

Fault Diagnosis Based on the Optimization of Characteristic Parameters and Neural Networks of Gearboxes

Huiling Liu^{1,*} and Kongwen Zhang²

¹School of Mechanics, Jinzhong University, Jinzhong 030619, China

²Department of Computer Information System, University of the Fraser Valley, Abbotsford, V2S 7M8, Canada

Received 9 January 2019; Accepted 23 May 2019

Abstract

Gearboxes are the most commonly used transmission components in heavy equipment such as helicopters, shearers, and ships. The failure rate of gearboxes is high, and the characteristic signals under faulty conditions tend to be extremely weak and are often overwhelmed by strong noise. Thus, extracting sensitive characteristic parameters is difficult. In order to optimize the characteristic parameters of gearboxes and improve diagnosis efficiency, this study proposed a method for fault diagnosis of gearboxes that combines empirical mode decomposition (EMD) with rough sets and neural networks. First, the principle of EMD was explored. The indicators for measuring characteristic parameters were identified to compare the feature set composed of energy values with those comprising approximate entropy parameters. Second, the conditional attribute reduction technique for rough sets was investigated. An algorithm for attribute reduction based on conditional equivalence classification was put forward for parameter optimization. Then, a neural network was employed to identify the feature sets before and after the attribute reduction. Results show that the energy characteristic set is the most sensitive to failures. The attribute reduction technique reduces the characteristic parameters from 6 to 4, thereby effectively lowering the input vectors of the neural network. The training time is also decreased from 1.024 s to 0.351 s, obviously promoting the efficiency of fault diagnosis. The study provides references for improving the performance of online real-time fault diagnosis.

Keywords: Fault diagnosis, Empirical mode decomposition, Rough set, Conditional attribute reduction, Neural network

1. Introduction

Gearboxes are the most important transmission components of heavy equipment, and they play an important role in transmitting torque and adjusting speed. Failures during the operation of heavy equipment are inevitable due to excessive loads or poor working environments. The characteristic signals under faulty conditions are extremely weak, and they are often overwhelmed by strong noise. Moreover, the signal-to-noise ratio is considerably low, and the degree, location, and type of faults exert a great influence on characteristic parameters. Therefore, extracting sensitive parameters and realizing online real-time fault diagnosis is difficult [1-2].

At present, the methods for extracting characteristic parameters include the time domain averaging method, cepstral analysis, refinement spectrum analysis, high-order cepstrum, envelope demodulation, etc., most of which are based on Fourier transform and are only applicable to the processing of stationary signals. Meanwhile, the vibration signals of gearboxes are non-stationary. The techniques to deal with non-stationary signals include short-time Fourier transform, Winger-Ville distribution, wavelet transform, etc. However, short-time Fourier transform could not guarantee the resolution in the time and frequency domains at the same time [3]. In Wigner-Ville distribution, cross-interference terms are apparent between components [4]. Moreover, wavelet transform is most suitable for processing with low-

frequency components [5]. By contrast, empirical mode decomposition (EMD) could display a full view of signals in the time and frequency domains simultaneously, especially during treatment of non-stationary and nonlinear signals [6-10]. Rough sets are a kind of mathematical tool to describe the incompleteness and uncertainty of data. This tool could effectively analyze and deal with incomplete information, discover hidden knowledge, and reveal potential laws [11]. The attribute reduction technique in rough sets can optimize characteristic parameters under faulty conditions. If combined with neural networks and rough sets, the method can promote the efficiency of fault diagnosis and lay a foundation for enhancing the performance of online real-time fault diagnosis.

2. State of the art

Scholars all over the world have extensively studied EMD and applied it to various fields, such as medicine, machinery, and geography. Rubén applied EMD to process and analyze physiological characteristic signals, which are beneficial for the pathological study of diseases [6]. EMD was also used to decompose seismic waves to obtain the intrinsic information of signals [7], which provided technical support for effectively identifying earthquake levels. Improved EMD algorithms have been developed and applied to fault diagnosis [8-9]. Tabrizi studied the problems of endpoint effects and false components and improved EMD in wind speed prediction [10]. These existing studies mainly focused

*E-mail address: hliu_cstu@163.com

ISSN: 1791-2377 © 2019 Eastern Macedonia and Thrace Institute of Technology. All rights reserved.

doi:10.25103/jestr.122.18

on the application of EMD, particularly its improvement, and could only slow down problems such as endpoint effects, false components, and modal aliasing according to signal characteristics. Solving these problems completely is impossible, and the related research on the faulty signals of gearboxes is rare. In line with the particularity and complexity of signals, a method combining window functions with mirror extension is utilized in the current study to relieve endpoint effects. Ensemble EMD is also used to prevent modal aliasing and ultimately obtain initial characteristic parameters accurately.

Conditional attribute reduction is the most critical technique in rough sets [11-12]. Minimal attribute reduction sets have been obtained by using attribute frequency in a difference matrix as heuristic information [13]. The information entropy of attributes has been employed as heuristic information to optimize feature sets [14-15]. To find attribute reduction sets quickly and efficiently, studies have applied the ant colony algorithm to attribute reduction [16], and some have combined particle swarm optimization and attribute reduction [17]. The genetic algorithm has been utilized to reduce the number of parameters and thereby obtain good results [18]. Although group-intelligent optimization algorithms could quickly obtain the attribute reduction sets of a decision table, they cannot ensure that the reduction sets are minimal. The present study proposes an attribute reduction algorithm that is based on conditional equivalence classification. The proposed method avoids the calculation of heuristic information and the problem of group-intelligent optimization algorithms falling into the local optimal solution while ensuring that all attribute reduction sets are minimal.

The remainder of this study is organized as follows. Section 3 presents the principle and improvement of EMD, rough sets, and neural networks. Section 4 describes the extraction of the characteristic parameters of a gearbox under faulty conditions. In this section, the feature set is optimized by an attribute reduction technique in rough sets, and a neural network is employed to identify fault patterns. A method for gearbox fault diagnosis that combines EMD with rough sets and neural networks is constructed and verified. Section 5 discusses the conclusions.

3. Methodology

3.1 EMD

EMD is a method that decomposes signals on a time scale into a number of finite intrinsic mode functions (IMFs). Each IMF represents one kind of vibration mode and describes the local characteristics of the original signals fully. During decomposition, problems such as endpoint effects, modal aliasing, and false components are inevitable. In this work, the method is raised to mirror the two ends of a signal first, and then a window function is added to the signal to ensure that the signal is not distorted. Finally, the extension part is cut off so as to avoid the endpoint effect. In addition, a random white noise sequence with finite amplitude is added to the original data every time before EMD. The sequences should be uncorrelated, the mean should be zero, and the variance should be equal. EMD is performed on the composite signal. Then, different white noise sequences are utilized each time, and complex signals are decomposed at least 100 times. Ultimately, the mean is used as the final decomposition result, and thus, modal aliasing is avoided. False IMF components may appear during EMD, and their

elimination has been studied previously [9-10]; however, now work has been able to completely remove them. Theoretically, false components are independent of the original signal, and therefore, the effective components are obtained by calculating the correlation coefficients. To obtain a valid IMF component, the current work sets the threshold to 0.06; that is, components with correlation coefficients greater than 0.06 are selected, and the others are removed.

3.1.1 Extraction technology of feature sets based on EMD

When a gearbox fails, the energy of each IMF changes, and the energy distribution of all IMFs can help distinguish the different kinds of fault types. Therefore, energy values can be extracted as a characteristic parameter set [19].

In 1991, Steven M. Pincus proposed the use of approximate entropy to measure the complexity of time series. In the current work, this technique is applied mainly to measure the probability of generating new patterns in vibration signals. A complex signal coincides with a large approximate entropy value [20], and thus, approximate entropy parameters are also constructed as a feature set.

Suppose that $\{u(i), i = 1, 2, \dots, N\}$ is the original data sequence, which contains N numbers. Assuming that the dimension is m and the similar capacity is r , the approximate entropy can be obtained by the following formula:

$$A_p^{En}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (1)$$

Equation (1) shows that the approximate entropy value A_p^{En} is related to the values of N , m , and r ; therefore, the appropriate parameters must be selected. N should be in the range of 100–5000 to ensure effective statistical characteristics and small errors. For the dimension, $m = 2$ is usually applied. The statistical characteristics are deemed reasonable when r is located between $0.1S_d$ and $0.25S_d$, where S_d is the standard deviation of the original signal.

3.1.2 Indicators for measuring characteristic parameters

Different characteristic parameters can be obtained by different methods. However, measuring characteristic parameters is difficult in the field of fault diagnosis. Characteristic parameters should satisfy two conditions: stability, that is, the values of each sample should be as close as possible under the same working condition; sensitivity, that is, the difference between the parameters of the samples under different working conditions should be large. The qualitative description of the characteristic parameters of faults has been provided previously. Diagnostic accuracy is the only indicator used to judge characteristic parameters. A quantitative method to gauge feature sets is proposed in this work. Two quantitative indicators are root mean square (RMS) and mean. RMS is also statistically called the standard deviation used to measure the degree of fluctuations of samples. A small RMS is indicative of good stability. The mean is used to survey the average degree of samples; a great difference between the mean values of samples under different working conditions indicates good sensitivity. Suppose that in a certain working condition, p delegates the number of samples ($i = 1:p$), q represents the

number of characteristic parameters in a sample ($j=1:q$), T_j is the mean of the j th characteristic parameter in all samples, and H_{ij} indicates the j th parameter of the i th sample. RMS is expressed as:

$$RMS = \sum_{j=1}^q \sqrt{\frac{\sum_{i=1}^p (H_{ij} - T_j)^2}{p}} / q \quad (2)$$

3.2 Rough sets

Rough sets are usually applied to analyze uncertainty, and they have been widely used in data mining, decision analysis, and other fields. The attribute reduction technology can particularly delete redundant components and simplify spatial dimensions so as to optimize characteristic parameters. $S = (U, \Omega, V_q, f_q)$ consists of U , a non-empty finite set called the universe; and Ω , a non-empty finite set of attributes, with $\Omega = CUD$, where C is a condition attribute set and D is a decision attribute set. For each $q \in \Omega$, V_q is called the domain of q , and f_q is an information function $f_q: U \rightarrow V_q$. The related concepts are as follows [11]:

Definition 1: If $S = (U, \Omega, V_q, f_q)$ is a decision table, $B \subseteq \Omega$ and $X \subseteq U$, then the B-lower and B-upper approximation of X are defined respectively as follows:

$$\underline{B}X = U \{Y \in U/B | Y \subseteq X\}. \quad (3)$$

$$\overline{B}X = U \{Y \in U/B | Y \cap X \neq \emptyset\}. \quad (4)$$

Definition 2: If $C \subseteq \Omega$ and $D \subseteq \Omega$, $POS_C(D)$ is a C positive region of D :

$$POS_C(D) = \bigcup_{X \in U/D} \underline{C}X \quad (5)$$

Definition 3: If $S \subseteq C$, only when S is D independent subset of C , and $POS_S(D) = POS_C(D)$, S is D reduction of C . The intersection of all reduction sets is the nuclear feature set.

3.2.1 Conditional attribute discretization

Rough set theory can only deal with discretized data. Discretization methods commonly include equidistant division, equal frequency division, naive scaler algorithm, semi naive scaler algorithm, etc. On the basis of the study of various discretization algorithms, an improved naive scaler algorithm is employed in the current work [19], and good results are obtained.

3.2.2 Attribute reduction algorithm based on conditional equivalence classification

At present, the heuristic reduction algorithm proposed by Pawlak. Z et al. [11] is popular. Two kinds of common heuristic information are dependency degree and information entropy of attributes. Minimal reduction sets may not always be available. According to the analysis of various algorithms, an attribute reduction algorithm based on conditional equivalence classification is proposed. This

algorithm avoids the calculation of heuristic information and the problem in which group-intelligent optimization algorithms fall into the local optimal solution. It also completely solves the attribute reduction sets from the perspective of classification ability, and it can find all the minimal reduction sets of a decision table.

$S = \langle U, C, D \rangle$ is the decision table, $U = \{x_1, x_2, \dots, x_N\}$ is the domain with N samples, C is the conditional attribute set, D is the decision set, and $D(x_i)$ is the value

of the decision attribute of sample x_i . The calculation process of the reduction algorithm based on the conditional equivalence classification is as follows. Algorithm 1 is used to calculate the kernel attributes. Algorithm 2 is utilized to obtain the minimal reduction set.

Algorithm 1:

Step 1) Define the kernel attribute set as an empty set $CORE = \emptyset$, and compute $POS_C(D)$ of the decision table.

Step 2) Remove any attribute C_i from C to obtain a new condition attribute set C' , and calculate $POS_{C'}(D)$. If $POS_{C'}(D) = POS_C(D)$, proceed to (2) directly. If $POS_{C'}(D) \neq POS_C(D)$, $CORE = CORE + C_i$, and then proceed to (2).

Step 3) Output $CORE$.

Algorithm 2:

Step 1) Calculate the kernel attribute set $CORE$ according to Algorithm 1.

Step 2) Compute the conditional equivalence classifications $U/CORE$ and decision classifications U/D . Find the conditional equivalence classifications that cannot be accurately classified into the decision classifications $\{x_a, x_b\}, \{x_c, x_d, x_e\}, \dots$

Step 3) Assume $D = C - CORE$, and find the conditional attribute sets E, F , which can differentiate the samples in $\{x_a, x_b\}, \{x_c, x_d, x_e\}, \dots$. Suppose $P = E \cap F$, and choose an element C_h from P so that the reduction set is $C_{min} = CORE \cup C_h$. If $P = \emptyset$, then choose any element C_i, C_j from E, F respectively so that the reduction set is $C_{min} = CORE \cup C_i \cup C_j$.

The method directly targets the condition classifications of core attributes that cannot be correctly classified into decision classifications by the kernel attribute set. It also finds the attributes that can distinguish unclassified condition classifications and finally constructs the minimal reduction set. This method can quickly find all the minimal reduction sets of a decision table, thereby reducing the calculation time and improving the efficiency of reduction.

3.3 Neural networks

The artificial neural network is a nonlinear dynamic system with strong parallel processing ability; approximation ability for nonlinear mapping; and self-organization, self-learning, and associative memory. Therefore, it is usually used to recognize failure modes in the field of fault diagnosis. Back propagation (BP) neural network is trained with the algorithm of back propagation of errors, and it is the most widely applied [21]. A BP neural network usually has three layers: an input layer, a hidden layer, and an output layer. Neurons in the same layer have no connections, and neurons

in different layers are connected by weights. Generally, a neural network can have only one input or output layer and multiple hidden layers. However, any nonlinear function can be approximated using an implicit layer. In this work, a BP neural network with three layers is adopted to pattern identification.

4. Analysis and Discussion

4.1 Signal acquisition and preprocessing

Signal acquisition is the most important step in fault diagnosis. The test apparatus of gearboxes is mainly composed of a motor, coupling, a gearbox, and a magnetic powder brake (Fig. 1). A gearbox consists of an input shaft, an intermediate shaft, an output shaft, three pairs of rolling bearings, two pairs of gears, and a case (Fig. 2). In the experiment, the rated speed is 1200 r/min, the load is 300 Nm, and the sampling frequency is 4000 Hz. According to the failure rates of the parts in the gearbox, five kinds of work conditions are set: tooth fracture, crack in the outer ring of the bearing, pitting in the inner ring of the bearing, tooth fracture and crack in the outer ring, and tooth fracture and pitting in the inner ring. One tooth of the driven gear on the intermediate shaft is broken. At the same time, the crack in the outer ring and the pitting in the inner ring are arranged on the rolling bearing of the intermediate shaft close to the magnetic powder brake. According to fault types, the measuring point V is selected for signal collection. The signals are analyzed in the time and frequency domains, as demonstrated in Fig. 3.

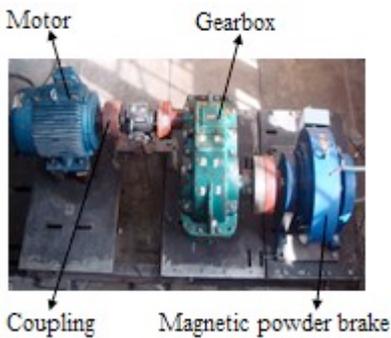


Fig.1. Test apparatus for fault diagnosis

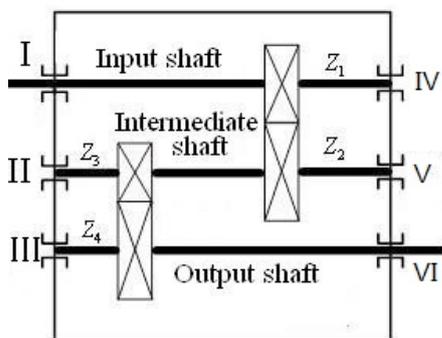
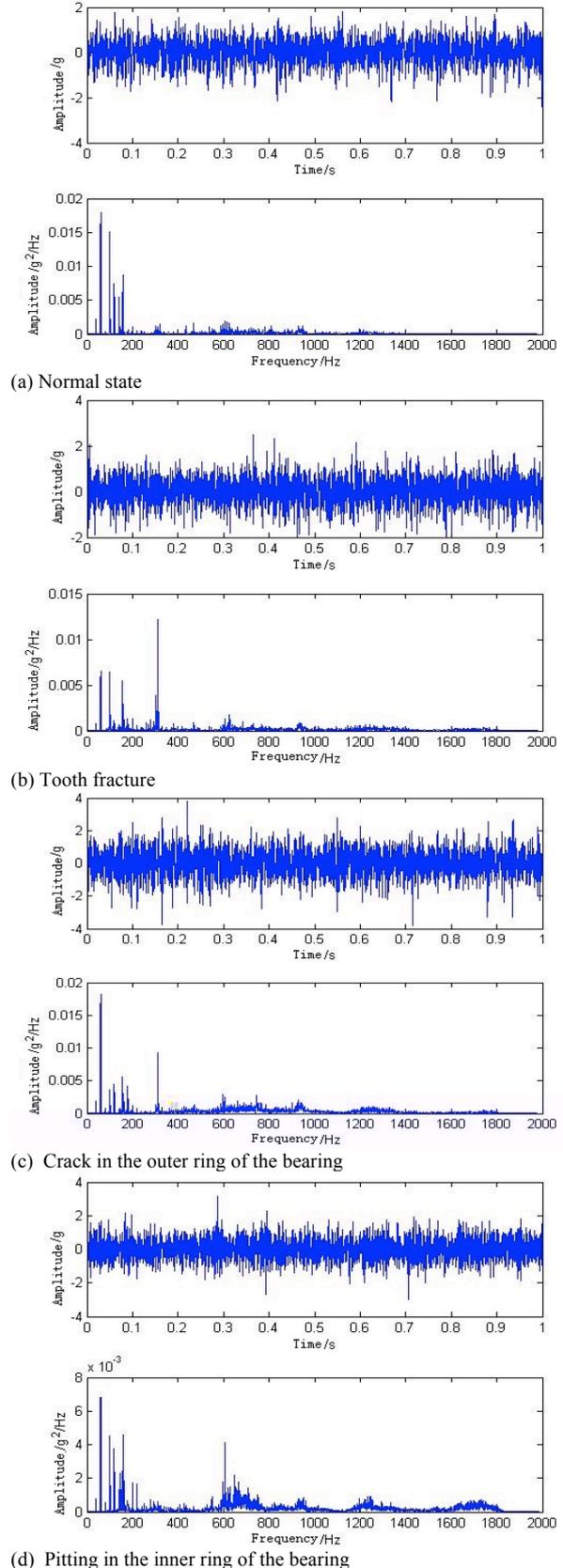


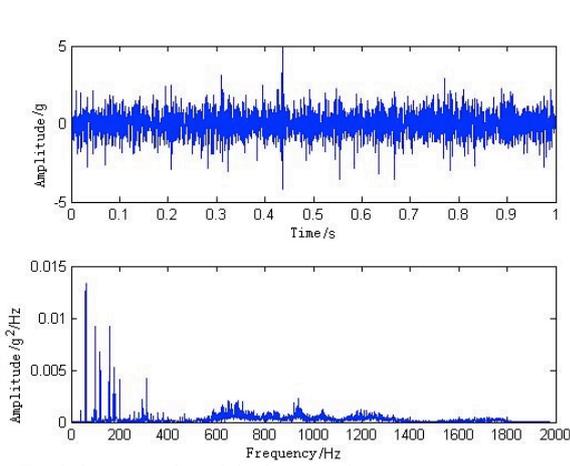
Fig. 2. Simplified diagram of a gearbox

4.2 Characteristic parameters

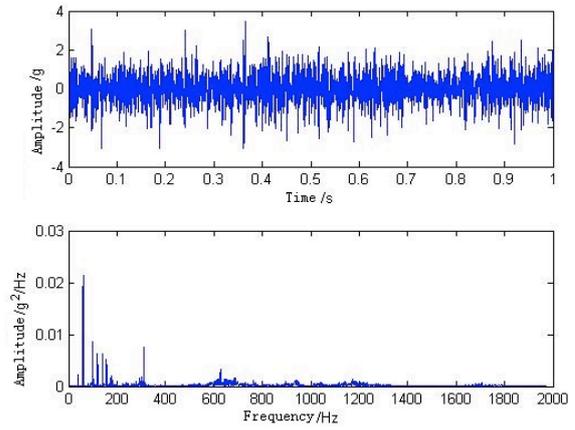
The signals in the time domain are processed by EMD according to the principle in Section 3.1. The first six components are selected to calculate the characteristic parameters. Table 1 shows the feature set composed of energy values based on EMD. Table 2 presents the set made up of the approximate entropy parameters based on EMD.

Five samples are taken for each work condition, and thus, the total number of samples is 30. *a, b, c, d, e, f* represent the characteristic parameters of each IMF component, and 1, 2, 3, 4, 5, 6 correspond to the six work conditions, namely, normal state, tooth fracture, crack in the outer ring of the bearing, pitting in the inner ring of the bearing, tooth fracture and crack in the outer ring, and tooth fracture and pitting in the inner ring.





(e) Tooth fracture and crack in the outer ring



(f) Tooth fracture and pitting in the inner ring

Fig. 3. Vibration signals in time and frequency domains

Table 1. Feature set composed of energy values

<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>D</i>
<i>x</i> ₁	0.8645	0.3926	0.1926	0.1902	0.1524	0.0446	1
<i>x</i> ₂	0.8372	0.3903	0.2182	0.2635	0.1691	0.0330	1
<i>x</i> ₃	0.8578	0.3683	0.1946	0.2466	0.1703	0.0309	1
<i>x</i> ₄	0.8768	0.3518	0.1885	0.2397	0.1162	0.0305	1
<i>x</i> ₅	0.8522	0.3935	0.1995	0.2394	0.1440	0.0322	1
<i>x</i> ₆	0.9140	0.3088	0.2080	0.1334	0.0824	0.0383	2
<i>x</i> ₇	0.9111	0.3063	0.2161	0.1510	0.0793	0.0183	2
<i>x</i> ₈	0.9193	0.2848	0.2182	0.1360	0.0792	0.0369	2
<i>x</i> ₉	0.9141	0.2997	0.2246	0.1341	0.0732	0.0300	2
<i>x</i> ₁₀	0.8913	0.3176	0.2596	0.1709	0.0863	0.0244	2
<i>x</i> ₁₁	0.8851	0.4151	0.1780	0.0840	0.0724	0.0194	3
<i>x</i> ₁₂	0.8655	0.4556	0.1595	0.1153	0.0590	0.0331	3
<i>x</i> ₁₃	0.9110	0.3739	0.1489	0.0730	0.0507	0.0134	3
<i>x</i> ₁₄	0.8791	0.4365	0.1519	0.0970	0.0597	0.0237	3
<i>x</i> ₁₅	0.9121	0.3657	0.1493	0.0945	0.0519	0.0191	3
<i>x</i> ₁₆	0.9475	0.2884	0.0919	0.0951	0.0370	0.0145	4
<i>x</i> ₁₇	0.9297	0.3345	0.1156	0.0881	0.0495	0.0130	4
<i>x</i> ₁₈	0.9221	0.3575	0.1040	0.0998	0.0318	0.0128	4
<i>x</i> ₁₉	0.9253	0.3369	0.1330	0.0948	0.0571	0.0217	4
<i>x</i> ₂₀	0.9345	0.3222	0.1143	0.0875	0.0408	0.0231	4
<i>x</i> ₂₁	0.9247	0.3331	0.1267	0.1245	0.0457	0.0164	5
<i>x</i> ₂₂	0.9391	0.3022	0.1394	0.0762	0.0360	0.0116	5
<i>x</i> ₂₃	0.9420	0.2895	0.1332	0.0933	0.0480	0.0097	5
<i>x</i> ₂₄	0.9089	0.3651	0.1590	0.1133	0.0492	0.0072	5
<i>x</i> ₂₅	0.9164	0.3582	0.1386	0.1031	0.0434	0.0112	5
<i>x</i> ₂₆	0.8517	0.4268	0.2028	0.1882	0.1177	0.0457	6
<i>x</i> ₂₇	0.8516	0.4407	0.1992	0.1710	0.0967	0.0477	6
<i>x</i> ₂₈	0.8612	0.4416	0.1951	0.1333	0.0803	0.0327	6
<i>x</i> ₂₉	0.8894	0.3706	0.2067	0.1396	0.0924	0.0283	6
<i>x</i> ₃₀	0.8788	0.3942	0.1765	0.1573	0.1195	0.0456	6

According to the method in Section 3.1.3, the RMS of each work condition is computed. The RMS values of the set in Table 1 are 0.0149, 0.0108, 0.0154, 0.0106, 0.0129, and 0.0161. The RMS values of the set in Table 2 are 0.0160, 0.0268, 0.0186, 0.0204, 0.0246, and 0.0203. The RMS values of the former are smaller than those of the latter and thus indicate better stability. Similarly, the mean value of each set is calculated separately, and the difference of the means between two operating conditions is obtained. Finally, the minimal value is taken as an indicator of sensitivity. For the six work conditions, the values of the first set are 0.0056, 0.0017, 0.0026, 0.0003, 0.0013, and 0.0046; and the corresponding values of the second set are 0.0007, 0.0016, 0.0059, 0.0037, 0.0009, and 0.0065. The sensitivities of the

two sets are basically the same. Therefore, the energy characteristic parameter set is the most suitable for gearbox fault diagnosis. Table 1 is used as basis to recognize faulty modes.

4.3 Characteristic parameter optimization based on attribute reduction

According to rough set theory (Section 3.2), the characteristic parameters in Table 1 are discretized, and the feature set is reduced with the algorithm based on conditional equivalence classifications. The kernel attribute set is , The minimal reduction sets are , . To verify the correctness of the reduction method, this study calculates the positive domains of the set before and after the reduction

according to Equation (5). The results show that the positive domains are exactly the same, and thus, the method proposed in this study is effective and feasible.

4.4 Fault pattern recognition based on neural network

The neural network theory described in Section 3.3 is adopted for pattern recognition. Table 1 shows five samples for each condition; the first four samples are the training samples and, the fifth sample is the test sample. The

structure of the BP neural network used to recognize the set before reduction is 6-12-6, and that after reduction is 4-8-6. The curves of the training error before reduction are shown in Fig. 4. The curves of C_{min1} and C_{min2} are shown in Figs. 5 and 6, respectively. Table 3 presents the test results of the neural network. $X_1, X_2, X_3, X_4, X_5, X_6$ represent the actual output.

Table 2. Feature set composed of approximate-entropy parameters

<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>D</i>
x_1	1.5721	1.3253	0.9758	0.6435	0.6277	0.5192	1
x_2	1.5405	1.3357	0.9710	0.6714	0.6246	0.5166	1
x_3	1.5742	1.2881	0.9783	0.6843	0.6295	0.5060	1
x_4	1.6127	1.2955	1.0183	0.6554	0.6125	0.5186	1
x_5	1.5883	1.3315	0.9355	0.6765	0.6299	0.5086	1
x_6	1.5486	1.3591	1.1657	0.8021	0.6093	0.5476	2
x_7	1.5693	1.3616	1.0315	0.7001	0.5948	0.4819	2
x_8	1.5986	1.3374	1.0609	0.7280	0.6024	0.5730	2
x_9	1.6015	1.3393	1.0133	0.6987	0.5980	0.5269	2
x_{10}	1.5614	1.3553	1.1233	0.7200	0.6011	0.5242	2
x_{11}	1.5935	1.4872	1.0182	0.6825	0.6303	0.5532	3
x_{12}	1.5785	1.4751	1.0553	0.7132	0.6557	0.5914	3
x_{13}	1.5999	1.4704	1.0122	0.6731	0.6236	0.5267	3
x_{14}	1.6205	1.4406	1.0078	0.7347	0.6286	0.5646	3
x_{15}	1.6097	1.4386	0.9976	0.6868	0.6200	0.5335	3
x_{16}	1.5111	1.4861	1.0647	0.7576	0.5735	0.5001	4
x_{17}	1.5078	1.5004	1.1065	0.7531	0.6117	0.4800	4
x_{18}	1.4382	1.5158	1.0879	0.7791	0.6140	0.5178	4
x_{19}	1.4821	1.4652	1.1032	0.7686	0.6147	0.5377	4
x_{20}	1.4204	1.5160	1.0619	0.7785	0.5805	0.5010	4
x_{21}	1.5564	1.4328	1.0535	0.7375	0.6037	0.5403	5
x_{22}	1.5776	1.4216	1.0200	0.6918	0.5926	0.4923	5
x_{23}	1.6002	1.4197	0.9912	0.7342	0.6029	0.4978	5
x_{24}	1.5873	1.4615	1.1305	0.7267	0.5980	0.4188	5
x_{25}	1.5628	1.4685	1.0650	0.6919	0.5929	0.4978	5
x_{26}	1.5719	1.4073	0.9840	0.7132	0.6362	0.5683	6
x_{27}	1.5984	1.3972	1.0033	0.7079	0.6141	0.5902	6
x_{28}	1.5996	1.3873	1.0105	0.7024	0.6378	0.5419	6
x_{29}	1.6256	1.3921	0.9824	0.6781	0.6245	0.5557	6
x_{30}	1.6395	1.3796	1.0814	0.7617	0.6271	0.5765	6

4.5 Discussion

The BP neural network is applied to recognize fault patterns. The results show that the accuracies of the three sets can reach 100%. However, Fig. 4 indicates that the training accuracy is 0.000301, the training frequency is 69 times, and the training time is 1.024 s. Moreover, the precision of C_{min1} is 0.000224, the frequency is 16 times, and the time is 0.476 s. Meanwhile, the accuracy of C_{min2} is 0.000248, the frequency is 13 times, and the time is 0.351 s. The training precision of the sets after reduction is smaller than that before reduction, and C_{min2} has the lowest training frequency and shortest time.

The simulation results of the BP neural network highlight the feasibility of the method for fault diagnosis of gearboxes that combines EMD with rough sets and neural networks. In guaranteeing the effects of online real-time diagnosis, the number of characteristic parameters should not be excessive. The algorithm for attribute reduction based on conditional equivalence classifications reduces the characteristic parameters from 6 to 4 and promotes the

efficiency of fault diagnosis, thereby benefitting online real-time diagnosis.

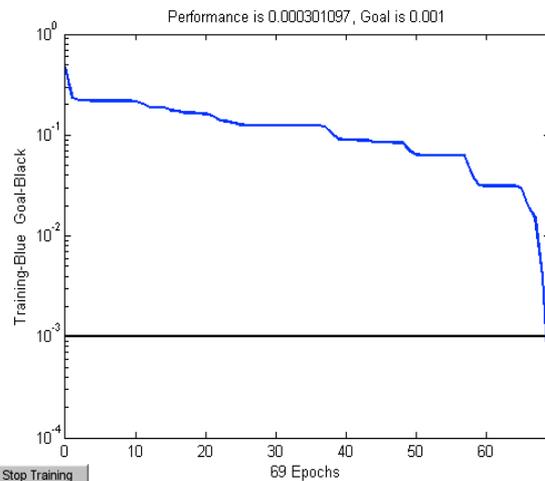


Fig. 4. Training error of C

Table 3. Testing results of neural network

Set	Condition	X_1	X_2	X_3	X_4	X_5	X_6
C	Normal state	0.9967	0.0098	0.0053	0.0028	0.0178	0.0238
	Tooth fracture	0.0124	0.9876	0.0045	0.0031	0.0249	0.0421
	Crack in the outer ring of the bearing	0.0678	0.0320	0.9605	0.1023	0.0726	0.0673
	Pitting in the inner ring of the bearing	0.0011	0.0025	0.0087	0.9975	0.0389	0.0154
	Tooth fracture and crack in the outer ring	0.0367	0.0410	0.0947	0.0012	0.9359	0.0069
	Tooth fracture and pitting in the inner ring	0.0148	0.0069	0.0387	0.0057	0.0276	0.9572
C_{min1}	Normal state	0.9537	0.0027	0.0074	0.0039	0.0756	0.0045
	Tooth fracture	0.0372	0.9840	0.0012	0.0008	0.0359	0.0249
	Crack in the outer ring of the bearing	0.0017	0.0001	0.9769	0.0248	0.0076	0.0642
	Pitting in the inner ring of the bearing	0.0021	0.0098	0.0102	0.9976	0.0327	0.0184
	Tooth fracture and crack in the outer ring	0.0357	0.0034	0.0508	0.0698	0.9746	0.0321
	Tooth fracture and pitting in the inner ring	0.0612	0.0076	0.0898	0.0872	0.0284	0.9653
C_{min2}	Normal state	0.9879	0.0101	0.0206	0.0007	0.0479	0.0064
	Tooth fracture	0.0203	0.9899	0.0013	0.0309	0.0236	0.0352
	Crack in the outer ring of the bearing	0.0011	0.0052	0.9893	0.0741	0.0687	0.0649
	Pitting in the inner ring of the bearing	0.0052	0.0029	0.0263	0.9742	0.0436	0.0228
	Tooth fracture and crack in the outer ring	0.0178	0.0231	0.0049	0.1047	0.9947	0.0147
	Tooth fracture and pitting in the inner ring	0.0083	0.0062	0.1262	0.0784	0.0345	0.9538

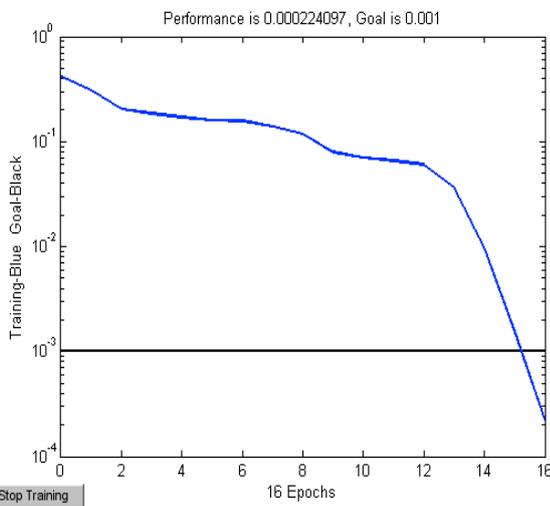


Fig. 5. Training error of C_{min1}

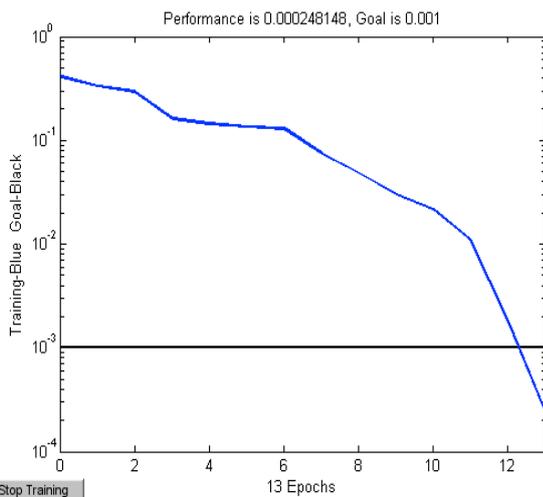


Fig. 6. Training error of C_{min2}

5. Conclusion

This study was aimed toward the nonlinear and non-stationary vibration signals of gearboxes. A feature set made

up of energy values based on EMD was obtained to identify the initial characteristic parameters. An algorithm for attribute reduction based on conditional equivalence classifications was established to obtain all the minimal reduction sets. A neural network was employed to identify the feature sets before and after reduction, and the following results are obtained.

- (1) According to the indicators for measuring the characteristic parameters of failures, the experimental results show that an energy characteristic set is more sensitive to fault modes than an approximate entropy set.
- (2) The algorithm based on conditional equivalence classifications for attribute reduction obtains the minimal attribute reduction set rapidly, effectively lessens the number of input vectors in neural networks, and simplifies the structure.
- (3) The minimal attribute reduction set can recognize failure modes accurately and promote the efficiency of fault diagnosis.

A method combining rough sets with neural networks was constructed on the basis of EMD for the fault diagnosis of gearboxes. The proposed method offers many advantages, such as strong anti-interference ability, fast convergence, and high diagnostic accuracy. Thus, it is helpful for online fault diagnosis. If the actual conditions of gearboxes are processed with the proposed method in the future, some modifications should be executed, and the mode identification should be made accurate.

Acknowledgment

This work was supported by the Applied Basic Research Programs of Shanxi Province (Grant No. 201701D221132), Teaching Reform and Innovation Project of Colleges and Universities in Shanxi Province (Grant No. J2018175), Bo Shi Ji Jin Programs of Jinzhong (Grant No. bsjj2015206), and Key Innovation Team Construction Plan of 1331 Project of Jinzhong University (Grant No. jzjxcytd2017005).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License



References

1. Chuan Li, René-Vinicio Sanchez, Grover Zurit, et al., "Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis". *Neurocomputing*, 168(11), 2015, pp.119-127.
3. Fannia Pacheco, José Valente de Oliveira, René-Vinicio Sánchez, et al., "A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions". *Neurocomputing*, 194(6), 2016, pp. 192-206.
3. Fulong Jing, Chunjie Zhang , Weijian Si, et al., "Polynomial Phase Estimation Based on Adaptive Short-Time Fourier Transform". *Sensors*, 18(2), 2018, pp.568-581.
4. TONG Ji-jun, LI Lin, LIN Qin-guang, et al., "SSVEP brain-computer interface (BCI) system using smoothed pseudo Wigner-Ville distribution". *Journal of Zhejiang University (Engineering Science)*, 51(3), 2017, pp.598-604.
5. Dipalee Gupta, Siddhartha Choubey, "Discrete Wavelet Transform for Image Processing". *International Journal of Emerging Technology and Advanced Engineering*, 4(3), 2015, pp.598-602.
6. Rubén San-Segundo, Manuel Gil-Martín, Luis Fernando D'Haro-Enríquez, et al., "Classification of epileptic EEG recordings using signal transforms and convolutional neural networks". *Computers in Biology and Medicine*, 109(6), 2019, pp.148-158.
7. Yasaman Maleki, Mostafa Allamehzadeh, "Time-Dependent Scaling Patterns in Sarpol-e Zahab Earthquakes". *Journal of Seismology and Earthquake Engineering*. 20(2), 2018, pp.21-27.
8. Jaouher Ben Ali, Nader Fnaiech, Lotfi Saidi, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals". *Applied Acoustics*, 89(3), 2015, pp.16-27.
9. Yong Lv, Rui Yuan, Gangbing Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing". *Mechanical Systems and Signal Processing*, 81(12), 2016, pp.219-234.
10. Tabrizi, A., Garibaldi, L., Fasana, A., "Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine". *Meccanica*, 50(3), 2015, pp. 865 - 874.
11. Pawlak Z, "Rough set approach to multi-attribute decision analysis". *European Journal of Operational Research*, 72(3), 1994, pp. 443-459.
12. Jelonek, J., "Particle swarm optimization: developments, applications and resources". *Computational Intelligence*, 11(2), 1995, pp. 339-347.
13. Wang G., Zhao J., An J., et al., "A Comparative Study of Algebra Viewpoint and Information Viewpoint in Attribute Reduction". *Funda-menta Informaticae*, 68(3), 2005, pp.289-301.
14. Xiao Zhang, Changlin Mei, Degang Chen, et al., "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy". *Pattern Recognition*, 56(8), 2016, pp.1-15.
15. Zhao, J., Zhang, Z., Han, C., et al., "Complement information entropy for uncertainty measure in fuzzy rough set and its applications". *Soft Computing*, 19(7), 2015, pp.1997 - 2010.
16. Youchuan Wan, Mingwei Wang, Zhiwei Ye, et al., "A feature selection method based on modified binary coded ant colony optimization algorithm". *Applied Soft Computing*, 49(12), 2016, pp. 248-258.
17. Jianhua Dai, Huifeng Han, Qinghua Hu, et al., "Discrete particle swarm optimization approach for cost sensitive attribute reduction". *Knowledge-Based Systems*, 102(6), 2016, pp.116-126.
18. Tamer Fetouh , Mohamed S. Zaky, "New approach to design SVC-based stabiliser using genetic algorithm and rough set theory". *IET Generation, Transmission & Distribution*, 11(2), 2017, pp.372-382.
19. Liu Huiling, Pan Hongxia, Liu Xiaojuan, "Fault Diagnosis of Gearbox Using EMD and Rough Set". *Journal of Convergence Information*, 8(1), 2013, pp.316-323.
20. Pincus, S.M., "Approximate entropy as a measure of system complexity". *Proceedings of the National Academy of Sciences of the United States of America*, 88 (6), 1991, pp.2297-2301.
21. Yu-Rong Zeng, Yi Zeng, Beomjin Choi, et al., " Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network". *Energy*, 127(5), 2017, pp.381-396.