CEEMDAN-MFE method for fault extraction of rolling bearing

Dong An¹², Bo Xu², Meng Shao³, Hao-Dong Li² and Li-Yan Wang²

¹ Research Center for Analysis and Detection Technology, Shenyang Jianzhu University, Shenyang 110168, China
² School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China
³ Research Institute for Micronano Detection and Motion Control, Shenyang Jianzhu University, Shenyang 110168, China

Corresponding author: Meng Shao; E-mail: mshao@sjzu.edu.cn

Abstract. Due to the harsh environment, the feature extraction of rolling bearings is not convenient. In this paper, a method for the CEEMDAN (Complementary Ensemble Empirical Mode Decomposition With Adaption Noise) and MFE (Multi-scale Fuzzy Entropy) is put forward. Firstly, we use CEEMDAN to analyse the original signal decomposition and its advantages, and then get the weight of MFE feature extracting, put the weight into the SVM (Support Vector Machine) and realize fault detection of rolling bearing. The experimental results show that the accuracy of the algorithm is 98.8%.

1. Introduction
Rolling bearings are currently used in many fields, such as machinery, electricity and so on, it has its own unique advantages. However, due to the strong impact energy, it is very vulnerable to damage [¹], so mechanical fault diagnosis is of great significance to ensure the safety and stable operation of mechanical equipment.

Huang et al. first proposed the concept of Intrinsic Mode Function (IMF) and Empirical Mode Decomposition (EMD) [²], but the shortcoming of this method is the endpoint effect. Based on this problem, HUANG and YEH put forward to the Ensemble Empirical Mode Decomposition (EEMD) and the Complementary Ensemble Empirical Mode Decomposition (CEEMD) [³]. However, the disadvantages of the two methods are that the calculation is large, and if the parameter selection is unreasonable, more false components will appear in the decomposition. CEEMDAN [⁴] is an improved algorithm for these problems and has certain advantages.

For vibration signal of mechanical equipment, because the complexity of vibration signal of faults is different, the probability of new mode is also different, and its entropy value is different. Multi-scale algorithm has the advantages of high efficiency [⁵][⁶][⁷], good convergence and high precision. SVM is considered as the best method to solve the classification problem of small samples, so we use it to identify mechanical fault and diagnose results [⁸].

2. Rolling bearing fault diagnosis key technology

2.1 Complementary ensemble empirical mode decomposition with adaption noise
CEEMDAN is proposed on the basis of EEMD. Firstly, EEMD algorithm is briefly introduced, EEMD is a noise-assisted signal processing method, which uses the statistical characteristics of Gaussian white noise with uniform frequency distribution to solve the problem of modal aliasing, and then we begin to introduce the specific steps of CEEMDAN formally:

(1) The decomposition modes of the CEEMDAN will be noted as $IMF_k$ and $x[t]$ is noted as the signal to be processed, firstly give a definition:

**Definition 1.** Obtain the $k$-th modal component of a given signal by EMD can be depicted as

$$E_k(W_i[t]) = (i, \ k = 1,2, ..., I)$$

In the formula (1), $E_k(.)$ is an operator symbol, $W_i$ means gaussian white noise with $N(0,1)$.

(2) The first modal component $IMF_1$ uses EEMD algorithm to decompose, see the EEMD references for details of the specific process.

(3) Calculate the first residue:

$$r_1[t] = x[t] - IMF_1$$

(4) Decompose to the first modal component:

$$r_1[t] + \varepsilon_1 E_1(W_i[t]) \quad (i = 1,2,...,I)$$

then the second modal component can be expressed as:

$$IMF_2 = \frac{1}{I} \sum_{i=1}^{I} E_i(r_1[t] + \varepsilon_1 E_1(W_i[t]))$$

(5) Calculate the k-th residue:

$$r_k[t] = r_{k-1}[t] - IMF_k$$

(6) Decompose $r_k[t] + \varepsilon_k E_k(W_i[t])$ to the first modal component, $\varepsilon_k$ means noise standard deviation, and the k+1 modal component can be expressed as:

$$IMF_{k+1} = \frac{1}{I} \sum_{i=1}^{I} E_i(r_k[t] + \varepsilon_k E_k(W_i[t]))$$

(7) Return k+1 to step 5, repeat step 5 to step 7 until the residual margin is not suitable for decomposition and stop decomposition. Assume that K is total number of the modes, the final margin time is:

$$R[t] = X[t] - \sum_{k=1}^{K} IMF_k$$

### 2.2 Multi-scale fuzzy entropy

Multi-scale fuzzy entropy is defined as the fuzzy entropy of different scales of time series. The calculation process is as follows:

(1) Give timing signal: $u(1), u(2), ..., u(N)$, and give some definitions:

**Definition 1.** An m-dimensional vector:

$$x_i^m = \{u(i), u(i+1), ..., u(i+m-1)\} - u_0(i), i, j = 1,2,...,N-m+1$$

In the formula (1), $u_0(i)$ represents the average:

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

**Definition 2.** For an initial sequence of length N:


\[ X = \{x_1, x_2, \ldots, x_n\} \quad (10) \]

**Definition 3.** Construct coarse-grained sequence:
\[ Y = \{y_j^{(\tau)}; y_j^{(\tau)} = \frac{1}{n} \sum_{i=j-1}^{j} x_i, 1 \leq j \leq \frac{N}{\tau} \} \quad (11) \]

In the formula (11), \( \tau \) is the scale factor, when \( \tau = 1 \), \( y_j^{(\tau)} \) is the original sequence, the length of each coarse grain sequence is equal to the length of the original time series divided by the scaling factor \( \tau \).

(2) Calculate the fuzzy entropy of coarse particle sequences with different scale factors: sequence is equal to the length of the original time series divided by the scaling factor \( \tau \).

**Definition 4.** The max distance between the vector \( X_i^m \) and \( X_j^m \):
\[ d_{ij} = d[X_i^m, X_j^m] = \max_{k=0(n-m)} \{ u[i+k] - u_0(i) - u(j+k) - u_0(j) \} \quad (12) \]

(3) **Definition 5.** \( D_{ij}^m \): sample familiarity about \( X_i^m \) and \( X_j^m \):
\[ D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-(d_{ij}^m/n)^r} \quad (13) \]

In the formula (13), \( \mu(d_{ij}^m, n, r) \) means fuzzy function, \( n \) and \( r \) are the gradients and widths of their boundaries;

(4) **Definition 6.** Function:
\[ \phi^n(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left( \frac{1}{N-m+1} \sum_{j=1}^{N-m} D_{ij}^m \right) \quad (14) \]

(5) Similarly, repeat the above steps (2) to (4), and get
\[ \phi^{n+1}(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left( \frac{1}{N-m+1} \sum_{j=1}^{N-m} D_{ij}^{m+1} \right) \quad (15) \]

(6) When \( N \) is a finite number, fuzzy entropy can be expressed as:
\[ \text{FuzzyEnt}(m, n, r, N) = \lim_{n \to \infty} [\ln\phi^n(n, r) - \ln\phi^{n+1}(n, r)] \quad (16) \]

Calculating the fuzzy entropy of different scale factors, and observing fuzzy entropy changing with different scales of the original time series, it is called Multi-scale fuzzy entropy analysis. When the fuzzy entropy is calculated for each coarse grain sequence, \( \tau \) is almost the same, the range of values is \( 0.1 \sim 0.25\,\text{SD} \), and \( \text{SD} \) means the standard deviation of the original data.

The multi-scale fuzzy entropy obtained by coarse granulation of time series reflects the complexity of multi-scale fuzzy entropy. Compared with the fuzzy entropy, the fuzzy entropy can only reflect the information at a single scale and can not truly reflect the essence, all the multi-scale fuzzy entropy is superior to the fuzzy entropy.

**3. Experimental equipment and data collection**

The experimental platform includes a 2 HP motor, a torque sensor, a power meter and electronic control equipment. The tested bearing supports the motor shaft. In the experiment, the vibration signal is collected by acceleration, and the sensor is placed on the motor housing with a magnetic base. Vibration signals are collected by 16-channel DAT recorder and processed in MATLAB environment. The sampling frequency of the digital signal is 12000Hz and 48000Hz, and the bearing fault data are collected according to this equipment.
The experimental data are based on the bearing data of Case Western Reserve University. The bearing status include normal bearing (NOR), inner ring fault (IRF), body rolling fault (BRF) and outer ring fault (ORF).

4. Experimental signal analysis

4.1 Comparison between CEEMDAN and EEMD

For fault signal diagnosis algorithm of rolling bearing, we use inner ring fault data as an example, CEEMDAN is used to process the fault signal of inner ring to obtain the decomposed mode signal, EEMD is used to process the same fault signal to obtain the decomposed mode signal. The figure displays a box graph of decomposed signals and filtering iterations for each pattern and the results are shown in the figure 1 and figure 2 below.

Figure 1. Decomposition (left) and box plot (right) of CEEMDAN

Figure 2. Decomposition (left) and box plot (right) of EEMD

It can be seen from the two figures, CEEMDAN has obvious advantages. It can filter useless IMF components and is more stable at the endpoint than EEMD. Moreover, from the box graph, the number of iterations of CEEMDAN filtering are much less than that of EEMD filtering.

4.2 Comparison between multi-scale fuzzy entropy and fuzzy entropy

Before the multi-scale fuzzy entropy calculation, determine the parameters of the fuzzy entropy optimal initial embedding dimension (m), the STD of the data (r), and the scale factor (τ), by looking up the related literature and experiment with the data that matches the optimal initial parameters: m = 2,R = 0.2,τ = 20.

After determining the initial parameters, the multi-scale fuzzy entropy is calculated for each sub-node signal after CEEMDAN decomposition. The inner circle fault data is also an example, and the scale factor function is drawn (Figure 3).
Figure 3. Multiscale fuzzy entropy values of four kinds of signal (m = 2, R = 0.2 , \tau = 20)

The signals decomposed by the same CEEMDAN are respectively solved for multi-scale fuzzy entropy and fuzzy entropy, as shown in the following:

The first three entropy values of multi-scale fuzzy entropy are: T1 = (1.8791, 1.2457, 0.8219); T2 = (2.7087, 1.4882, 2.7087); T3 = (1.8863, 1.2873, 1.8863); T4 = (1.8482, 1.2888, 1.8482).

The fuzzy entropy obtained by the same CEEMDAN decomposition and calculates that the entropy values of fuzzy entropy are: Z1 = 1.8791; Z2 = 2.7087; Z3 = 1.8863; Z4 = 1.8482.

From the above entropy comparison, signal contains important time information at other scales from MFE, it can be concluded that multi-scale fuzzy entropy is obviously superior to fuzzy entropy and has obvious characteristics.

4.3 Support vector machine fault feature recognition

It can be seen from the figure that the first three fuzzy entropy values can effectively distinguish fault types, so the first three entropy values are selected as the characteristic quantity, that is T = [EN1, EN2, EN3]. The feature vector T is input into the SVM for training and testing, so as to realize the diagnosis of rolling bearing fault categories. In each of the four status mentioned above, 30 sets of data are selected, a total of 120 sets of data are collected, and a total of 120 fault feature vectors are obtained by the above method. In each status, 10 sets of eigenvectors are used for training, and the remaining 20 sets are used as test samples. Four status need to establish four second class support vector classification machine, Input the characteristic vector T of the test data sample and train four support vector machines successively and the rest of the data as test data input into the SVM, own the SVM based on MATLAB function as a criterion, SVM can output discriminant f (x) for + 1 or other, if a corresponding discriminant classifier output is the largest, it is judged as this classification [9]. The statistical results are shown in Table 1. We can be seen from the table that the CEEMDAN-MFE method has a good effect and greatly improves the accuracy of fault diagnosis of rolling bearings.

| Status | Training Sample | Test Sample | Correctly Identify | Accuracy | Overall Accuracy |
|--------|-----------------|-------------|--------------------|----------|------------------|
| NOR    | 10              | 20          | 20                 | 100%     |                  |
| IRF    | 10              | 20          | 20                 | 100%     | 98.8%            |
| BRF    | 10              | 20          | 19                 | 95%      |                  |
| ORF    | 10              | 20          | 20                 | 100%     |                  |

5. Conclusion

The fault feature extraction method based on CEEMDAN-MFE can extract the hidden fault information in fault vibration signal effectively. By comparing with EEMD, CEEMDAN is found to be better in
timeliness and endpoint inhibition. By comparing MFE with fuzzy entropy, multi-scale fuzzy entropy contains deeper mode information of time series than fuzzy entropy and has the advantages of short data required for calculation, resistance to noise and so on. Experimental research shows that the overall recognition accuracy of rolling bearing fault diagnosis algorithm based on CEEMDAN-MFE for four types of bearing status reaches 98.8%, indicating that this algorithm is effective for rolling bearing fault diagnosis.

Acknowledgment
Acknowledgements: This study is supported by the Basic Research Projects for Colleges and Universities of Liaoning Province (LJZ2017035); Science and Technology fund of Liaoning Province (20180550002); Liaoning Province Key Research and Development Project（2017225016）

References
[1] Gang C, Chen X, Li H et al Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition 2016 Measurement. 91 140-154
[2] N.E. Huang et al The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis 1998 Proc. R. Soc vol (London: Proceedings A) pp 903–995
[3] Dong L and Wang D Signal enhancement based on complex curvelet transform and complementary ensemble empirical mode decomposition 2017 J. Appl. Phys. 144
[4] Luukko P, Helske J and Rässänen E Introducing libeemd: a program package for performing the ensemble empirical mode decomposition 2016 Comput. Stat. 31(2) 545-557
[5] Xin, Y.H., Zhou J. and Chen Y.S. Dual multi-scale filter with SSS and GW for infrared small target detection 2016 Infrared. Phys. Technolo 81 97-108
[6] Azami H and Rodriguez J.E . Refined composite multivariate generalized multiscale fuzzy entropy: A Tool for Complexity Analysis of Multichannel Signals 2017 Physical A 465 261-276
[7] Zheng J, Cheng J, Yu Y, et al. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination 2014 Mech. Mach. Theory 78(16) 187-200
[8] Zheng J , Pan H , Cheng J . Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines 2017 Mech. Syst. Signal Pr, 85 746-759.
[9] Zhang M, Tang J and Zhang X Intelligent diagnosis of short hydraulic signal based on improved EEMD and SVM with few low-dimensional training samples 2016 Chin. J. Mech. Eng-En 29(2) 396-405.