TELKOMNIKA, Vol. 11, No. 9, September 2013, pp. 5471~5479
ISSN: 2302-4046

Rough Sets Algorithm and its Application in Fault Diagnosis

Huiling Liu[1,2], Hongxia Pan*[1], Aiyu Wang[1]
[1]School of Mechanical Engineering and Automation, North University of China, Xueyuan Road, Taiyuan 030051, China
[2]School of Mechanics, JinZhong University, Jinzhong 030600, China
*Corresponding author, e-mail: panhx1015@sohu.com

Abstract

Gearbox is one of the most complicate rotary mechanical apparatus, the fault signal shows non-linear and non-stationary, and how to recognize the faults effectively is a key issue. A novel method based on wavelet packet transform and rough sets theory was presented for fault diagnosis of gearbox. First, the vibration signals were decomposed into eight bands from low frequency to high frequency by wavelet packet transform, energy characteristics were extracted as the condition attributes. Second, an improved NaiveScaler algorithm was put forward to discrete continuous attributes in the case of assuring classification ability. A new reduction algorithm based on condition equivalence classifications was proposed to delete the redundant features, which could improve the reduction efficiency. Lastly the decision rules were drawn and utilized to test the samples. The results show that the method could obtain more sensitive fault characteristic parameters and have better classification ability accordingly.

Keywords: *rough sets algorithm, fault diagnosis, wavelet packet transform, decision rules*

1. Introduction

Gearbox is usually used to power transmission and velocity distribution in modern machinery equipment, once the faults occur, the whole transmission system would be interrupted. Because of the poor working environment and lower signal-noise rate, the vibration signal of gearbox is usually non-linear and non-stationary, most of the traditional signal processing methods in core of Fourier Transform are not suitable to deal with the signal, so how to recongnize the faults effectively and quickly is a key issue [1].

Z. Pawlak proposed rough sets theory(RST) in 1982, which is a newer mathematical tool to process the fuzzy and uncertainty knowledge [2]. RST can analyze and deal with all kinds of incomplete data without any prior knowledge, and reveal the internal laws, so it is used in many domains such as stock market forecast and medical diagnosis [3-5]. In recent years, RST has been applied in fault diagnosis field gradually for reducting the characteristic parameters and extracting decision rules [6-8]. Wavelet packet transform(WPT) is developed based on wavelet transform [9,10], which would decompose both in low and high frequences at the same time and has a great result in time and frequence domain analyzation, so it is feasible to process the gearbox signal. At present, WPT is mainly used to extract the characteristic parameters of the faults through reconstructing the time-domain signal of every frequency band.

In the paper, the composite method of WPT and RST would be applied to fault diagnosis [11]. The signal collected from gearbox would be decomposed by orthogonal wavelet in full-scale, the energetic feature parameters are got by analyzing and computing the energy distributed in every frequence band. Attributes discretization and reduction Algorithms are studied to extract more sensitive features and decision rules.

2. Rough Sets Theory

2.1. Basic Concepts

Rough Sets Theory is used to process the fuzzy and uncertainty knowledge. RST can analyze and deal with all kinds of incomplete data without any prior knowledge. The detail information is described in the works of Pawlak and Yasdi [12]. Here are some related concepts.

Received February 25, 2013; Revised June 10, 2013; Accepted June 26, 2013

$S=\left(U,\Omega,V_q,f_q\right)$ represents an information system, where, U is the universe, Ω is a nonempty finite set of condition attributes C and decision attributes D. If $q\in\Omega$, V_q is the domain of q, f_q is regarded as a function from U to V_q. Suppose $B\subseteq\Omega$ and $X\subseteq U$, then B-upper and B-lower approximation of X are defined respectively : $\overline{B}X=\bigcup\left\{Y\in U/B\middle|Y\bigcap X\neq\phi\right\}$, $\underline{B}X=\bigcup\left\{Y\in U/B\middle|Y\subseteq X\right\}$. If $C\subseteq\Omega$ and $D\subseteq\Omega$, C positive region of D is $POS_c(D)=\bigcup_{X\in U/D}\underline{C}X$. If $S\subseteq C$, and $POS_S(D)=POS_C(D)$, S is D reduction of C.

2.2. Attributes Discretization Algorithm

Generally the value got by kinds of signal processing methods is continuous, while rough sets only can deal with discrete data, so data discretization properly is very important and would lay the foundation for attributes reduction. At present there are several discrete algorithms: Naïve Scaler (NS) algorithm, equidistant and equifrequent classification, etc. Usually the candidate breakpoint set is got by NS algorithm, then kinds of majorization algorithms such as genetic algorithm and particle swarm optimization are employed to optimize the breakpoint set, but the method often results in more breakpoints. The principle of discretization is the least breakpoints under the condition of keeping classification ability. Here an improved NS algorithm is put forward as follows:

(1) Calculating $POS_C(D)$ according to 2.1 before discretization.

(2) Gaining the candidate breakpoint sets of all the attributes by NS algorithm.

(3) Choosing a breakpoint from the sets respectively to discrete condition attributes and computing $POS_C(D)$, If equal to that before discretization, go to (5), else to (4).

(4) Adding another breakpoint to discrete, calculating and comparing, If equal to (5), else to (4).

(5) Outputting the final breakpoint sets and the discrete decision table.

In the discretization algorithm, when adding a new breakpoint into the set, it is necessary to divide the universe evenly as much as possible in order to ensure the least breakpoints in the case of keeping classification ability.

2.3. Reduction Algorithm based on Condition Equivalence Classifications

Condition attributes reduction technology is to find the minimal feature vector through deleting the redundant features in the case of keeping classification ability, which is a NP hard problem. At present, the basic idea of attribute reduction algorithms is that computing core attributes first, then adding new features according to the heuristic information. Two kinds of common heuristic information are the attribute dependence and attribute information entropy, the principles of adding attributes are dependence of condition attributes to decision attribute and probability of the sample occurrence respectively, which result in retaining more attributes.

Here, A new reduction algorithm based on condition Equivalence classifications is proposed, the samples not assigned to decision classes only by the core property can be classified accurately through adding the less attributes. The basic idea is: first obtaining the core attributes and computing condition equivalence classifications, then finding the attributes in the rest of the conditions which can distinguish the samples in the condition equivalence classifications not assigned to decision classes properly by the core attributes, lastly the attributes and the core attributes construct the final reduction set. The method could divide the condition equivalence classifications more finely, so it can ensure all the samples classified correctly. Additionally the method could find the core attributes quickly and add the new attributes purposefully, which greatly save computing time and improve the reduction efficiency.

An information system $S=(U,C,D)$, $D\left(x_i\right)$ represents the decision attribute value of x_i, the specific algorithm is as follows:

Core attributes algorithm:

(1) Defining that core attribute set is empty $CORE=\varnothing$ and computing $POS_C(D)$ of the whole decision table.

(2) Removing C_i from $C\ (i=1, i\le n, i++)$ to get a new condition attributes set C', computing $POS_C(D)$, if $POS_{C'}(D)=POS_C(D)$, go to (2), else to (3).

(3) $CORE=CORE+C_i$, go to (2).

(4) Outputting $CORE$.

Reduction attributes algorithm:

(1) Obtaining core attributes set $CORE$ according to core attributes algorithm above;

(2) Computing condition equivalence classifications $U/CORE$ and decision equivalence classifications U/D, selecting all the condition classifications not classified into decision classifications correctly: $\{x_a, x_b\}, \{x_c, x_d, x_e\}\cdots$.

(3) Assuming $D=C-CORE$, finding the property sets E, F which can differentiate the samples in $\{x_a, x_b\}, \{x_c, x_d, x_e\}\cdots$. Supposing $P=E\cap F$, choosing a element C_k arbitrarily from P ,so the reduction set is $C_{\min}=CORE\cup C_k$. If $P=\phi$, choosing elements C_i, C_j from E, F respectively, so the reduction set is $C_{\min}=CORE\cup C_i\cup C_j$.

(4) When computing F, if $D(x_c)\ne D(x_d)\ne D(x_e)$, F should distinguish three samples at the same time. If $D(x_c)=D(x_d)\ne D(x_e)$, F only need to distinguish x_e and x_c, x_d, so F may include one element or more than two attributes. Additionally if there are more than three samples in the condition equivalence classifications, the way is the same as above.

3. Fault Diagnosis of Gearbox based on RST and WPT

3.1. Characteristics of Gearbox

Gearbox is one of the most complicate rotary mechanical apparatus, the fault signal shows non-linear and non-stationary. Most of the traditional signal processing methods in core of Fourier Transform are not suitable to deal with the fault signal, so it is very difficult to acquire the sensitive characteristics. The tests are done on JZQ-250 gearbox, which is made up of three pairs of rolling bearings, two pairs of straight gears, the input shaft, intermediate shaft and output shaft. There are totally six kinds of states for research: normal state, tooth fracture, crackle of inner ring, cage fracture, composite fault of tooth fracture and crackle of inner ring, composite fault of tooth fracture and cage fracture. The experimental parameters are: the rated speed of the input shaft 1200r/min, the sampling frequency 4000Hz. The vibration signals in time domain are collected by the acceleration sensors, seen in Figure 1.

3.2. Energy Features Extraction based on WPT

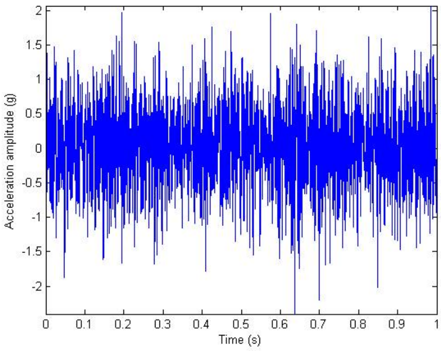
Wavelet packet transform is developed based on wavelet transform, but when dealing with the vibration signal, WPT would decompose both in low and high frequences at the same time, which has a great result in time and frequence domain analyzation, so it is suitable to process the non-stationary signals.

In wavelet packet transform, the discrete signal is made convolution with a low pass filter and a high pass filter respectively, the approximate and detail coefficients would be got. The former represents low-frequence component, while the latter expresses high-frequence of the signal. When the faults occur, energy in each frequency band would change: some frequence bands increase, some frequence bands decrease, others keep. The variation of energy distribution can reflect the different fault patterns, so it is reasonable to extract the energy characteristic parameters from every frequence band as the condition attributes in decision table.

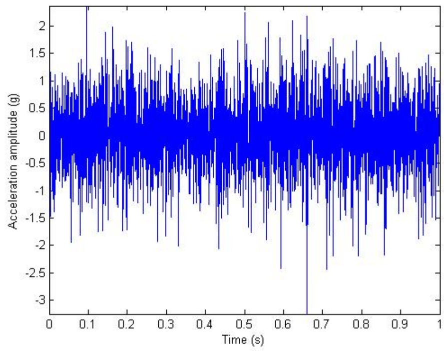
Here, the signals are decomposed into three layers with “db4”, there are eight frequence bands totally, seen in Figure 2. Every band would be processed further to extract energy feature vectors, the main steps are described as follows:

(1) Decomposing the signal and getting the amplitude of every discrete point in each frequence band.

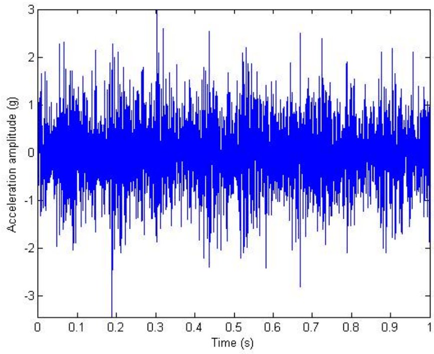
(2) Computing energy of each frequence band.



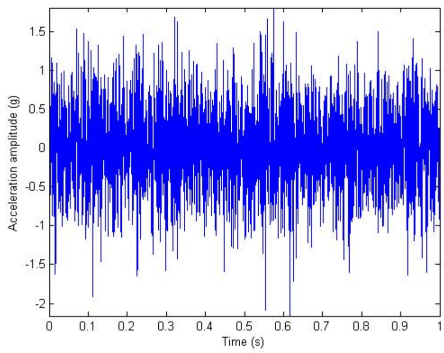
a. Normal state



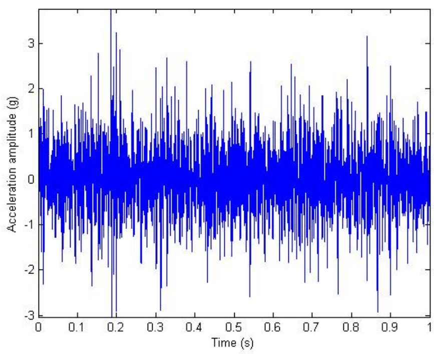
b. Tooth fracture



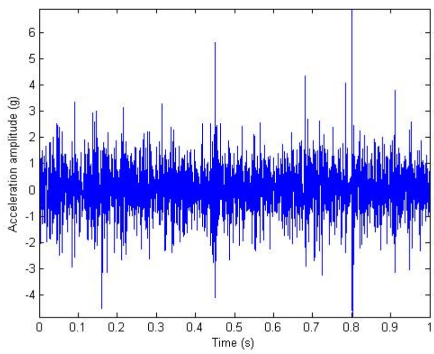
c. Crackle of inner ring



d. Cage fracture



e. Tooth fracture and crackle of inner ring



f. Tooth fracture and cage fracture

Figure 1. Vibration Signals in Time Domain

$$E_{j,r}=\int\left|S_{j,r}(t)\right|^2 \mathrm{dt}=\sum_{\mathrm{k}=1}^{m}\left|x_{j,r}^k\right|^2 \tag{1}$$

Where, j,r and m are the number of layer, node and discrete point respectively, $x_{j,r}^k$ is the amplitude of k discrete point on j layer r node, $S_{j,r}(t)$ represents the reconfiguration signal, $E_{j,r}$ is regarded as the egergy.

(3) Constructing the feature vectors T.

$$T=\left[E_{j,1}, \mathrm{E}_{j,2},\cdots,\mathrm{E}_{j,r},\cdots E_{j,2^j-1}\right] \tag{2}$$

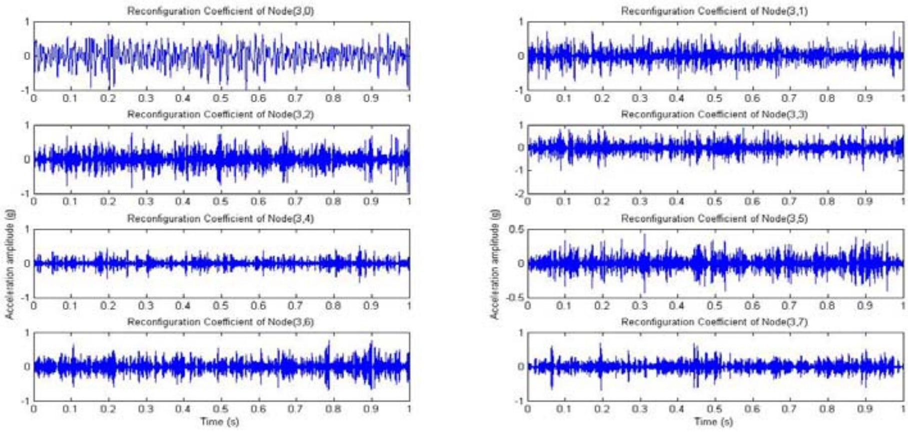
For $E_{j,r}$ is large, it is necessary to normalize according to the equations below.

$$E = \left(\sum_{r=0}^{2^j-1} \left| E_{j,r} \right|^2 \right)^{1/2} \quad (3)$$

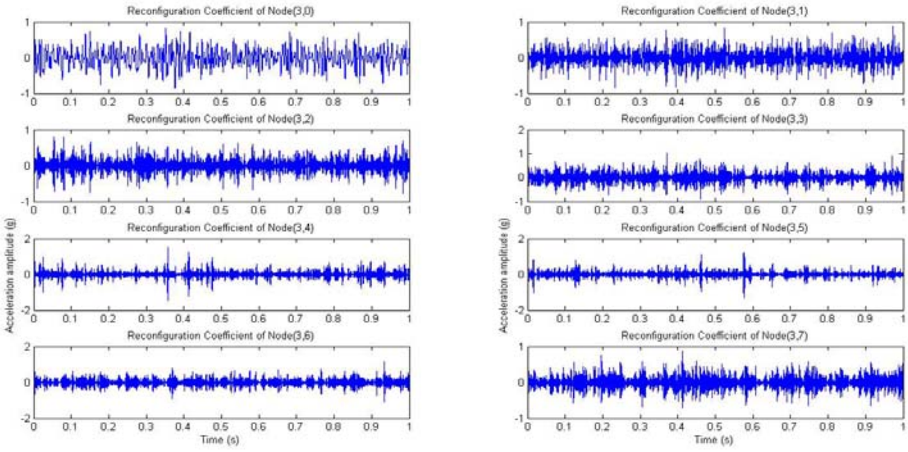
$$e_r = E_{j,r} / E \quad (4)$$

$$T' = \left[e_1, e_2, \cdots, e_r, \cdots, e_{2^j-1} \right] \quad (5)$$

Where T' is the normalized feature vector.



a. Normal state



b. Tooth fracture

Figure 2. Wavelet Packet Reconfiguration Signal

3.3. Attributes Discretization and Reduction based on RST

In order to facilitate computing, six workstates are noted as 1,2,3,4,5,6 respectively and each takes six samples, thirty-six samples in all. First , every sample is decomposed by WPT, the normalized feature vector $T' = \left[e_1, e_2, \cdots, e_r, \cdots, e_{2^j-1} \right]$ is obtained as the condition attributes. Then the universe $U = \{x_1, x_2, \cdots, x_{36}\}$ are constituted by thirty-six samples, and the decision attributes D are made up of six states. Lastly the decision table are constructed by all above, seen in Table 1.

Table 1. Decision Table

U	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	D
x_1	0.6194	0.3040	0.3981	0.4802	0.1170	0.0872	0.2920	0.1682	1
x_2	0.6319	0.3055	0.3886	0.4681	0.1189	0.0877	0.2942	0.1697	1
x_3	0.6422	0.2873	0.3961	0.4571	0.1252	0.0849	0.2976	0.1668	1
x_4	0.6357	0.2953	0.4015	0.4562	0.1331	0.0915	0.2896	0.1715	1
x_5	0.6321	0.2924	0.3864	0.4759	0.1249	0.0859	0.2918	0.1764	1
x_6	0.6398	0.2979	0.3713	0.4769	0.1206	0.0902	0.2958	0.1626	1
x_7	0.4433	0.3844	0.3467	0.3498	0.3765	0.2404	0.3557	0.2952	2
x_8	0.4481	0.3787	0.3484	0.3637	0.3656	0.2353	0.3550	0.2952	2
x_9	0.4387	0.3841	0.3480	0.3604	0.3827	0.2334	0.3402	0.3039	2
x_{10}	0.4295	0.3993	0.3352	0.3627	0.3836	0.2401	0.3497	0.2915	2
x_{11}	0.4280	0.3965	0.3456	0.3629	0.3923	0.2276	0.3406	0.2944	2
x_{12}	0.4381	0.3941	0.3517	0.3600	0.3666	0.2418	0.3436	0.2976	2
x_{13}	0.2453	0.1599	0.2375	0.3761	0.7431	0.2712	0.2217	0.2039	3
x_{14}	0.2454	0.1662	0.2520	0.3737	0.7380	0.2678	0.2241	0.2058	3
x_{15}	0.2569	0.1683	0.2689	0.3759	0.7264	0.2609	0.2315	0.2068	3
x_{16}	0.2465	0.1638	0.2576	0.3874	0.7261	0.2713	0.2309	0.2048	3
x_{17}	0.2491	0.1582	0.2528	0.3641	0.7405	0.2716	0.2270	0.2069	3
x_{18}	0.2491	0.1678	0.2599	0.3844	0.7245	0.2672	0.2383	0.2036	3
x_{19}	0.4801	0.2951	0.3437	0.5760	0.3348	0.1859	0.2222	0.1909	4
x_{20}	0.4867	0.2812	0.3720	0.5558	0.3384	0.1903	0.2291	0.1830	4
x_{21}	0.4871	0.2863	0.3522	0.5689	0.3390	0.1848	0.2198	0.1890	4
x_{22}	0.4990	0.2840	0.3543	0.5737	0.3195	0.1880	0.2178	0.1753	4
x_{23}	0.5046	0.2956	0.3601	0.5438	0.3302	0.1967	0.2257	0.1844	4
x_{24}	0.4893	0.2762	0.3396	0.5885	0.3251	0.1880	0.2139	0.1891	4
x_{25}	0.4833	0.2810	0.3736	0.5190	0.2011	0.1778	0.4050	0.2061	5
x_{26}	0.4824	0.2951	0.3729	0.5020	0.2046	0.1654	0.4231	0.2024	5
x_{27}	0.4848	0.2984	0.3790	0.4918	0.2084	0.1766	0.4129	0.2127	5
x_{28}	0.4754	0.2873	0.3734	0.5134	0.2033	0.1775	0.4171	0.2040	5
x_{29}	0.4915	0.2884	0.3783	0.4870	0.1971	0.1783	0.4279	0.2031	5
x_{30}	0.4901	0.2869	0.3629	0.5023	0.2109	0.1822	0.4151	0.2086	5
x_{31}	0.4056	0.2335	0.4365	0.4457	0.2291	0.1898	0.4748	0.2789	6
x_{32}	0.4166	0.2475	0.4363	0.4437	0.2344	0.1874	0.4653	0.2670	6
x_{33}	0.4216	0.2408	0.4395	0.4439	0.2222	0.1822	0.4605	0.2818	6
x_{34}	0.4227	0.2386	0.4382	0.4449	0.2251	0.1796	0.4682	0.2688	6
x_{35}	0.4083	0.2382	0.4403	0.4627	0.2264	0.1781	0.4597	0.2726	6
x_{36}	0.4087	0.2397	0.4326	0.4481	0.2152	0.1834	0.4792	0.2793	6

For the purpose of realization real-time and online fault diagnosis, the characteristic vector should include as little elements as possible, so the rough sets algorithm is used to reduct features here. First, the improved NaiveScaler algorithm in 2.2 is utilized to discrete Table 1, the key procedures are:

(1) Computing condition equivalence classifications before discretization:

$$X_1 = \{x_1\}, X_2 = \{x_2\}, X_3 = \{x_3\}, X_4 = \{x_4\}, X_5 = \{x_5\}, X_6 = \{x_6\}, X_7 = \{x_7\}, X_8 = \{x_8\}, X_9 = \{x_9\},$$
$$X_{10} = \{x_{10}\}, X_{11} = \{x_{11}\}, X_{12} = \{x_{12}\}, X_{13} = \{x_{13}\}, X_{14} = \{x_{14}\}, X_{15} = \{x_{15}\}, X_{16} = \{x_{16}\}, X_{17} = \{x_{17}\}, X_{18} = \{x_{18}\}, X_{19} = \{x_{19}\},$$
$$X_{20} = \{x_{20}\}, X_{21} = \{x_{21}\}, X_{22} = \{x_{22}\}, X_{23} = \{x_{23}\}, X_{24} = \{x_{24}\}, X_{25} = \{x_{25}\}, X_{26} = \{x_{26}\}, X_{27} = \{x_{27}\}, X_{28} = \{x_{28}\},$$
$$X_{29} = \{x_{29}\}, X_{30} = \{x_{30}\}, X_{31} = \{x_{31}\}, X_{32} = \{x_{32}\}, X_{33} = \{x_{33}\}, X_{34} = \{x_{34}\}, X_{35} = \{x_{35}\}, X_{36} = \{x_{36}\}.$$

Computing decision equivalence classifications and $POS_C(D)$:

$$D_1 = \{x_1, x_2, x_3, x_4, x_5, x_6\}, D_2 = \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}, D_3 = \{x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\},$$
$$D_4 = \{x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}\}, D_5 = \{x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}\}, D_6 = \{x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}\}.$$

$$POS_C(D) = \begin{Bmatrix} x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, \\ x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36} \end{Bmatrix}$$

(2) Calculating all the breakpoints of every attribute and choosing one to discrete the condition attribute, eight breakpoints successively are: 0.47775, 0.28785, 0.3189, 0.20645, 0.27695, 0.1877, 0.3189, 0.20645. For each attribute, if the value is greater than the breakpoint, the discrete value would be “1”, else “0”, so the discrete decision table is got.

(3) Computing $POS_C(D)$ of the discrete table. It is equal to that before discretization, so the discrete decision table is final.

Next the same lines and rows in the discrete decision table are deleted, only seventeen lines are left. The algorithm based on condition equivalence classifications in 2.3 is adopted to delete the redundant attributes. The main steps are:

(1) Obtaining $CORE = \{e_1, e_7\}$ according to the core attributes algorithm;

(2) Computing condition equivalence classifications by core attributes:

$$X_1 = \{x_1, x_2, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}\}, X_2 = \{x_7, x_{31}, x_{32}, x_{33}\}, X_3 = \{x_{13}, x_{14}\}, X_4 = \{x_{25}, x_{26}, x_{27}, x_{28}\}.$$

Computing decision equivalence classifications accordingly:

$$D_1 = \{x_1, x_2\}, D_2 = \{x_7\}, D_3 = \{x_{13}, x_{14}\}, D_4 = \{x_{19}, x_{20}, x_{21}, x_{22}, x_{23}\}, D_5 = \{x_{25}, x_{26}, x_{27}, x_{28}\},$$
$$D_6 = \{x_{31}, x_{32}, x_{33}\}.$$

(3) For X_1 and X_2 can not be classified correctly, it is necessary to search for new attributes in the rest attributes that could distinguish the samples. According to the algorithm, the sets E and F are got: $E = \{e_5\}$, $F = \{e_2, e_3, e_5\}$, $P = E \cap F = \{e_5\}$, so the reduction set is $S = CORE \cup P = \{e_1, e_5, e_7\}$. The decision table after reduction is seen in Table 2.

(4) Calculating S positive region of D. According to 2.1, $POS_S(D) = POS_C(D)$, S is D reduction of C, which verifies the effectiveness and accuracy of the method.

Table 2. Decision Table after Reduction

U	e_1	e_5	e_7	D	U	e_1	e_5	e_7	D
x_1	1	0	0	1	x_{23}	1	1	0	4
x_2	1	0	0	1	x_{25}	1	0	1	5
x_7	0	1	1	2	x_{26}	1	0	1	5
x_{13}	0	1	0	3	x_{27}	1	0	1	5
x_{14}	0	1	0	3	x_{28}	1	0	1	5
x_{19}	1	1	0	4	x_{31}	0	0	1	6
x_{20}	1	1	0	4	x_{32}	0	0	1	6
x_{21}	1	1	0	4	x_{33}	0	0	1	6
x_{22}	1	1	0	4					

4. Results and Discussion

In the light of the relation of the condition attributes and decision attribute in Table 2, the decision rules for fault diagnosis can be drawn as follows:

Rule 1: IF $\{e_1, e_5, e_7\} = \{1,0,0\}$, THEN $D = 1$.

Rule 2: IF $\{e_1, e_5, e_7\} = \{0,1,1\}$, THEN $D = 2$.

Rule 3: IF $\{e_1, e_5, e_7\} = \{0,1,0\}$, THEN $D = 3$.

Rule 4: IF $\{e_1, e_5, e_7\} = \{1,1,0\}$, THEN $D = 4$.

Rule 5: IF $\{e_1, e_5, e_7\} = \{1,0,1\}$, THEN $D = 5$.

Rule 6: IF $\{e_1, e_5, e_7\} = \{0,0,1\}$, THEN $D = 6$.

In order to detect whether the algorithm based on rough sets is reliable, two samples every state are extracted from the original signal for testing, seen in Table 3. Three breakpoints of e_1, e_5, e_7 are still used to discrete Table 3 here, then the decision rules above are used to classify the samples, the results of fault diagnosis are listed in Table 4.

Table 3. Testing Samples

U	e_1	e_5	e_7	D
x_1	0.6404	0.1222	0.2946	1
x_2	0.6370	0.1214	0.2922	1
x_3	0.4285	0.3669	0.3372	2
x_4	0.4292	0.3721	0.3593	2
x_5	0.2520	0.7329	0.2283	3
x_6	0.2555	0.7273	0.2328	3
x_7	0.4845	0.3245	0.2319	4
x_8	0.4848	0.3319	0.2231	4
x_9	0.4817	0.1922	0.4149	5
x_{10}	0.4970	0.2046	0.4132	5
x_{11}	0.4080	0.2269	0.4678	6
x_{12}	0.4169	0.2340	0.4690	6

Table 4. Diagnosis Results

U	e_1	e_5	e_7	Result
x_1	1	0	0	1
x_2	1	0	0	1
x_3	0	1	1	2
x_4	0	1	1	2
x_5	0	1	0	3
x_6	0	1	0	3
x_7	1	1	0	4
x_8	1	1	0	4
x_9	1	0	1	5
x_{10}	1	0	1	5
x_{11}	0	0	1	6
x_{12}	0	0	1	6

All the samples are classified accurately from Table 4, which proves that it is feasible to utilize the method based on wavelet packet transform and rough sets theory to diagnose the gearbox. First of all, WPT is available to the gearbox signals, the variation of energy distribution in each frequence band can reflect the different fault patterns. Second, in this work only one breakpoint is used to discrete the decision table under the condition of keeping classification ability, which verifies that the improved Naïve Scaler algorithm could reduce the complexity of discretization and lay the foundation for attributes reduction. The final reduction set includes three attributes, and the samples not assigned to decision classes by the core properties can be classified accurately through adding one attribute, which certifies that the reduction algorithm based on condition equivalence classifications could improve efficiency. Lastly the characteristic vector should include as little elements as possible in order to realize real-time and online fault diagnosis, the method can delete a lot of redundant features, so it is appropriate to online diagnosis.

5. Conclusion

Because of the poor working environment and lower signal-noise rate, the vibration signal of gearbox is usually non-linear and non-stationary, most of the traditional signal processing methods in core of Fourier Transform are not suitable to deal with the signal, so how to recongnize the faults effectively and quickly is a key issue. A novel method based on wavelet packet transform and rough sets theory is presented for fault diagnosis of gearbox in the paper. First, the fault signals are decomposed into eight bands from low frequency to high frequency by wavelet packet transform, energy characteristics are extracted from each band as the condition attributes in the decision table. Second, an improved NaiveScaler algorithm is put forward to discrete the continuous attribute values in the case of assuring classification ability. A new reduction algorithm based on condition equivalence classifications is proposed to delete the redundant features, which could save computing time and improve the reduction efficiency. Lastly the decision rules were drawn by the reduction table and utilized to test samples. The experimental results show that the method could delete many redundant features and obtain more sensitive fault characteristic parameters. Additionally, research on more efficient reduction algorithms should be done in future.

Acknowledgement

This work is supported by the National Natural Science Fund Project (No. 51175480).

References

[1] Chen Fafa, Tang Baoping, Chen Renxiang. A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm. *Journal of the International Measurement Confederation*. 2013; 46(1): 220-232.
[2] Z Pawlak. Rough Sets. *Computer and Information Science*. 1982; 11(2): 341-356.
[3] Xiaohui Chen. A Rough Neural Network Algorithm for Multisensor Information Fusion. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(2): 747-753.
[4] Jing Siyuan, She Kun, Ali Shahzad. A Universal neighbourhood rough sets model for knowledge discovering from incomplete heterogeneous data. *Expert Systems*. 2013; 30(1): 89-96.
[5] Liu Jianmin, Li Xiaolei, Zhang Xiaoming, et al. Misfire Diagnosis of Diesel Engine based on Rough Set and Neural Network. *Procedia Engineering*. 2011; 16: 224-229.
[6] Zhao Bai-Ting, Jia Xiao-Fen, Zeng Qing-Shuang. Design of rough-neural network fault diagnosis system for turntable. *Journal of Chinese Inertial Technology*. 2012; 20(4): 501-504.
[7] Feng Yongxin, Yang Tao, Yang Wenguang, Jiang Dong xiang. Study of Fault Diagnosis Method for Wind Turbine with Decision Classification Algorithms and Expert System. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 905-910.
[8] Shen Lixiang, Francis E.H.T, Qu Liangsheng, et al. Fault Diagnosis Using Rough Sets Theory. *Computers in Industry*. 2000; 43(2): 61-72.
[9] Zhao Zhihong, Yang Shaopu. Roller bearing fault diagnosis based on wavelet packet transform and sample entropy. *Journal of Vibration, Measurement and Diagnosis*. 2012; 32(4): 640-644.
[10] Keskes Hassen, Braham Ahmed, Lachiri. Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM. *Electric Power Systems Research*. 2013; 97:151-157.
[11] Li Ning, Zhou Rui, Hu Qinghua, Liu Xiaohang. Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine. *Mechanical Systems and Signal Processing*. 2012; 28: 608-621.
[12] R Yasdi. Combining Rough Sets Learning and Neural Learning method to deal with uncertain and imprecise information. *Neurocomputing*. 1995; (7): 61-84.