Intelligence Diagnosis Method Based on Particle Swarm Optimized Neural Network for Roller Bearings

ZUOYI DONG[1], HUAQING WANG[1, *], SHUMING WANG[2], WEI HOU[1],QINGLIANG ZHAO[1]
1 School of Mechanical & Electrical Engineering
Beijing University of Chemical Technology
Chaoyang District, Beijing, 100029
CHINA
2 College of Mechanical Engineering
Jiamusi University
Jiamusi, 156100
CHINA
2011200693@grad.buct.edu.cn, *corresponding author: hqwang@mail.buct.edu.cn,
wsmwsm2000@163.com, 2012200658@grad.buct.edu.cn, zhaoql@mail.buct.edu.cn

Abstract: - This paper presents an intelligent diagnosis approach based on the particle swarm optimized BP (PSO-BP) neural network and the rough sets to detect roller bearings faults and distinguish fault types, using symptom parameters of acoustic emission signals. The rough sets algorithm is used to reduce details of time-domain symptom parameters for the training of the neural network instead of principal component analysis. The PSO-BP neural network, which used for condition diagnosis of roller bearing, can obtain good convergence using the symptom parameters acquired by the rough sets when learning, and can automatically distinguish fault types when diagnosing. Using the PSO-BP neural network can increase the learning rate and the subtracting capability of the neural network. Practical examples are provided to verify the efficiency of the proposed method.

Key-Words: - Intelligence diagnosis, Particle Swarm Optimization, BP neural network, PCA, Rough set, Fault Diagnosis, Roller Bearings

1 Introduction

The rolling bearings are important components in rotating machines, which are wildly used in industrial production. The event of failure, it will cause serious damage to persons and property. The fault diagnosis technology has played a very important role for quality and life of machines. Neural networks (NN) have potential applications in auto-mated detection and diagnosis of machine failures [1]. However, NN usually will converge slowly, when the symptom parameters, input to the first layer of the NN, have the same values in different states. In order to solve these problems and improve the efficiency of the fault diagnosis, this paper proposes a method of condition diagnosis of rolling bearings in rotating machinery using the rough sets and the Particle Swarm Optimized BP neural network to detect faults and distinguish fault types, on the basis of symptom parameters of AE signals. Particle swarm optimization (PSO) has undergone many changes since its introduction in 1995. Researchers have learned about the technique, derived new versions, developed new applications, and published theoretical studies of the effects of the various parameters and aspects of the algorithm. Practical examples of fault diagnosis for rotating machinery have verified that the proposed method is effective [2]. Quite a few works have been done in this field [3, 4].

The traditional diagnosis methods can play a better role of single process, single fault and gradually developed fault of simple systems. However, it has larger limitations for complex process and fault, abrupt faults and highly automated equipment. Intelligence diagnosis methods especially neural network, which not depended on the control object and mathematical model, have a good advantage of solving these problems. Otherwise, as too many of the training sample parameters inputted, it will cause the training with slow convergent speed and low identification accuracy when we use neural network alone. Therefore, attribute reduction usually works before using neural network to make pattern recognition in order to improve recognition accuracy and efficiency. By now, there are many attribute reduction methods and each method has its own advantages and disadvantages[5].

2 Intelligence Diagnosis Method by Particle Swarm Optimized Neural Network

In order to make better use of the advantages of various methods and avoided the limitations of single method at the same time. This paper presents a method for roller bearings, using the way of the combination of the rough sets algorithm, to reduce details of time-domain symptom parameters for the training of the neural network, which process shown as Fig.1.

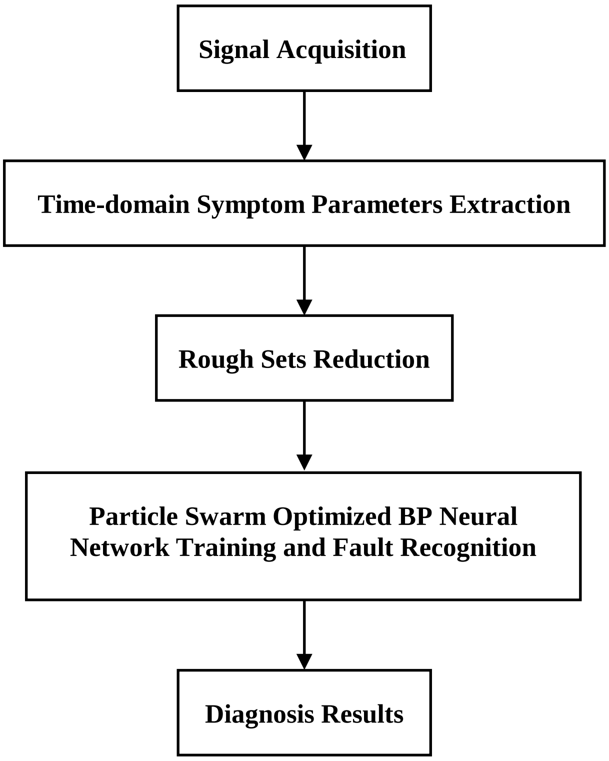


Fig.1. Process of the proposed method

2.1 Symptom Parameters Extraction

For automatic diagnosis, symptom parameters are needed that can sensitively distinguish the fault types. A large set of symptom parameters has been defined in the pattern recognition field [6]. Here, the dimensional, non-dimensional and acoustic emission symptom parameters in the amplitude domain, commonly used for the fault diagnosis of rolling bearing, are considered. Using the normalized signals, the 11 symptom parameters in the amplitude domain are calculated as follows, respectively [7].

$$p_1 = \frac{1}{N}\sum_{i=1}^{N} x_i \text{(Mean value)} \tag{1}$$

$$p_2 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \text{ (Root mean square)} \tag{2}$$

$$p_3 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \text{ (Variance)} \tag{3}$$

$$p_4 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \text{ (Standard deviation)} \tag{4}$$

$$p_5 = \max\left(\left|x_i\right|\right) \text{(Peak value)} \tag{5}$$

$$p_6 = \frac{1}{N}\sum_{i=1}^{N}\left|x_i\right| \text{(Absolute mean value)} \tag{6}$$

$$p_7 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \Bigg/ \frac{1}{N}\sum_{i=1}^{N}\left|x_i\right| \text{(Shape factor)} \tag{7}$$

$$p_8 = \max\left(\left|x_i\right|\right) \Bigg/ \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \text{ (Crest factor)} \tag{8}$$

$$p_9 = \max\left(\left|x_i\right|\right) \Bigg/ \frac{1}{N}\sum_{i=1}^{N}\left|x_i\right| \text{(Impulse factor)} \tag{9}$$

$$p_{10} = \max\left(\left|x_i\right|\right) \Big/ \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left|x_i\right|} \text{ (Clearance factor)} \quad (10)$$

$$p_{11} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 \text{ (Kurtosis value)} \quad (11)$$

2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components [8] [9]. The more Characteristic parameter, the more profound we recognize the system. Number of characteristic parameter contains many interrelated factors. Too much data takes a lot of storage space and processing time. According to the linear mapping principle, we can start from a number of original features, reducing dimensions through mapping method, to construct a handful of new features. The number of principal components is less than or equal to the number of original variables.

This transformation is defined in such a way that the first principal component has as high a variance as possible, and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components [10]. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. PCA technology can overcome the difficulties in modelling such as nonlinear factors. The feature extraction can be done and the PCA model can be established through dimensional reduction of sample features sets of different state test data.

(1).The first step of PCA is data standardization of original feature sets using the formula as

$$P_{ij}^* = \frac{p_{ij} - \overline{p_j}}{\sqrt{\operatorname{var}\left(p_j\right)}}\left(i = 1, 2, \cdots\cdots, n; j = 1, 2, \cdots\cdots, p\right)$$

(2).Then make eigenvalue decomposition to the related matrix P^*P^{*T} of P^*. Ordering C for orthogonal matrix and meet $CC^T = I$, where I is an identity matrix, making $CP^*P^{*T}C^T = \Lambda$.

(3).Define the transformation matrix as

$$\begin{cases} y_1 = c_{11}P_1^* + c_{21}P_2^* + \ldots + c_{p1}P_p^* \\ y_2 = c_{12}P_1^* + c_{22}P_2^* + \ldots + c_{p2}P_p^* \\ \quad\quad \cdots\cdots \\ y_p = c_{1p}P_1^* + c_{2p}P_2^* + \ldots + c_{pp}P_p^* \end{cases} \quad (12)$$

Short as $Y = CP^*$ and $Y = \left[y_1, y_2, y_3, \cdots\cdots, y_p\right]$, $y_1, y_2, y_3, \cdots\cdots, y_p$ each of the principal component is independent.

(4).The contribution rate of the i principal component as $n_i = \frac{\lambda_i}{\sum_{k=1}^{p}\lambda_k}\left(i = 1, 2, \cdots\cdots, p\right)$, the accumulation contribution rate of front m principal components as $\sum_{k=1}^{m} n_k$. Contribution rate is higher, indicating that the corresponding principal component reflects the stronger the ability of integrated information. At last, selecting front m (m<p) principal components to multiply primitive symptom matrix, when accumulation contribution rate reached the specified requirements, and getting a new matrix $Z = P * Y_m$ which can achieve the purpose of dimension reduction of feature set. The original feature set can also be reduced as $Z = \left[z_1, z_2, z_3, \cdots\cdots, z_m\right]$.

2.3 Rough Sets (RS)

Rough set theory [11] is a new mathematical tools which processing the fuzzy and uncertainty knowledge. The main idea is that exporting the decision-making classification rules of the issue, through the knowledge reduction in the premise of keeping the same classification ability. Rough set theory thinks that knowledge is based on the object classification ability [12]. For any information system K = (U, R), $U = \left\{x_1, x_2, x_3, \cdots\cdots, x_n\right\}$ is all of discussing which called domain and $R = \left\{r_1, r_2, r_3, \cdots\cdots, r_n\right\}$ called attribute set. To any $P \subseteq R$ and $P \neq \varnothing$, the elements of the U about the equivalence class of Z constitute an indiscernibility relation signed as ind (Z). The equivalence class formed by Z can be expressed as $U \mid ind\left(Z\right) = \left\{X_1, X_2, \cdots\cdots, X_n\right\}$. We call that set X related to the relationship Z can be precise defined, when X can be expressed as the union set of the equivalence class. It can only be depicted through the upper and lower approximation if not. The upper and lower approximation of set X about Z are expressed as $Z_{-}(X)$ and $Z^{-}(X)$ which meet $Z_{-}(X)=\bigcup(X_i \mid X_i \subseteq X)$ and $Z^{-}(X)=\bigcup(X_i \mid X_i \cap X)\neq\varnothing$. The lower approximation is also called as the positive field of Z and signed as $pos_z(X)$. Supposing that Z and S are two equivalence relation in U ($U \mid S=\{X_1, X_2, \cdots\cdots X_n\}$), the Z positive field of S is signed as $pos_z(S)$ which meet $pos_z(S)=\bigcup_{i=1}^{n} Z_{-}(X_i)$. If there exist $r \in Z$ which meet $pos_z(S)=pos_{z_{-}\{r\}}(S)$, we call that r is S omitted in Z and $Z_{-}\{r\}$ is S relative simplified of Z.

2.4 The Particle Swarm Optimized BP Neural Network

2.4.1 The BP Neural Network

BP neural network is a multilayer feed forward neural network model. The transfer function of the neuron is *S* function and the inputs is a continuously quantity between 0 and 1. It can realize a nonlinear mapping from the input to the output. We call it BP neural network because of the adjustment of the weights with back propagation learning algorithm. The advantages of BP neural network are fast calculation speed and low consumption of memory [13] [14]. The BP neural network model consists of the input layer, hidden layer and output layer. Each layer has several nodes, which are some neurons. The most important factor of the neural network is the number of each layer nodes [15]. The number of input and output layer nodes is determined by the actual problem. At present, the vast majority of neural network model is adopted by the BP network and its change forms in the artificial neural network of practical application. It is a central part of the feed forward network which reflecting the essence of artificial network [16]. After determining the structure of BP neural network, we use the input / output sample to set learning and training, that is, learning and adjustment the weights and biases of the network. It makes the network achieve the given input / output mapping relationship, and complete the system identification. The construction of normal BP neural network is shown as Fig.2.

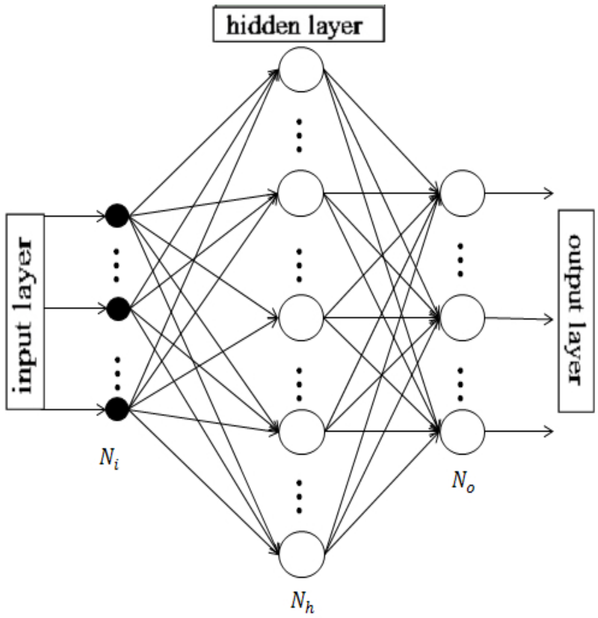


Fig.2. The construction of BPNN

2.4.2 The Particle Swarm Algorithm

The Particle Swarm optimization (PSO) can be described in math as follows: determine an n-dimensional solution space firstly, and particle swarm search in this solution space. Initialize m particles X={X1, X2, …, Xm} and the speed of m particles Vi={vi1, vi2, …, vin} in the solution space randomly, particles adapt their positions to search the new solutions[17] [18]. Every particle can lock its optimal solution- p_{id}, and the optimal solution of the whole particle swarm- p_{gd}. Getting both of the optimal solution, every particle adjusts its speed on the basis of formula below:

$$v_{id}(t+1) = wv_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)) \quad (13)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (14)$$

From adjustable speed formula, we can see that it consists of three parts: its original speed $v_{id}(t)$, the distance to its optimal particle $p_{id} - x_{id}(t)$, and the distance to the optimal particle in the group.

The basic PSO described above has a small number of parameters that need to be fixed. One parameter is the size of the population. This is often set empirically on the basis of the dimensionality and perceived difficulty of a problem[19]. Many parameters in related to particle swarm, the specific meaning are as follows:

(1).Dimension of the particle vector *n*: dimension of problem solution space.

(2).The scale of particle swarm *m*: the number of particle in the group. There is no theory to follow in determining the number of particle swarm, generally determined by experiment many times. Typically the number of particles of 20-40 will be able to solve most of the problems the optimal solution. For some simple questions, 10 particles can achieve quite good results.

(3).Learning factor c_1, c_2. In the speed regulation formula of the particle swarm, the learning factor is very important. c_1, c_2 represents the degree of influence of individual and group experience on the speed adjustment respectively. When c1 is 0, the particle velocity is completely free from the impact of individual experience. At this time, it will soon convergence, but easy to fall into local minimum, not the global optimum. When c2 is 0, the particle velocity is completely free from the impact of group experience. This allows each particle adjust the position according to their own experience, difficult convergence, nor get an effective solution. Therefore, in order to ensure global convergence of particle swarm speed and fitness, we need appropriate c1 and c2.

(4).Inertia weight *w*. It affects the capability of the global search. In general, when w is large, the speed of particles in the iterative process will be large, global search ability will be stronger. Conversely, when the w is small, the speed of particles in the iterative process will be small, local search capability enhanced, and global search capability weakened. Therefore, appropriate inertia weight can make PSO has better global search capability, while also fast convergence rate. In the early PSO, has no inertia weight, means, inertia weight is a fixed value of 1. However, in the process of the convergence of the particle, we need inertia weight to be a dynamic value. At the beginning of the PSO, we hope the particle has a large global search capability, so that allow the particle swarm converges rapidly and find the approximate location of the optimal solution. In late, we need to increase the local search ability of particles, so that make particles search around the optimal solution, and get the optimal solution. Based on this, often use linear decrease inertia weight, the mathematical formula is as follows:

$$w = (w_{\max} - w_{\min}) \times \frac{t_{\max} - t}{t_{\max}} + w_{\min} \qquad (15)$$

$t_{\max}$ is the PSO iterations of maximum number, *t* is the PSO iterations of present. $w_{\max}$ is the inertia weight at the beginning of iterative algorithm, $w_{\min}$ is the inertia weight of maximum number of iterations the algorithm. Typically inertia weight of initial value is 0.9, the minimum value is 0.4.

(5).Random number r_1 and r_2. It used to keep impressing particles the randomness of the flight, to make sure particles can jump out of bondage, enlarging the searching scale.

(6).Error precision *e*. It used to control error range.

(7).Maximum iterations. Setting maximum iterating times can prevent vibration in convergence.

2.4.3 Particle Swarm Optimized Neural Network

The particle swarm algorithm has the advantage of high speed of convergence and powerful searching ability, therefore, the Particle Swarm to Optimize BP neural network was proposed[20]. The optimization flow chart is as shown in Fig.3.

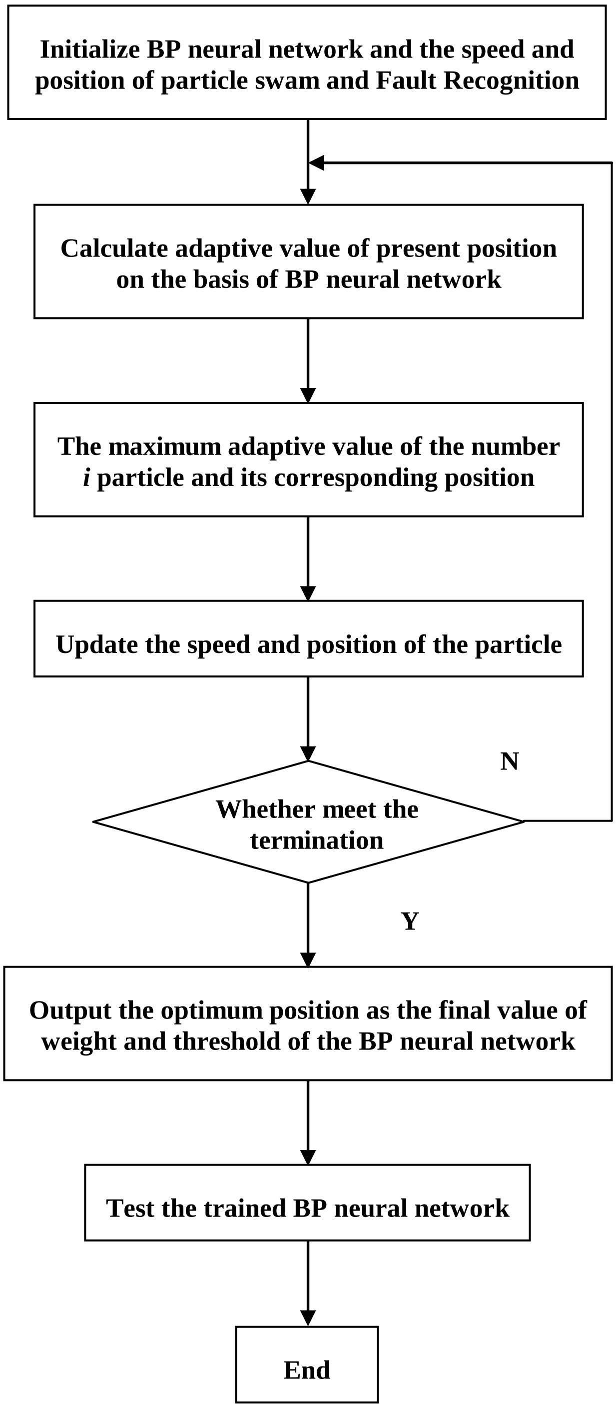


Fig.3. Process of particle swarm optimized BP neural network

As shown in Fig.3, in the process of using the particle swarm to equip BP neural network, first step is initializing the particle swam, different dimension of the particle swarm contributes to different weights and bias,. The number of weights and bias of the neural is equal to the number of dimensions of the particle. The site of the particle concerns the weights and bias of the particle, which contribute to the output of the neural network to satisfy the given precision [21]. Whether the site of the particle is the optimized solution, it is need an accommodated function to decide. The used accommodated function adopts mean-square error of the neural network. The value of weight and threshold of the neural network is determined by the number of unit in the neural network[22] [23][24]. In this paper, the number of unit in the hidden layer is 12. Initialized particle swam iterative optimizes on the basis of velocity formula and position formula. When it reaches the given iteration, stops searching and get the optimum solution. Using the optimum solution, and substituting it into the formula, applies the pattern recognition.

3 Experimental Verification

Fig.4 shows the rotating machine and the bearing for diagnosis. An AE sensor is used to measure AE signals for the bearing diagnosis. Fig.5 shows bearing flaws made artificially for diagnosis purposes. The types of bearing faults are: An outer race flaw (O) shown in Fig.5, (a) and an inner race flaw (I) as in Fig.5, (b). The AE signals are measured at a rotational speed of 600 rpm (10Hz). The sampling frequency is 1 MHz, and the sampling time is 2s or 20 cycles[25][26].

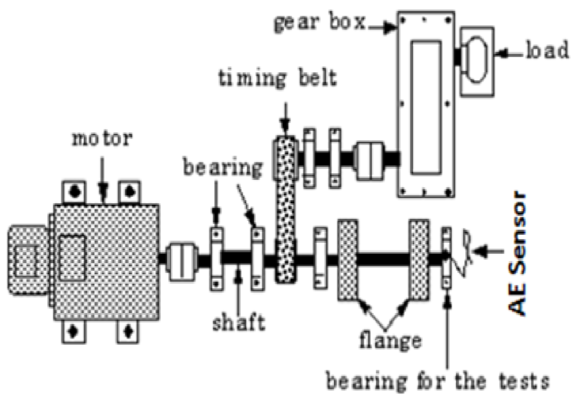


Fig.4. Experiment system of bearing flaw

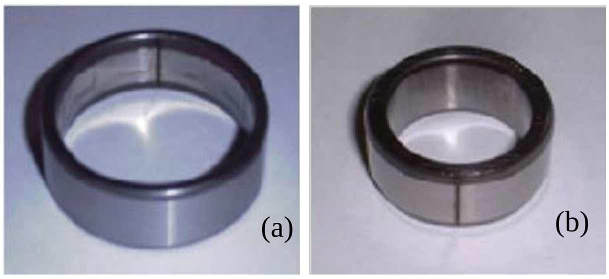


Fig.5. Bearing flaws, (a), outer race flaw (O); (b), inner race flaw (I).

The acoustic emission signals were measured under three conditions which are normal (N), outer (O) and inner (I) race fault. After collecting, the time-domain data are divided into 20 parts each containing 500,000 (5 cycles) samples to calculate the symptom parameters. Then the symptom parameters of each part are calculated respectively. Finally, the original characteristic matrixes of three different conditions are divided into 2 groups averagely and making the data of first and second groups together respectively form the sample and test data. The first step is normalizing the original features data. The partial sample and test data after standardization process are shown in Table 1. Now the raw data should be processed by two ways respectively (PCA and RS), to compare the pros and cons of both approaches.

Table 1. Data after standardization process

	Sample Data			Test Data		
	1	…	30	1	…	30
Mean	0.0005	…	0.0002	-0.0002	…	0.0001
RMS	0.7999	…	0.7700	0.8056	…	0.7700
Var.	0.6398	…	0.5929	0.6490	…	0.5929
SD	0.7998	…	0.7700	0.8056	…	0.7700
MAX	2.7575	…	3.5833	2.7683	…	3.5833
ABS	0.6391	…	0.4784	0.6482	…	0.4784
SF	1.2514	…	1.6093	1.2427	…	1.6093
CF	3.4475	…	4.6536	3.4362	…	4.6536
IF	4.3142	…	7.4890	4.2703	…	7.4890
ClF	5.0959	…	9.8934	5.0147	…	9.8934
KV	2.9345	…	10.37	2.8487	…	10.37

We reduced the data in Table 1 by PCA. We made eigenvalue decomposition to the relevant matrix of it, in order to get the projection matrix which namely the transformation matrix. Then we choose the front 8 principal component as the transformation matrix based on accumulative contribution rate (99.9%). The data after PCA is shown in Table 2. Characteristic matrixes are named as Z_1 to Z_8.

Table 2. Data after PCA process for NN

Training Data				State Data		
	Z_1	…	Z_8	N	O	I
1	5.2811	…	1.998	1	0	0
2	5.1537	…	1.9288	1	0	0
3	4.9335	…	1.8145	1	0	0
…	…	…	…	…	…	…
8	4.7809	…	1.7368	1	0	0
9	5.2597	…	1.963	1	0	0
10	5.3287	…	2.016	1	0	0
11	19.034	…	7.9955	0	1	0
12	19.493	…	8.3609	0	1	0
13	19.815	…	8.5592	0	1	0
…	…	…	…	…	…	…
18	20.046	…	8.4142	0	1	0
19	19.07	…	8.1058	0	1	0
20	16.576	…	7.1874	0	1	0
21	11.457	…	4.6861	0	0	1
22	10.057	…	4.0106	0	0	1
23	11.514	…	4.6277	0	0	1
…	…	…	…	…	…	…
28	10.682	…	4.2888	0	0	1
29	10.964	…	4.5738	0	0	1
30	11.423	…	4.6871	0	0	1

For RS, we make continuous attributes discretization to the decision table, which consists of the sample and test data after reduction, using the method of equidistance. Decision table is as shown in Table 3. And the test data is in Table 4. Here, the section number is 10.

Table 3. Decision table

Training Data						State Data		
	Z_1	Z_2	…	Z_{10}	Z_{11}	N	O	I
1	7	8	…	1	1	1	0	0
2	3	8	…	1	1	1	0	0

3	10	9	…	1	1	1	0	0
…	…	…	…	…	…	…	…	…
8	7	10	…	1	1	1	0	0
9	1	8	…	1	1	1	0	0
10	7	8	…	1	1	1	0	0
11	7	1	…	10	9	0	1	0
12	3	1	…	10	10	0	1	0
13	6	1	…	10	10	0	1	0
…	…	…	…	…	…	…	…	…
18	6	1	…	10	10	0	1	0
19	5	1	…	10	9	0	1	0
20	5	2	…	8	8	0	1	0
21	1	6	…	5	4	0	0	1
22	10	8	…	4	4	0	0	1
23	5	5	…	5	4	0	0	1
…	…	…	…	…	…	…	…	…
28	6	7	…	4	4	0	0	1
29	8	7	…	5	4	0	0	1
30	6	6	…	5	4	0	0	1

Table 4. Test table

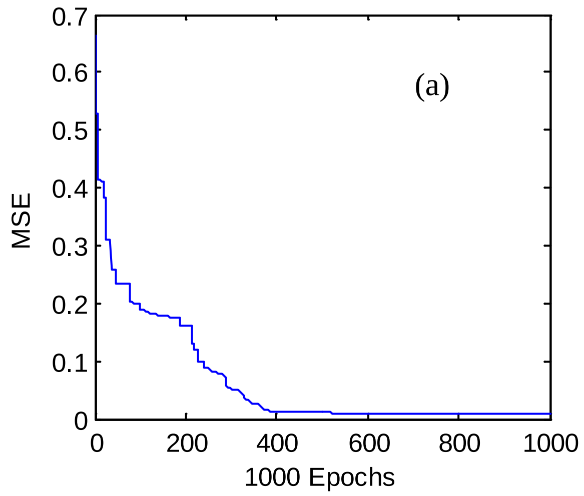
			Test Data					
	Z_1	Z_2	Z_3	Z_4	…	Z_9	Z_{10}	Z_{11}
1	5	8	8	8	…	1	1	1
2	3	8	7	8	…	1	1	1
3	10	10	10	10	…	1	1	1
…	…	…	…	…	…	…	…	…
8	9	10	10	10	…	1	1	1
9	1	9	8	9	…	1	1	1
10	6	9	8	9	…	1	1	1
11	5	2	1	2	…	9	9	8
12	5	3	2	3	…	8	9	8
13	6	2	2	2	…	9	9	8
…	…	…	…	…	…	…	…	…
18	4	1	1	1	…	10	10	10
19	7	1	1	1	…	10	10	10
20	6	1	1	1	…	9	9	9
21	4	7	6	7	…	5	5	4
22	3	8	7	8	…	4	4	4
23	9	7	6	7	…	5	5	4
…	…	…	…	…	…	…	…	…
28	1	7	6	7	…	4	4	4
29	2	7	6	7	…	5	5	4
30	6	7	7	7	…	5	5	4

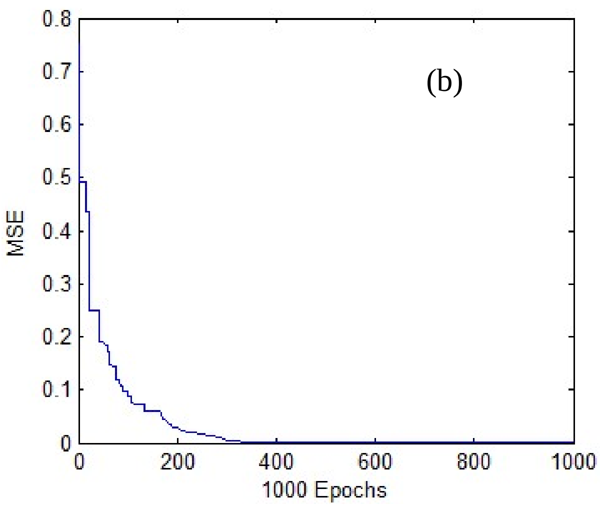
Then observing each condition attributes of the sample data in decision table by line. We can delete the line which attributes completely the same with all the other lines. After that we can delete the column which each attribute values are not all the same between two lines of the others column attribute when it is deleted. The partial training and state data after RS process for PSP-BP are shown in Table 5. Characteristic matrixes are named as Z_1 to Z_{11}.

Table 5. Data after RS process for NN

Training Data						State Data		
	Z_1	Z_5	Z_6	Z_8	Z_{11}	N	O	I
1	7	2	8	1	1	1	0	0
2	3	2	9	1	1	1	0	0
3	10	2	9	1	1	1	0	0
…	…	…	…	…	…	…	…	…
8	7	3	10	1	1	1	0	0
9	1	2	9	1	1	1	0	0
10	7	2	9	1	1	1	0	0
11	7	9	1	9	10	0	1	0
12	3	9	1	9	10	0	1	0
13	6	8	1	9	10	0	1	0
…	…	…	…	…	…	…	…	…
18	6	8	1	10	10	0	1	0
19	5	9	1	9	9	0	1	0
20	5	9	2	8	8	0	1	0
21	1	10	5	5	4	0	0	1
22	10	10	6	6	4	0	0	1
23	5	10	5	6	4	0	0	1
…	…	…	…	…	…	…	…	…
28	6	10	5	5	4	0	0	1
29	8	10	5	5	4	0	0	1
30	6	10	5	5	4	0	0	1

In this paper, the numbers of the input layer, hidden layer and output layer for the PSO-BP neural network are 5, 12 and 3 based on the actual situation and set iterations as 1000. For the data after PCA, training accuracy reaches 0.0085, which process shown in Fig.6, (a). For the data after RS, training accuracy reaches 0.0003, which process shown in Fig.6, (b). Moreover, training accuracy reaches 0.0037 when using PSO-BP, which process shown in Fig.6, (c). From the better result of RS above, we can conclude that RS is more suitable for this condition. So here, we use the RS for reduction, and renounce the use of PCA.





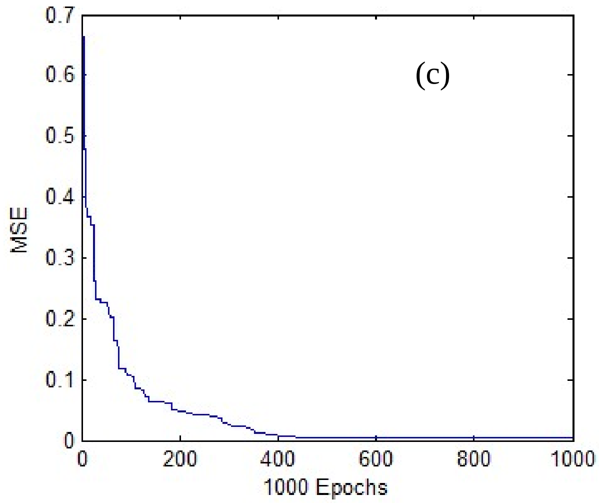


Fig.6. Training process of BPNN, (a) with PCA reduction; (b) with RS reduction; (c) without any reduction.

Table 6 shows the partial diagnosis results for each state by the RS reduction, which the recognition rate is 100%. According to the test results, the probability grades output by the BPNN show the correct judgment in each state. Therefore, the BPNN can precisely distinguish the type of bearing fault, more efficiency on the basis of the symptom parameters of signal after RS process.

Table 6. Verification results

	Possibility grade in each state			State
	N	O	I	
1	1.0069	-0.0116	0.0057	Normal
2	0.9241	-0.0076	0.0515	Normal
3	1.0038	0.0041	9.0056e-06	Normal
…	…	…	…	…
8	1.0016	0.0015	0.0002	Normal
9	0.8046	0.2718	-0.0731	Normal
10	1.0042	-0.0062	0.0033	Normal
11	-0.0126	0.9948	0.0193	Outer race flaw
12	0.0009	0.9535	0.0426	Outer race flaw
13	0.0007	0.9902	0.0052	Outer race flaw
…	…	…	…	…
18	0.0690	1.0313	-0.0599	Outer race flaw
19	0.0018	1.0019	-0.0094	Outer race flaw
20	-0.0108	1.0044	-0.0019	Outer

				race flaw
21	-0.0021	0.0073	1.0018	Inner race flaw
22	-0.0462	-0.0434	1.0028	Inner race flaw
23	0.0011	0.0090	0.9997	Inner race flaw
…	…	…	…	…
28	-0.0644	-0.0495	1.0189	Inner race flaw
29	-0.0225	-0.0087	1.0166	Inner race flaw
30	0.0007	0.0089	0.9996	Inner race flaw

4 Conclusion

The rolling bearings are important components in rotating machines, which are wildly used in industrial production. The intelligent diagnosis technology has played a very important role for quality and life monitoring of machines. To effectively diagnose faults and discriminate fault types for rotating machinery at early stages, this paper proposes an intelligent diagnosis method for rolling bearings using features of acoustic emission signals. The diagnosis approach is constructed on the basis of the rough sets and the PSO-BP. First, use the RS reduce the data. And then, the diagnosis knowledge used for PSO-BP learning can be acquired by the rough sets. It is proved that the effect of RS is better than PCA here by experiment. The PSO-BP can quickly converge when learning, and when diagnosing can quickly and automatically distinguish fault types with high accuracy. This method is suitable for different type of rotating machinery and has been successfully applied to the condition diagnosis of a bearing experiment system. It is significant to apply this method in equipment's bearing failure.

5 Acknowledgements

This project is supported by National Natural Science Foundation of China (Grant No. 51375037, 51075023) National Program on Key Basic Research Project (Grant No. 2012CB026000) and Program for New Century Excellent Talents in University (NCET-12-0759), and SRF for ROCS, SEM.

References

[1] Samanta, B. and Al-Balushi, K. R., Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features, *Mechanical Systems and Signal Processing*, Vol.17, 2003, pp. 317-328.

[2] Riccardo P., James K. and Tim B., Particle Swarm Optimization, *Swarm Intell*, Vol.1, 2007, pp. 33-57.

[3] Mahamed G.H., Andries P. and Ayed S., Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification, *Transactions on Engineering, Computing and Technology*, Vol.9, 2005, pp. 199-204.

[4] Van der Merwe, D. W. and Engelbrecht, A. P., Data Clustering Using Particle Swarm Optimization, *Proceedings of IEEE Congress on Evolutionary Computation 2003 (CEC 2003)*, Canbella, 2003 , pp. 215-220.

[5] Hou Wei, Zuoyi Dong and Huaqing Wang, Research of Diagnosis Method for Fault Bearings Based on Particle Swarm Optimization BP Neural Network, *Fifteenth National Equipment Monitoring and Diagnostic Conference Papers*, Vol.1, 2012, pp. 565-567.

[6] Fukunaga K., *Introduction to Statistical Pattern Recognition*, Academic Press, 1972.

[7] Jia Pan, Huaqing Wang, Feng Wang, Jianfeng Yang and Wenbin Liu, Intelligence Diagnosis Method for Roller Bearings Using Features of AE Signal, *Journal of Physics: Conference, 25th International Congress on Condition Monitoring and Diagnostic Engineering*, Vol.364, No.1, 2012, pp. 1-8.

[8] El-Bakry, Hazem M, Mamoon and Mamoon H, Integrating Neural Networks and PCA for Fast Covert Surveillance, *Proceedings of the 4th WSEAS International Conference on Computational Intelligence, CI` 10*, 2010, pp. 161-167.

[9] Xiufen Fang, Guisong Liu and Tingzhu Huang, A Principal Components Analysis Neural Gas Algorithm for Anomalies Clustering, *WSEAS Transactions on Systems*, Vol.9, No.1, 2010, pp. 42-51.

[10] Minghai Yao, Xinyu Qu and Qinlong Gu, Hand Gesture Recognition Based on Online PCA with Adaptive Subspace, *Proceedings of the 10th WSEAS International Conference on Multimedia Systems and Signal Processing, MUSP '10*, 2010, pp. 211-216.

[11] Huaqing Wang and Peng Chen, Condition Diagnosis of Blower System Using Rough Sets

and A Fuzzy Neural Network, *WSEAS Transactions on Business And Economics*, Vol.5, 2008 , pp. 66-71.

[12] Weizheng Zhang, Yanming Li and Lanjun Yang, The Fault Diagnosis of Power Transformer Based on Rough Set Theory, *WSEAS Transactions on Circuits and Systems*, Vol.4, No.8, 2005, pp. 912-916.

[13] Stastny Jiri and Skorpil Vladislav, Neural Networks Learning Methods Comparison, *WSEAS Transactions on Circuits and Systems*, Vol.4, No.4, 2005, pp. 325-330.

[14] Jatmiko Wisnu, Nulad W. P. , Elly M. I., Setiawan I. Made Agus and Mursanto P., Heart Beat Classification Using Wavelet Feature Based on Neural Network, *WSEAS Transactions on Systems*, Vol.10, No.1, 2011, pp. 17-26.

[15] Yonglin Wang, Yan Liu and Shengbing Che, Research on Pattern Recognition Based on BP Neural Network, *Advanced Research on Material Engineering*, Vol.282-283, 2011, pp. 161-164.

[16] Quan Zhang, Nanyu Chen, Jun Huang and ZhijunMeng, Application of Expert System Fuzzy BP Neural Network in Fault Diagnosis of Piston Engine, *Proceedings - 2012 International Conference on Computer Science and Electronics Engineering, ICCSEE*, Vol.3, 2012, pp. 604-607.

[17] Kalos Alex, Discrete Particle Swarm Optimization for Automating The Design of Multivariate Autoregressive Neural Networks, *WSEAS Transactions on Information Science and Applications*, Vol.2, No.11, 2005, pp. 1807-1815.

[18] Huaching Chen, Donghui Guo and Hsuanming Feng, Fuzzy Embedded Mobile Robot Systems Design through The Evolutionary PSO Learning Algorithm, *WSEAS Transactions on Systems*, Vol.10, No.8, 2011, pp. 259-269.

[19] Tuppadung Yutthapong and Kurutach Weasak, Particle Swarm Optimization Application for Phase Balancing, *WSEAS Transactions on Systems*, Vol.5, No.8, 2006, pp. 1976–1981.

[20] Neri F., Empirical investigation of word-of-mouth phenomena in markets: a software agent approach, *WSEAS Transactions Computers*, Vol.4, No.8, 2005, pp. 987-994.

[21] Jianyan Tian, Wenting Li, Shuangqing Yang and Yanfei Hou, The Research on Neural Network Predictive Model for Hot Rolling Strip Width Based on Improved Particle Swarm Optimization, *Journal of Information and Computational Science*, Vol.8, No.16, 2011, pp. 4237-4335.

[22] Kexu Zou, Haodong Ma, Hongzheng Fang and Dawei Yi, Study of Prognostics for Spacecraft Based-on Particle Swarm Optimized Neural Network, *Prognostics System Health Management Conference, PHM-Shenzhen*, 2011, pp. 1-5.

[23] Guojin Tan, Hanbing Liu, Yongchun Cheng, Bin Liu and Yin Zhang, Prediction Method for The Deformation of Deep Foundation Pit Based on Neural Network Algorithm Optimized By Particle Swarm, *Proceedings International Conference on Transportation, Mechanical, and Electrical Engineering, TMEE*, 2011, pp. 1407-1410.

[24] Youpeng Zhang and Hongsheng Su, Turbo-Generator Vibration Fault Diagnosis Based on PSO-BP Neural Networks, *WSEAS Transactions on Systems and Control*, Vol.5, No.1, 2010, pp. 37-47.

[25] Zuoyi Dong, Hongwei Qu, Qingliang Zhao and Huaqing Wang, Confirmatory Experiment of Relation of Influence Coefficient between Multiplane and Static-Couple Methods, *Journal of Vibration and Shock*, Vol.31, 2012, pp. 469-471.

[26] Jing Cui, Huaqing Wang, Fei Xiao and Zuoyi Dong, Finite Element Modeling and Analysis of Rolling Bearing Based on Explicit dynamics, *Applied Mechanics and Materials*, Vol.397-400, 2013, pp. 589-592.