature dataset with different fault severities is shown in Table 1 (dataset 4). The analyzed vibration data containing the normal state and three-faults state with fault diameters of 0.1778 millimeters, 0.3556 millimeters and 0.5334 millimeters, and we select the vibration data at the motor loads of 1 horsepower as an example. The features of normal state, inner-race fault state, rolling element fault state and outer-race fault state are extracted respectively by DTCWT-PEs, EEMD-PEs and VMD-PEs. The scatter plots of the fault features at different fault severities are shown respectively in Fig. 7-9.

Fig. 7 shows the scatter plots of the fault features extracted from the rolling element fault with different fault severities and normal state using DTCWT-PEs, EEMD-PEs and VMD-PEs where the scatter plots of the feature vectors obtained from EEMD-PEs and DTCWT-PEs are gathered well and are far away from each other. In particular, the features vectors of Fig. 7(c) are more compact, and the separation of the different states is very clear. By contrast, in Fig. 7(b), the scatter plots of VMD-PEs have some mixing between the fault severities with 0.1778 millimeters and 0.5334 millimeters.

The scatter plots of feature vectors are obtained from the inner-race fault with different fault severities and normal state, as shown in Fig. 8. In Fig. 8(a) and (b), where the features obtained by EEMD-PEs and VMD-PEs of inner-race fault between the fault severities with 0.1778 millimeters and 0.5334 millimeters, the mixing phenomena occur. From the scatter plots of DTCWT-PEs in Fig. 8(c), we can see that the features of different fault severities cluster well and are far away from the health state, three fault severities and normal states can be separated effectively.

Fig. 9 shows the scatter plots of feature vectors gained from the outer-race fault with different fault severities and normal state by three techniques. From Fig. 9(a) and (b), it can be seen that the features of four states do not cluster very well, the features extracted by EEMD-PEs and VMD-PEs are mixed for different fault severities, especially in Fig. 9(b). The scatter plots of

DTCWT-PEs are shown in Fig. 9(c). We can see that the feature vectors cluster well, and there is no mixing phenomena happened.

From the above analysis, we can determine that the feature are extracted from different fault severities by the EEMD-PEs and VMD-PEs occurred with mixing phenomena. While the features extracted from the DTCWT-PEs also have the capacity to be distinguished as different fault severities of three fault types, the features of different fault severities cluster well and are far away from each other. Therefore, the features extraction performance of DTCWT-PEs is better than that of EEMD-PEs and VMD-PEs for different fault severities.

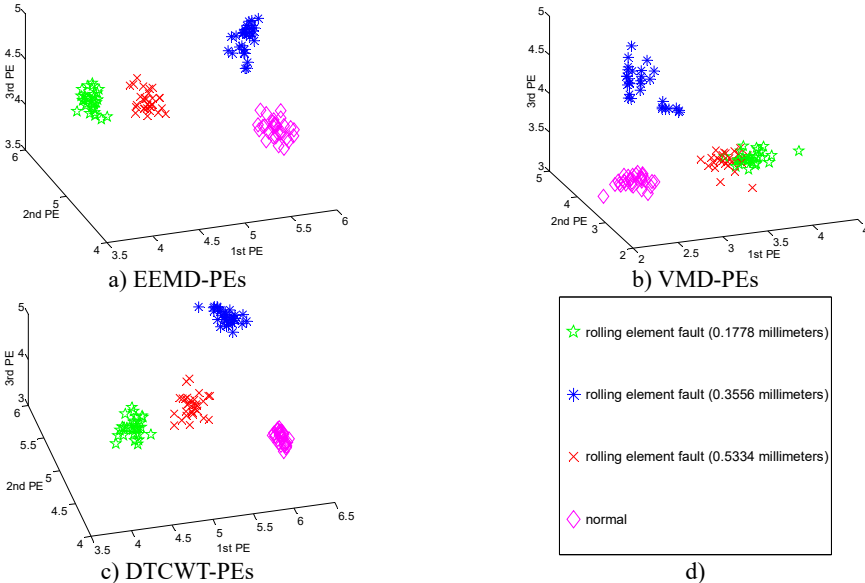


Fig. 7. Scatter plot of: a) EEMD-PEs, b) VMD-PEs and, c) DTCWT-PEs for fault feature extraction results with rolling element fault of dataset 4

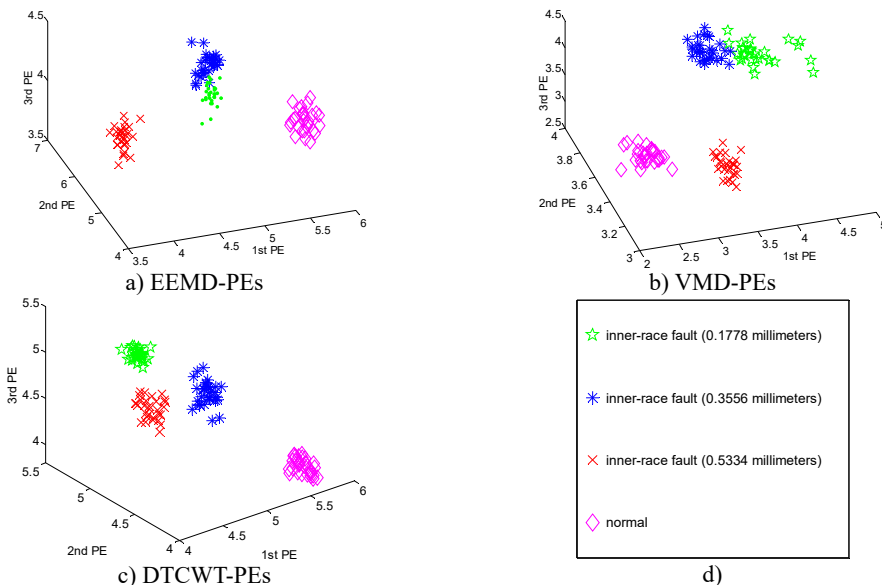


Fig. 8. Scatter plot of: a) EEMD-PEs, b) VMD-Pes and, c) DTCWT-PEs for fault feature extraction results with inner-race fault of dataset 4

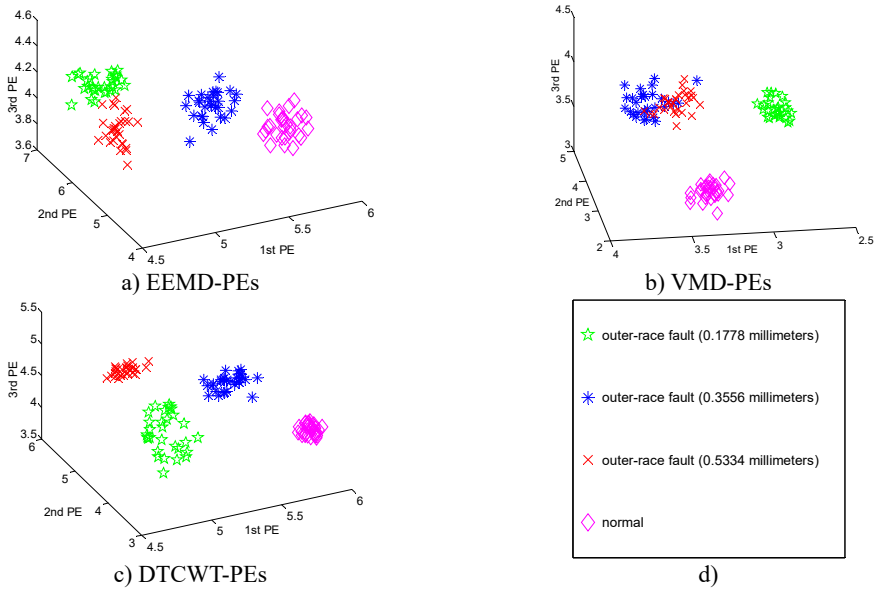


Fig. 9. Scatter plot of: a) EEMD-PEs, b) VMD-PEs and, c) DTCWT-PEs for fault feature extraction with outer-race fault results of dataset 4

4.2. Fault diagnosis by optimized FCM

In order to verify the performance of DTCWT-PEs in the fault identification, the optimized FCM is introduced to calculate the classification accuracy. In this step, we consider the diagnosis of all kinds of data under the same loads. Meanwhile, we use the fault data under known load conditions to identify the fault types of other load conditions. The detailed information of the data sample is shown in Table 2.

Table 2. Detail of dataset

	Health state	Fault width	Motor load (working condition)			
		Millimeters	0 hp	1 hp	2 hp	3 hp
1	Normal	—	40 grs	—	—	—
	Inner-race fault	0.1778	40 grs	40 grs	40 grs	40 grs
	Rolling element fault	0.1778	40 grs	40 grs	40 grs	40 grs
	Outer-race fault	0.1778	40 grs	40 grs	40 grs	40 grs

grs: groups, hp: horsepower

4.2.1. Fault diagnosis under same load

Under the motor load of 1 hp, 40 groups of data sample are extracted in turn from the normal, inner-race fault, outer-race fault and rolling element fault. Among them, the former 20 groups of four states are regarded as the known samples, using the FCM to get the standard cluster center separately, the last 20 groups of data are regarded as test samples, through the hamming approach degree calculation used to identify the fault types. The experimental results show that the fault recognition rates of extracted features using three techniques are 100 %. In order to further prove the effectiveness of the proposed method, the hamming approach degree between test samples and four known different states is calculated.

Fig. 10 shows the columnar distribution of the hamming approach degree between test samples and four different known states. From Fig. 10(a) and (b), we can see that the hamming approach degrees between test samples and corresponding state based on the EEMD-PEs and VMD-PEs are smaller than that ones based on the DTCWT-PEs. In addition, it can be seen in Fig. 10(b) that the

maximum hamming approach degree is obviously close to the second largest hamming approach degree. In Fig. 10(c), when the type of test sample is the rolling element fault, inner-race fault, outer-race fault and normal, the maximum hamming approach degree between test samples and corresponding state is close to 1, that is obviously higher than the second largest hamming approach degree. The accurate recognition and clustering abilities of DTCWT-PEs are demonstrated. Therefore, the method of DTCWT-PEs can more accurately extract the characteristics of each state, the clustering effect is better, and it is easier to classify the faults automatically.

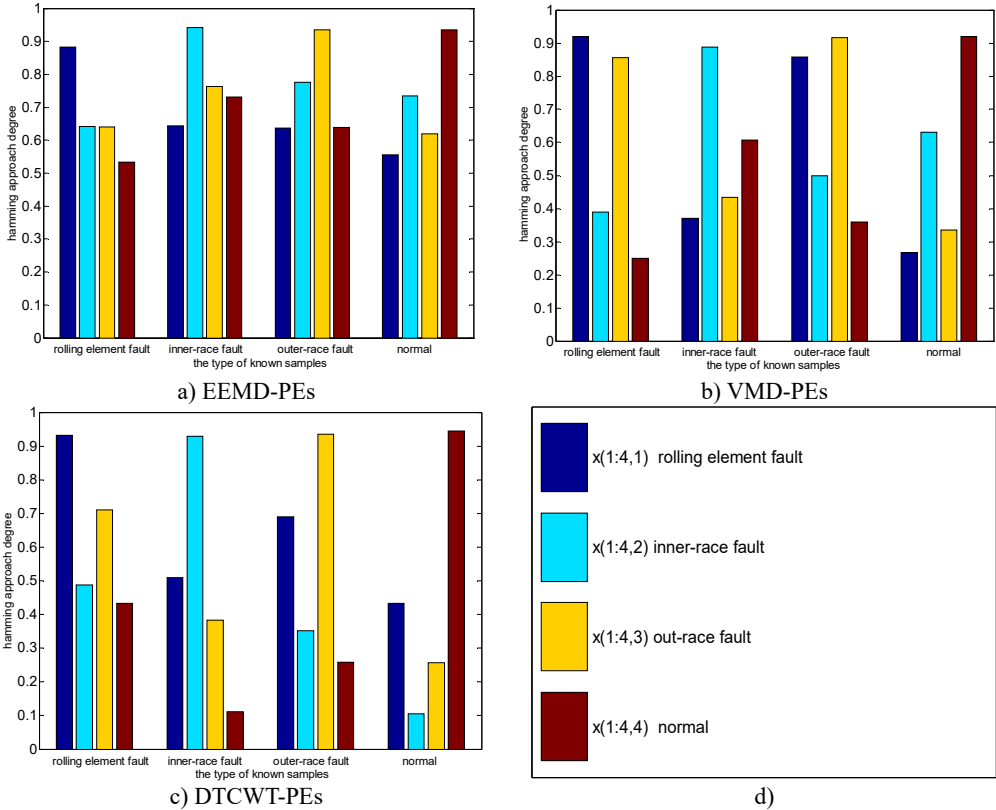


Fig. 10. Distribution of average nearness

4.2.2. Fault diagnosis under variable load

Due to that the equipment safety is relatively high in engineering applications, the small probability of failures leads to the shortage of fault samples. The fault samples can only be collected under a certain condition sometimes. Therefore, it is significant to identify the fault types under different operating conditions by using the existing fault data under this specific condition. In this paper, the fault diagnosis of other three load conditions is carried out with the help of the clustering center of existing and known fault data under load 0. Test samples are taken from the load 1, load 2 and load 3, and each of them has 40 groups, and uses the hamming approach degree between test samples and existing known fault data to identify the fault type of a test sample. The classification results show that the proposed approach can achieve a good classification effect under all loads, and the fault recognition rate is still 100 %, e.g. the fault diagnosis results of test samples under load 2 by using the clustering center of existing known fault data under load 0 are listed, as shown in Table 3. Due to the fault signal changed under different loads, in EEMD-PEs, there are 2 outer-race fault data which are mistakenly diagnosed as an inner-race fault, and 3

inner-race fault data are diagnosed as a rolling element fault in the VMD-PEs. And we can see that the DTCWT usage can be effectively identified as the four-states type under load 1 with the help of the clustering center of existing fault data under load 0.

Table 3. Fault diagnosis for load 2 by EEMD-PEs, VMD-PEs and DTCWT-PEs

		Normal	Inner-race fault	Outer-race fault	Rolling element fault
EEMD-PEs	Normal	40	0	0	0
	Inner-race fault	0	40	0	0
	Outer-race fault	0	2	38	0
	Rolling element fault	0	0	0	40
VMD-PEs	Normal	40	0	0	0
	Inner-race fault	0	37	0	3
	Outer-race fault	0	0	40	0
	Rolling element fault	0	0	0	40
DTCWT-PEs	Normal	40	0	0	0
	Inner-race fault	0	40	0	0
	Outer-race fault	0	0	40	0
	Rolling element fault	0	0	0	40

In this section, for a vibration signal under different loads, the PEs of three signal components decomposed from DTCWT-PEs are linked together, that forms the characteristic line of the test sample. Then the change trend of the characteristic line is analyzed under different loads relative to the characteristic line of the original cluster center obtained under the load of 0 hp, as shown in Fig. 11.

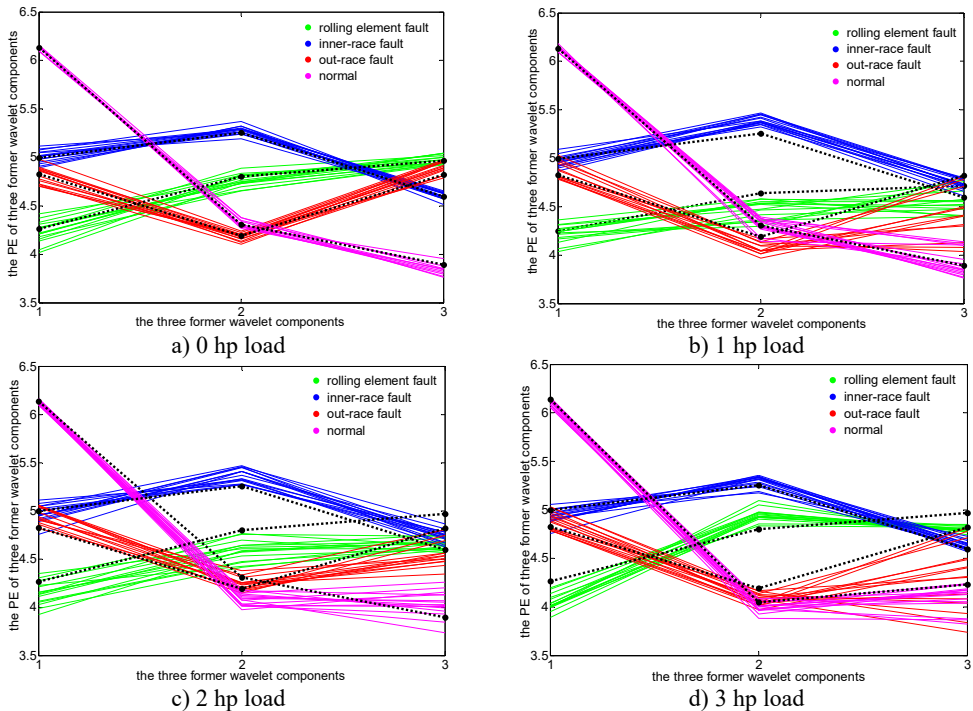


Fig. 11. Run chart of original cluster centers and characteristics line of test samples

It can be seen that, under load 0, the characteristic lines of four states closely follow the black characteristic lines of a standard clustering center, thus the accuracy of clustering center extraction is proved. Under load 2, the characteristic line of rolling element fault is deviated from the

clustering center firstly, and the characteristic line of normal state has some deviations. Under 1 hp load and 3 hp load, the characteristic lines of 4 states have a varying degree of deviation. Among them, the degree of characteristic lines of outer-race fault deviated from the clustering center is the biggest, and the normal state is the second one. Overall, when the motor load changes, the characteristic lines of a test sample under different operating conditions are all around the characteristic lines of clustering center under 0 hp load, which makes the fault recognition rate of rolling bearings be maintained at 100 %.

5. Conclusions

In this work, an approach for bearing fault identification is proposed based on the DTCWT, PE and optimized FCM. The basic idea is that the vibration signal is decomposed into several wavelet components by the DTCWT, calculating the PE of former three wavelet components. This approach regards them as the fault feature vectors, and then the algorithm of optimized FCM is introduced into the identified fault type. The experimental results verified that a more accurate classification had been obtained using DTCWT-PEs compared with EEMD-PEs and VMD-PEs. It was demonstrated that the proposed method had better ability to identify different faults types and different fault severities under different operating conditions.

It is shown that the proposed approach can accurately identify different fault severities, and has a good performance to resist variable operating conditions. Practically, the proposed feature extraction approach based on the DTCWT and PE in this paper is suitable for diagnosing a bearing fault due to that it has a strong performance of clustering and recognition. But the efficiency of this approach may be affected when the operating condition has some big changes. Therefore, a further research will mainly focus on improving the fault diagnosis capability under variable operating conditions.

Acknowledgements

This work is supported by the National Science Foundation of China (Nos. 51767022, 51575469) and the outstanding Doctor Graduate Student Innovation Project (No. XJUBSCX-2016017).

References

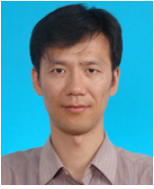
- [1] **Liu W. Y., Han J. G., Jiang J. L.** A novel ball bearing fault diagnosis approach based on auto term window method. *Measurement*, Vol. 46, Issue 10, 2013, p. 4032-4037.
- [2] **Tian Y., Ma J., Lu C., et al.** Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mechanism and Machine Theory*, Vol. 90, 2015, p. 175-186.
- [3] **Liu H., Han M.** A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mechanism and Machine Theory*, Vol. 75, 2014, p. 67-78.
- [4] **Lei Y., Lin J., He Z., et al.** Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, Vol. 25, Issue 5, 2011, p. 1738-1749.
- [5] **Lei Y., Lin J., He Z., et al.** A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, Vol. 35, Issue 1, 2013, p. 108-126.
- [6] **Bin G. F., Gao J. J., Li X. J., et al.** Early fault diagnosis of rotating machinery based on wavelet packets-empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, Vol. 27, 2012, p. 696-711.
- [7] **Cheng J., Yang Y., Yang Y.** A rotating machinery fault diagnosis method based on local mean decomposition. *Digital Signal Processing*, Vol. 22, Issue 2, 2012, p. 356-366.
- [8] **Tabrizi A., Garibaldi L., Fasana A., et al.** Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Meccanica*, Vol. 50, Issue 3, 2015, p. 865-874.

- [9] **Dragomiretskiy K., Zosso D.** Variational mode decomposition. *IEEE Transactions on Signal Processing*, Vol. 62, Issue 3, 2014, p. 531-544.
- [10] **Huang N. E., Shen Z., Long S. R., et al.** The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, Vol. 454, Issue 1971, 1998, p. 903-995.
- [11] **Smith J. S.** The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface*, Vol. 2, Issue 5, 2005, p. 443-454.
- [12] **Wu Z., Huang N. E.** Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, Vol. 1, 2009, p. 1-41.
- [13] **Ma Z. Q., Li Y. C., Liu Z., et al.** Rolling bearings fault feature extraction based on variational mode decomposition and Teager energy operator. *Journal of Vibration and Shock*, Vol. 35, Issue 13, 2016, p. 134-139.
- [14] **Cheng J., Yu D., Tang J., et al.** Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery. *Shock and Vibration*, Vol. 16, Issue 1, 2009, p. 89-98.
- [15] **Zhang S. Q., Sun G. X., Li L., et al.** Study on mechanical fault diagnosis method based on LMD approximate entropy and fuzzy C-means clustering. *Chinese Journal of Scientific Instrument*, Vol. 34, Issue 3, 2013, p. 714-720.
- [16] **Zhang X., Liang Y., Zhou J.** A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, Vol. 69, 2015, p. 164-179.
- [17] **Yang Y., Cheng J., Zhang K.** An ensemble local means decomposition method and its application to local rub-impact fault diagnosis of the rotor systems. *Measurement*, Vol. 45, Issue 3, 2012, p. 561-570.
- [18] **Tang G. J., Wang X. L.** Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing. *Journal of Xi'an Jiaotong University*, Vol. 49, Issue 5, 2015, p. 73-81.
- [19] **Yan R., Gao R. X., Chen X.** Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Processing*, Vol. 96, 2014, p. 1-15.
- [20] **Djebala A., Ouelaa N., Hamzaoui N.** Detection of rolling bearing defects using discrete wavelet analysis. *Meccanica*, Vol. 43, Issue 3, 2008, p. 339-348.
- [21] **Li Z., He Z., Zi Y., et al.** Rotating machinery fault diagnosis using signal-adapted lifting scheme. *Mechanical Systems and Signal Processing*, Vol. 22, Issue 3, 2008, p. 542-556.
- [22] **Selesnick I. W.** The design of approximate Hilbert transform pairs of wavelet bases. *IEEE Transactions on Signal Processing*, Vol. 50, Issue 5, 2002, p. 1144-1152.
- [23] **Kingsbury N. G.** The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. *IEEE Digital Signal Processing Workshop*, Bryce Canyon, Vol. 86, 1998, p. 120-131.
- [24] **Selesnick I. W., Baraniuk R. G., Kingsbury N. C.** The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, Vol. 22, Issue 6, 2005, p. 123-151.
- [25] **Lo E. H. S., Pickering M. R., Frater M. R., et al.** Image segmentation from scale and rotation invariant texture features from the double dyadic dual-tree complex wavelet transform. *Image and Vision Computing*, Vol. 29, Issue 1, 2011, p. 15-28.
- [26] **Wang Y., He Z., Zi Y.** Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform. *Mechanical Systems and Signal Processing*, Vol. 24, Issue 1, 2010, p. 119-137.
- [27] **Wang F. T., Chen S. H., Yan D. W., et al.** Noise reduction based on dual tree complex wavelet transform unfolding and its application in fault diagnosis. *Journal of Mechanical Engineering*, Vol. 50, Issue 21, 2014, p. 159-163.
- [28] **Bandt C., Pompe B.** Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, Vol. 88, Issue 17, 2002, p. 174102.
- [29] **Zanin M., Zunino L., Rosso O. A., et al.** Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy*, Vol. 14, Issue 8, 2012, p. 1553-1577.
- [30] **Bezdek J. C.** *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer Science and Business Media, 2013.
- [31] **Gou M. F., Xu L. L., Liao X. R., et al.** A vibration signal feature extraction method for distribution switches based on singular value decomposition of time-frequency matrix. *Proceedings of the CSEE*, Vol. 34, Issue 28, 2014, p. 4990-4997.

- [32] **Jahromi A. T., Er M. J., Li X., et al.** Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis. *Neurocomputing*, Vol. 196, 2016, p. 31-41.
- [33] **Kingsbury N.** Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, Vol. 10, Issue 3, 2001, p. 234-253.
- [34] <http://www.eecs.case.edu/laboratory/bearing>.



Ping Ma currently is a doctoral student at the School of Electrical Engineering, Xinjiang University, Urumqi, China. Her current research interests include control, dynamics and fault diagnosis.



Hongli Zhang received the Ph.D. degree from Beijing Institute of Technology in 2009. He is currently an Professor in Xinjiang University, Urumqi, China. His current research interests include control, dynamics and fault diagnosis.



Wenhui Fan received the Ph.D. degree from Zhejiang University in 1998. He is currently an Professor in Tsinghua University, Beijing, China. His current research interests include control, dynamics and fault diagnosis.



Cong Wang currently a Ph.D. candidate at Xinjiang University, Urumqi, China. Her research interests include machine learning, intelligent computing, and system modeling and control.