Fault diagnosis method of rolling bearing based on EMD-Hilbert envelope spectrum and BPNN

Tao Wu*

School of Environmental and Material Engineering, Shanghai Polytechnic University, Shanghai Pudong, China

*Corresponding author: taowu@sspu.edu.cn

Abstract. Rolling bearings are an important component of rotating machinery, and accurate diagnosis of their faults is also very important. This paper proposes a rolling bearing fault diagnosis method combining empirical mode decomposition (EMD)-Hilbert envelope spectrum analysis and BP neural network (BPNN). First, EMD is used to decompose the intrinsic modal function (IMF) component containing the bearing fault feature information from the original vibration signal data of the rolling bearing, and then the IMF component is processed in combination with the Hilbert envelope analysis method to obtain a clear fault feature frequency. The processed fault characteristic frequency is subjected to dimensionality reduction by principal component analysis (PCA) to propose redundant data. Finally, the dimensionality-reduced feature data is input into BPNN to establish a fault diagnosis model. The results show that the rolling bearing fault diagnosis method based on EMD-Hilbert envelope spectrum analysis and BPNN can effectively identify different fault states of rolling bearings.

1. Introduction
Rolling bearings are important parts in rotating machinery and the most vulnerable parts. Rolling bearing failure will have a very large impact on the equipment [1]. Therefore, it is of great significance to study the fault diagnosis technology of rolling bearings [2]. Commonly used diagnostic methods for rolling bearing faults include empirical mode decomposition (EMD) method, short-time Fourier method, wavelet analysis method, etc. Huang [3] and others proposed EMD on non-stationary signals in 1998 and divided them into a series of intrinsic modal functions (IMF), and then obtained the entire signal spectrum by Hilbert transform. Rao [4] proposed a rolling bearing fault diagnosis method combining empirical mode decomposition (EMD) and Hilbert envelope spectrum analysis, which can effectively identify the inner ring faults of rolling bearings. Hu [5] proposed a rolling bearing fault recognition method that combines the Hilbert-Huang Transform (HHT), Levenberg-Marquardt (LM) algorithm and BP neural network. This method can effectively extract the fault feature information of the bearing, and can accurately identify different faults. Huang [6] proposed a bearing fault diagnosis method based on singular value decomposition (SVD), ensemble empirical mode decomposition (EEMD) and BP neural network. This method can effectively identify the fault type of rolling bearings and can be used for bearing fault diagnosis. This paper uses EMD-Hilbert envelope spectrum, PCA and BP neural network to diagnose the rolling bearing fault condition. After processing by EMD, the IMF component is obtained, and the envelope spectrum is obtained by applying Hilbert transform. Then use PCA method to reduce the
dimension of the signal. Finally, the characteristic parameters are used as the input of the BP neural network, and the BP neural network is used to classify the vibration signal of the bearing.

2. Introduction to algorithm principles

2.1. EMD

The essence of the EMD method is the smoothing of the signal, especially for the processing of nonlinear and non-stationary signals [7]. This method can decompose a complex nonlinear signal into a finite number of IMF components. Each IMF component must meet two conditions: First, the number of poles and zeros in the entire signal time domain are equal, or the difference is at most 1; The second is to arbitrarily take a point on the signal, the average value of the envelope determined by the local maximum and the envelope determined by the local minimum is zero.

When EMD is applied to any signal, all the maximum and minimum points on the original signal \( x(t) \) are respectively fitted with cubic spline functions, and the resulting two function fitting curves are used as the original signal \( x(t) \) the upper and lower envelopes. Calculate their average value and record it as \( m^1(t) \). Subtract the original signal \( x(t) \) and \( m^1(t) \) to get a new signal, denoted \( h^1(t) \):

\[
h^1(t) = x(t) - m^1(t) \tag{1}
\]

If the signal \( h^1(t) \) does not satisfy the two conditions of the IMF, then the above steps need to be repeated with \( h^1(t) \) as the original signal. Filter \( k \) times until the signal \( h^1_k(t) \) meets the two conditions of the IMF component, and the signal \( h^1_k(t) \) becomes the first IMF component:

\[
h^1_k(t) = h^1_{k-1}(t) - m^1_k(t) \tag{2}
\]

Where: \( h^1_{k-1}(t) \) is the signal after screening \( k-1 \) times; \( m^1_k(t) \) is the average of the upper and lower envelope functions after screening \( k \) times.

The first order IMF component is decomposed from the original signal, which is denoted as \( c_1(t) \):

\[
c_1(t) = h^1_1(t) \tag{3}
\]

Subtract \( c_1(t) \) from the original signal \( x(t) \) to get the first-order residual signal, denoted \( r_1(t) \):

\[
r_1(t) = x(t) - c_1(t) \tag{4}
\]

Then take \( r_1(t) \) as the original signal and recalculate according to equations (1) to (3) to obtain the second IMF component \( c_2(t) \).

Repeat the above steps, you can get \( c_2(t) \), \( c_3(t) \), …, \( c_n(t) \), until \( r_n(t) \) can no longer be decomposed or meet the given termination conditions.

Usually \( r_n(t) \) becomes a monotone residual function and exits the loop, so there are:

\[
x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{5}
\]

In the formula: \( r_n(t) \) represents the average trend of the signal.
2.2. Hilbert envelope spectrum

The EMD method decomposes several IMF components based on the local characteristic time scale of the signal, and can calculate the instantaneous frequency and instantaneous amplitude of each IMF component [8]. Make Hilbert transformation for each $c_i(t)$ needed as:

$$H[c_i(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t - \tau} d\tau \tag{6}$$

Construct the analytical signal as:

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t)e^{j\phi_i(t)} \tag{7}$$

Obtain the corresponding amplitude function of $c_i(t)$, namely Hilbert envelope spectrum $a_i(t)$ and phase function $\phi_i(t)$, as:

$$a_i(t) = \sqrt{c_i^2(t) + H^2[c_i(t)]} \tag{8}$$

$$\phi_i(t) = \arctan \frac{H[c_i(t)]}{c_i(t)} \tag{9}$$

It can be known from the above principle that the IMF component decomposed by the EMD method can be amplitude or frequency modulated. By adjusting the envelope signal and refining the spectrum analysis, the rubbing fault information can be effectively extracted.

2.3. BPNN diagnosis

BP neural network is a multi-layer feedforward neural network. It is generally composed of three parts: input layer, hidden layer and output layer, as shown in Figure 1. The adjacent layers are connected by interconnection, there is no connection between neurons in the same layer, and there is no direct connection between the output layer and the input layer [9].

![Figure 1. BP neural network structure.](image-url)
This paper uses a 3-layer BP neural network structure, the first layer of the network is the input layer, the number of input layers depends on the number of feature vectors; the second layer of the network is the hidden layer. There is no unified principle for the selection of the number of hidden layer nodes. Generally, it is selected based on experiments and experience; the third layer of the network is the output layer, and the number of neurons in the output of the neural network depends on the number of failure modes. In the process of creating a neural network, the input samples need to be normalized. Since the range after normalization is in the interval \( [0,1] \), the maximum value of the input of the designed neural network is 1, and the minimum value is 0. In addition, the transfer function and training function between the output layer and the hidden layer need to be determined.

3. Case analysis

3.1. Experimental description

In order to verify the effectiveness of the method proposed in this paper, the experimental data of the Western Reserve University Bearing Data Center was used to verify [10]. The test bearings are installed at the drive end and the fan end respectively. The test bearings are SKF6205-2RSJEM deep groove ball bearings with a sampling frequency of 12 kHz. The data of 10 different states including normal, outer ring fault, inner ring fault, rolling element fault, and fault states with different degrees of damage were selected for experimental verification. Bearing data is shown in Table 1.

| Bearing status       | Failure level (inch) | Abbreviation |
|----------------------|----------------------|--------------|
| Normal               | 0                    | Normal       |
| Inner ring failure   | 0.007                | IR07         |
|                      | 0.014                | IR14         |
|                      | 0.027                | IR21         |
| Outer ring failure   | 0.007                | OR07         |
|                      | 0.014                | OR14         |
|                      | 0.027                | OR21         |
| Rolling element failure | 0.007              | B07          |
|                      | 0.014                | B14          |
|                      | 0.027                | B21          |

3.2. Signal processing

![Figure 2. Original signal.](image-url)
Figure 3. Time domain decomposition diagram.

Figure 4. Envelope spectrum.
Figure 2 shows the outer ring bearing with a fault diameter of 0.5334 mm. Figure 3 shows the signal diagram of the time domain signal after VMD decomposition. Figure 4 shows the envelope spectrum after VMD decomposition. It can be seen from Figure 3 that the features after the envelope spectrum have been able to distinguish their octave well, which also provides good preparation for the subsequent fault diagnosis.

3.3. Fault diagnosis results
From the original data, each sample is sampled with 2048 points, 100 samples for each fault, and the original signal is decomposed into multiple inherent modal components using EMD. In this paper, the envelope spectrum corresponding to the first 5 IMFs is taken as the fault feature. Since there are a total of 10 kinds of faulty rolling bearings, there are a total of 1,000 sets of data. Therefore, the fault characteristic data finally obtained is 1000 * 5120. Then, the dimension of the extracted feature data is reduced. In this paper, PCA is used to reduce the dimension of the feature data, eliminating redundant data, and finally reduced to 100 dimensions. There are 700 training sets and 300 test sets. Figures 5 and 6 are the results of fault diagnosis using BPNN in the training set and test set, respectively. It can be seen from the figures that the accuracy of the training set is 79.71%, and the accuracy of the test set is 71.67%.

Comparison of BP neural network output and expected output - training set

![Comparison of BP neural network output and expected output - training set](image_url)

**Figure 5.** Training set fault diagnosis results.
4. Conclusion
This paper combines the EMD-Hilbert envelope spectrum, PCA and BP neural network to study the fault diagnosis method of rolling bearings. In this method, the nonlinear and non-stationary vibration signals are processed by EMD, the IMF components reflecting the original signals are decomposed, and the envelope spectrum is obtained by applying the Hilbert transform. Then use PCA method to reduce the dimension of the signal. Finally, the characteristic parameters are used as the input of the BP neural network, and the BP neural network is used to classify the bearing vibration signals. The energy characteristics of the IMF component obtained by EMD decomposition of the original vibration signal are combined with the energy characteristics of the Hilbert marginal spectrum region, and the effectiveness of the method is proved by an example. Through the results of the test set and the training set, it can be found that this method can more effectively obtain the fault information of the rolling bearing. Ten fault states can be effectively distinguished. Although the accuracy is not very high, there are still many areas for improvement. In future work, the feature extraction will use energy entropy to extract fault features. In terms of fault diagnosis, we will optimize the BP neural network to improve its running speed and accuracy.

Acknowledgments
This work is supported by Shanghai Polytechnic University Graduate Key Project Fund (Fund No. EG D19YJ0024).

References
[1] W.J. Min, G.P. Li, T.P. Han, S.T. Xiang, W.F. Lai, Fault diagnosis of rolling bearings based on EEMD energy moment and improved quantum particle swarm neural network. Journal of Ni
[2] C. Liu, Y.X. Wang, J.W. Yang, Application of variational modal decomposition based on FOA in bearing fault diagnosis. Mechanical Transmission, 44 (05) (2020) 146-154.

[3] R. Gu, J. Chen, R.J. Hong, Y.B. Pan, Y.Y. Li, Weak fault diagnosis of rolling bearings based on improved adaptive variational mode decomposition. Vibration and Shock, 39 (08) (2020) 1-7 + 22.

[4] Z.R. Rao, Y. Hu, Research on fault diagnosis of rolling bearing based on EMD and Hilbert envelope spectrum analysis. Equipment Machinery, (02) (2019) 58-61.

[5] Z. Hu, Z.B. Zhang, X.J. Wang, Y.C. Wu, X.R. Xie, Fault diagnosis of rolling bearing based on Hilbert-Huang transform and neural network. Power Tools, (01) (2020) 11-18.

[6] J.N. Huang, S.H. Wang, C. Ma, Fault diagnosis of rolling bearings based on SVD-EEMD and BP neural network. Journal of Beijing Information Science and Technology University (Natural Science Edition), 34 (02) (2019) 69-74.

[7] Y.T. Ai, Y. Fang, J. Tian, Extraction of rolling bearing fault features based on the combination of kurtosis criterion EMD and spatial correlation. Mechanical Design and Manufacturing, (12) (2019) 213-216.

[8] M.Y. Wang, Planetary gear fault monitoring based on wavelet and Hilbert transform. Mechanical Design and Manufacturing Engineering, 48 (05) (2019) 105-108.

[9] Y.F. Feng, H.Q. Lu, H. Yin, L. Cao, Research on fault diagnosis model based on BP neural network. Computer Engineering and Applications, 55 (06) (2019) 24-30.

[10] Bearing Data Center. Case Western Reserve University, Cleveland, OH. [EB/OL]. (2010), (2018-10-20). Available: http://www.eecs.case.edu/laboratory/Bearing.