Rolling Bearing Fault Identification Based on Variational Mode Decomposition and Machine Learning

Junhua Zhang [a], Sha Xu [b]

Department of Mechanical Engineering, North China Electric Power University, Baoding, 071003, China

[a]15165928173@163.com,[b]xusha889@163.com

Abstract

Aiming at the difficulty of extracting fault characteristics of rolling bearings and the low accuracy of fault recognition rate, a bearing fault identification method based on variational mode decomposition (VMD) fuzzy entropy (FE) and extreme learning machine (ELM) is proposed. First, the vibration signal of the rolling bearing is subjected to VMD decomposition to obtain a number of intrinsic mode function (IMF). Then, the IMF is correlated with the original signal, and the first four IMF with main feature information are selected, and the FE of the selected IMF component is used as the feature vector.Finally, the feature vector is input into the ELM classifier for identification. Compared with Empirical mode decomposition (EMD) fuzzy entropy and VMD approximate entropy (AE), the results show that the proposed method can effectively improve the diagnostic accuracy of bearing faults.

Keywords

Variational mode decomposition; fuzzy entropy; ELM recognition; fault diagnosis.

1. Introduction

The key to the fault diagnosis of rolling bearings is to extract the fault features from the collected vibration signals, so as to accurately classify the faults. The vibration signal of the bearing is decomposed by the EMD, and the singular value of the IMF is extracted as the fault feature for bearing fault diagnosis[1]. The EMD and nonlinear dynamic parameter entropy are combined to extract the fault characteristics of rolling bearing. The fault diagnosis of rolling bearing is realized by support vector machine[2]. The vibration signal of the cylindrical roller bearing is weak, and the fault feature is extracted by wavelet packet approximation entropy. Fault identification of cylindrical roller bearings is realized by support vector machine[3]; fault feature extraction is performed by using eigentime scale decomposition fuzzy entropy, fault diagnosis of rolling bearings is carried out by Gath-Geva (GG) clustering and FE, sample entropy (SE) and approximation are obtained. Entropy contrast, it is proved that FE can better characterize the fault signal[4].

FE is a method to measure the complexity of time series based on the concepts of AE and SE. It has a good effect when analyzing nonlinear and non-stationary signals[5]. Artificial neural network (ANN) is a classic classifier in fault diagnosis. Most of the research work uses this classifier for fault identification[6], but the structure of artificial neural network is not fixed, the convergence speed is not fast and easy. Over-fitting occurs, which causes the method to run slowly and the classification accuracy needs to be improved. Support vector machine (SVM) is a classifier for small samples and nonlinear problems. It has been effectively promoted in fault diagnosis[7], but this method requires preset parameters and slow calculation speed. Disadvantages. The ELM is a learning algorithm for a class of single hidden layer feedforward neural networks. By randomly selecting the input weights during training, the earth improves the convergence speed and generalization performance of the network. ELM applied to fault diagnosis[8,9].

2. Introduction to the basic principles

2.1 VMD method

VMD achieves adaptive decomposition of signals by placing the signal decomposition process within the variational framework and searching for the optimal solution of the constrained variational model. The non-stationary signal y can be decomposed into a series of component signals $\{y_k\}$, $k=1,2,\cdots,K$ of limited bandwidth, and the center frequency ω_k of y_k can be determined during the decomposition process.

The component signal bandwidth is determined by first calculating the corresponding analytical signal of y_k by Hilbert transform, then shifting the signal spectrum to the base frequency band by multiplying function $e^{-j\omega_k t}$, and finally estimating the bandwidth of the frequency shifted signal by Gaussian smoothing method, and its corresponding constraint. The variational model is as follows:

$$\min_{\{y_k\},\{\omega_k\}} \left(\sum_{k=1}^{K} \partial_t ((\sigma(t)+\frac{j}{\pi t}) y_k(t)) e^{-i\omega_k t} \|_2^2\right)$$

$$s.t. \sum_{k=1}^{K} y_k = y \tag{1}$$

Wherein $\{y_k\}$ and $\{\omega_k\}$ respectively represent a set of component signals and their corresponding center frequencies; K is the number of components.

The equation (1) is transformed into an unconstrained optimization model using the second penalty term and the Lagrange multiplier, which is expressed as follows:

$$\mathcal{L}(\{y_k\}\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \partial_t ((\sigma(t)+\frac{j}{\pi t}) y_k(t)) e^{-i\omega_k t} \|_2^2 +$$

$$y(t) - \sum_{k=1}^{K} y_k(t) \|_2^2 + \left\langle \lambda(t), y(t) - \sum_{k=1}^{K} y_k(t) \right\rangle \tag{2}$$

In the formula, α is the disciplinary parameter, this paper takes 2000; λ is the Lagrange multiplier.

The model of equation (1) is transformed into an optimization model of equation (2). Then use the alternating multiplier algorithm to find the optimal solution for equation (2), and the results are as follows:

$$\hat{y}_k^{n+1}(\omega) = \frac{\hat{y}(\omega) - \sum_{i\neq k} \hat{y}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1+2\alpha\left(\omega-\omega_k^n\right)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left|\hat{y}_k^n(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{y}_k^n(\omega)\right|^2 d\omega} \tag{4}$$

In the formula, $\hat{y}_k^{n+1}(\omega)$, $\hat{y}(\omega)$, $\hat{y}_i^n(\omega)$, $\hat{\lambda}^n(\omega)$ are respectively Fourier transforms of $y_k^{n+1}(t)$, $y(t)$, $y_i^n(t)$, $\lambda^n(t)$; n is the number of iterations. The specific implementation process is as follows.

Initialize $\{\hat{y}_k^1\}$, $\{\hat{\omega}_k^1\}$, $\hat{\lambda}^1$ and n to 1;

Update $\hat{y}_k^{n+1}(\omega)$, ω_k^{n+1} according to equations (3) and (4);

Updated according to formula (5);

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{y}(\omega) + \sum_{k=1}^{K} \hat{y}_k^{n+1}(\omega)) \tag{5}$$

It is judged whether the convergence condition is satisfied, and if it is satisfied, the decomposition process ends; otherwise, the number of iterations is increased by 1, and the process returns to step 2).

$$\sum_{k=1}^{K} \| y_k^{n+1} - \hat{y}_k^n \|_2^2 / \| \hat{y}_k^n \|_2^2 < \varepsilon \tag{6}$$

2.2 Fuzzy entropy

Fuzzy entropy has the advantages of stability and continuity, and small dependence on data length. The definition is as follows:

Construct a m-dimensional vector in order for the N-point time series $\{\mathrm{u(i)} : 1 \le \mathrm{i} \le \mathrm{N}\}$:

$$X_i^m = (u(i) - u_0(i), u(i+1) - u_0(i), \cdots, u(i+m\text{-}1)\text{-}u_0(i)) \tag{7}$$

$$u_0(i) = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j) \tag{8}$$

The selection of m is determined according to the length of the data, and m=2 is generally selected.

Calculate the distance between any two time series, and define d_{ij}^m as the distance between X_i^m and X_j^m.

$$d_{ij}^m = \max_{k\in(0,m-1)} (|u(i+k) - u(i) - (u(j+k) - u(j))|) \quad i, j = 1,2,\cdots, N-m, i \ne j .$$

Through the fuzzy function $\mu(d_{ij}^m, n, r) = \exp(-(d_{ij}^m / r)^n)$, where n and r are the similarity calculation parameters, this paper takes $n = 2$, $r = 0.2 \times SD$,, where SD is the standard deviation of the original data.

Defining function:

$$\phi^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m-1}\sum_{j=1, j=i}^{N-m} D_{ij}^{m+1})$$

Similarly, in the dimension m+1, repeat the above steps 1-4,

$$\phi^{m+1}(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}(\frac{1}{N-m-1}\sum_{j=1, j=i}^{N-m} D_{ij}^{m+1}) \tag{9}$$

Define the FE as:

$$FuzzyEn(m,n,r) = \ln \phi^m(n,r) - \ln \phi^{m+1}(n,r)$$

2.3 Extreme learning machine

The connection weight of the input layer of the ELM algorithm and the threshold of the hidden layer neurons are randomly generated. In the training process, only the number of hidden layers needs to be set, that is, the global optimal solution can be obtained, so the network learning speed can be obviously improved. And generalization ability.

Let the number of input and output data be Q sets.

$$\theta = \{(x_i, y_i), i = 1,2,\cdots, Q, x_i \in R^n, y_i \in R^m\} \tag{10}$$

Where: x_i is the i-th input data, and y_i is the corresponding output data. Assuming the number of hidden layer neurons is L, the ELM network output is:

$$f(x_i) = \sum_{j=1}^{L} \beta_j g(\omega_j \bullet x_i + b_j) \qquad i = 1,2,\cdots, Q \tag{11}$$

In the middle: $\omega_j = [\omega_{1j}, \omega_{2j}, \cdots, \omega_{xj}]^T and \beta_j = [\beta_{1j}, \beta_{2j}, \cdots, \beta_{xj}]^T$ respectively represent the connection weight vector between the input node and the hidden layer neuron, the hidden layer neuron and the output node; b_j is the threshold of the j-th hidden layer neuron; $g(x)$ is the hidden layer Neuron activation function.

Simplify equation (11) to the following linear equations:

$$H\beta = Y \tag{12}$$

If the hidden layer neurons have the same number as the training data, the single hidden layer feedforward neural network can approximate the training data with zero error, and does not care about the values of ω and b . However, if the number of training data Q is large, the number of neurons in the hidden layer is usually smaller than Q.When the activation function $g(x)$ is infinitely divisible, it is not necessary to adjust all parameters of the single implicit layer feedforward neural network. ω and b are arbitrarily chosen and unchanged during training.

By obtaining the least squares solution of the system of equations, the connection weight β of the hidden layer and the output layer can be calculated, that is $\min_{\beta} \| H\beta = Y \|$, and its solution is $\hat{\beta} = \| H^{+}Y \|$, where H^{+} is the $Moor - Penrose$ generalized inverse of the hidden layer output matrix H .

3. Experiment analysis

Experimental data was derived from rolling bearing experimental data from the Electrical Engineering Laboratory at Case Westen Reserve University. Its bearing partial failure is a single point of failure manually processed by EDM technology.The fault code and fault category are shown in the Table 1.

Table 1 Fault codes and fault classification of bearing

Error code	Fault category[μm]	Remarks
F1	Normal status	6 o'clock
F2	Inner ring failure177.8	
F3	Inner ring failure355.6	
F4	Rolling element failure177.8	
F5	Rolling element failure355.6	
F6	Outer ring fault177.8	
F7	Outer ring fault355.6	

Its time domain waveform diagram is shown in Fig. 1.

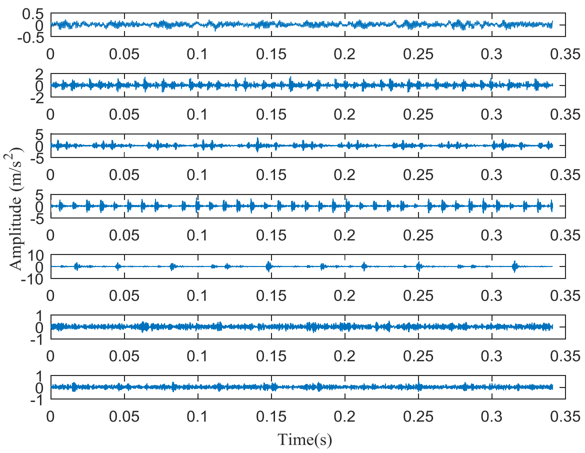


Fig. 1 Time domain waveform

The fault diameters are 177.8 μm and 355.6 μm, and the data sampling frequency is 12 k Hz. Each type takes 40 sets of data samples, and the sample length is N=2048. Taking the inner ring fault signal $x(t)$ as an example, the signal is decomposed by VMD. The center frequencies corresponding to different K values are shown in Table 2.

Table. 2 Center frequency corresponding to different K

Modal number	Center frequency[Hz]					
2	1334.4	3486.0	—	—	—	—
3	1280.4	2582.4	3573.6	—	—	—
4	608.4	1351.2	2588.4	3574.8	—	—
5	606.0	1356.0	2564.4	3381.6	3644.4	—
6	602.4	1348.8	2488.8	2836.8	3414.0	3648

Among them, starting from K is 5, there is a mode with similar center frequency. This paper thinks that over-decomposition occurs. Therefore, the number of modes is chosen to be 4, and the decomposition result is shown in Fig. 2.

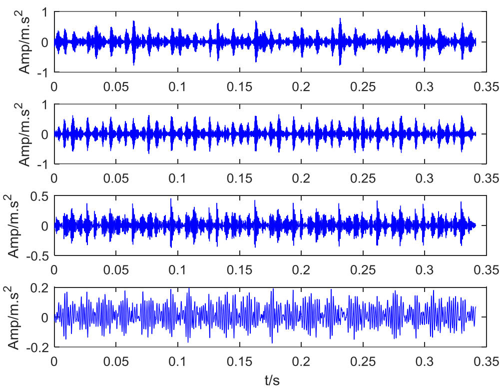


Fig. 2 Results of decomposition of VMD

Calculate the mutual information MI_i of each modal signal and the original signal, and normalize it.

$$\beta = MI_i / \max(MI_i) \tag{13}$$

The β threshold is taken as 0.02[10]. The results are shown in Table 3.

Table. 3 Mutual information between each mode and the original signal

Algorithm	Modal 1	Modal 2	Modal 3	Modal 4
VMD	0.3127	0.3353	0.1557	0.0657
EMD	0.6833	0.1166	0.0480	0.0107

Using the same method as above, other fault signals are decomposed, and finally K=4 in the VMD is determined, and the first 3 components with larger mutual information are used in the EMD for feature extraction.

The VMD decomposition is performed on the 7 kinds of data, and the first 4 components are taken to form 7 sets of IMF component data, and the dimension of each group is 2048×4×40. By obtaining the FE of the 7 sets of IMF component data, 7 sets of 4×40 fuzzy entropy are obtained, and the normalized mean is shown in Table 4.

In the ELM algorithm of this paper, the number of hidden nodes is set to 20, and the activation function applies the Sigmoid function. 20 sets of data were randomly selected to train the ELM to establish a fault diagnosis model, and the established diagnostic model was tested with another 20 sets of test data. The diagnosis results are show in Fig. 3.

Table. 4 Fuzzy entropy average value

Signal type	FE			
	Mode1	Mode2	Mode3	Mode4
F1	0.0031	0.0247	0.0004	0.0003
F2	0.2052	0.0963	0.3153	0.4678
F3	0.5099	0.8075	0.6615	0.7475
F4	0.0262	0.0011	0.0817	0.2716
F5	0.0075	0.0087	0.2101	0.0152
F6	0.0294	0.9555	0.9524	0.9418
F7	0.8750	0.0686	0.2944	0.3636

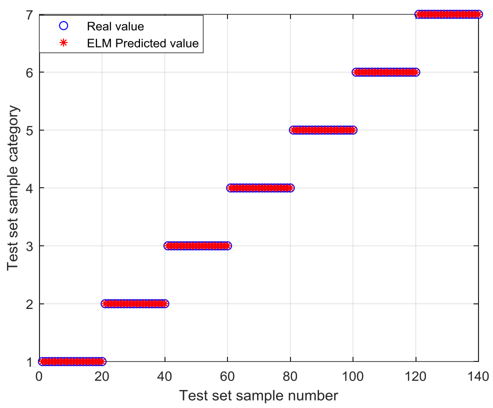
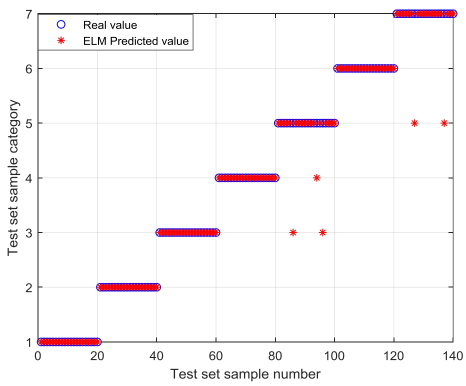


Fig. 3 VMD fuzzy entropy recognition result

For the VMD_FE, the fault categories F1, F2, F4, F5, F6, and F7 can be completely distinguished by the ELM classifier, and the fault diagnosis accuracy rate can reach 100%. As a comparative experiment, the EMD_FE and the VMD_AE are separately performed. Fault identification, the recognition result are shown in Fig. 4 and Fig. 5.



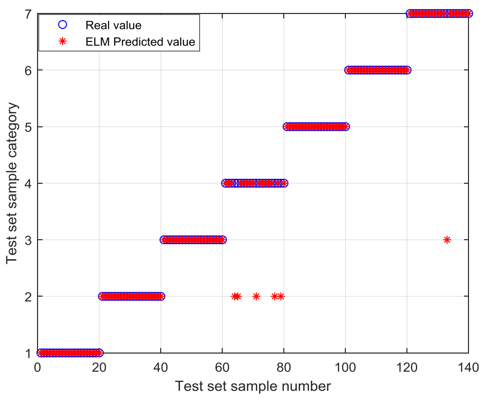


Fig. 4 EMD fuzzy entropy Fig. 5 VMD approximate entropy

EMD_FE is used for fault identification. One sample of F5 is diagnosed as F4and two sample of F5 is diagnosed as F3, F5 fault recognition accuracy is 85%, F7 has two sample recognition errors, and fault recognition accuracy is 90%. VMD_AE is used for fault. Identification, in which F4 has five sample recognition errors, the fault recognition accuracy rate is 75%,F7 have one sample recognition error respectively, and the fault recognition accuracy rate is 95%, and the VMD_FE recognition result is

100%. With higher recognition accuracy, the overall diagnostic accuracy of the bearing fault category is shown in the Table 5.

Table. 5 Accuracy of bearing fault diagnosis

Algorithm	Training accuracy	Test accuracy
VMD_FE	100%	100%
VMD_AE	100%	95.7143%
EMD_FE	99.2857%	96.4286%

4. Conclusion

In this paper, an error diagnosis method for ELM bearing fault based on VMD_FE is proposed. The method first decomposes the time domain signal of bearing in various working states, and selects the sub-signal with large correlation by mutual information method to calculate FE as the fault feature. Thus, the different bearing fault categories apply to establish an ELM fault diagnosis model to distinguish. Compared with the other two methods, the proposed method has better completed the identification of its various fault categories, and this method has a clear recognition accuracy.

References

[1] Y. Yang ,D.J. Yu , Cheng Jun-sheng:Application of EMD based singular value decomposition technique to fault diagnosis for roller bearing,Journal of Vibration and Shock, Vol.24 (2005) No.02, p.12-15.

[2] D. Xiang ,J. Cen :Method of roller bearing fault diagnosis based on feature fusion of EMD entropy. Journal of Aerospace Power,Vol.30 (2015) No.05, p.1149-1155.

[3] X.J. Li , N.S, He, et al.Cylindrical Roller Bearing Diagnosis Based on Wavelet Packet Approximate Entropy and Support Vector Machines.Journal of Vibration,Measurement & Diagnosis,Vol.35 (2015) No.06, p.1031-1036.

[4] L.G. Zhang , P. Li,et al.Fault diagnosis of rolling bearing based on ITD fuzzy entropy and GG clustering.Chinese Journal of Scientific Instrument, Vol.11 (2014) No.35, p.132-141.

[5] Z.B. Wang,F.C. Li,et al.Rolling Bearing Intelligent Diagnosis Based on Fuzzy Entropy and ITD.Noise and Vibration Control.Vol.36 (2016) No.01, p.142-146.

[6] R.Z. Ye,W.H. Li,Fault diagnosis for bearings based on wavelt packet decomposition and BP neural network.Bearing.Vol.10 (2012) No.01, p.53-56.

[7] C.H. Zhang,J.J. Hen,diagnosing faults of bearings with LMD approximate entropy and SVM.Mechanical Science and Technology for Aerospace Engineering,Vol.1 (2012) No.09, p.1539-1542.

[8] P.K. Wong,Z. Yang,C.M. Vong,et al. Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. Neurocomputing,Vol.128 (2014) No.5, p.249-257.

[9] Y. Xu,Y.J. Hen,Q.X. Hu.An Extension sample classification-based extreme learning machine ensemble method for process fault diagnosis. Chemical Engineering＆Technology,Vol.37 (2014) No.06, p.911-918.

[10] A.J. Hu. Research on the application of Hilbert-Huang Transform in vibration signal analysis of rotating machinery. Baoding,North China Electric Power University.