Intelligence diagnosis method for roller bearings using features of AE signal

J Pan\(^1\), H Q Wang\(^{1,2}\), F Wang\(^1\), J F Yang\(^1\) and W B Liu\(^1\)

\(^1\) School of Mechanical & Electrical Engineering, Beijing University of Chemical Technology,
15 Beisanhuan East Road, ChaoYang district, Beijing, 100029, P. R. China

E-mail: wanghq_buct@hotmail.com

Abstract. Rolling bearings are important components in rotating machines, which are wildly used in industrial production. The fault diagnosis technology plays a very important role for quality and life of machines. Based on symptom parameters of acoustic emission (AE) signals, this paper presents an intelligent diagnosis method for roller bearings using the principal component analysis, rough sets, and BP neural network to detect faults and distinguish fault types. The principal component analysis and the rough sets algorithm are used to reduce details of time-domain symptom parameters for training the BP neural network. The BP neural network, which is used for condition diagnosis of roller bearings, can obtain good convergence using the symptom parameters acquired by the principal component analysis and the rough sets during learning, and automatically distinguish fault types during diagnosing. Practical examples are provided to verify the efficiency of the proposed method.

1. Introduction

Condition monitoring of rotating machinery is important in terms of system maintenance and process automation. In industrial applications, rolling bearings are considered as critical mechanical components and a defect in such a bearing will cause malfunction and even lead to catastrophic failure of the machinery. Bearing failure is one of the foremost causes of failures in rotating machinery. Therefore, bearing failures have received much attention in fault diagnosis because they represent an area where much can be gained from the early detection of faults.

Different methods have been studied in the last two decades. Among them, vibration measurements and analyses are widely applied. Quite a few works have been done in this field \([1-2]\). However, acoustic emission (AE) can be used as a non-destructive testing method to monitor mechanical events and processes in rotating machinery and diagnose faults during operation. One major advantage of using AE monitoring and diagnosis is that the signal has a very high signal-to-noise ratio because the signal frequency is very high, usually in the range of 100 kHz - 1 MHz. Acoustic emission is the phenomenon whereby transient elastic waves generate due to a rapid release of strain energy from localized sources inner or on the surface of material. Several applications of the acoustic emission technology for machine condition monitoring have been presented \([3-4]\).

Although fault diagnosis of rolling bearings is often artificially carried out using time or frequency analysis of AE signals, a reliable, fast automated diagnosis method is still needed. Neural networks...
(NN) have potential applications in automated detection and diagnosis of machine failures [5]. However, NN will usually converge slowly, when the symptom parameters that are input to the first layer of the NN have the same values in different states. To solve these problems and improve the efficiency of fault diagnosis, this paper proposes a method of condition diagnosis of rolling bearings in rotating machinery using the principal component analysis, the rough sets, and the BP neural network to detect faults and distinguish fault types on the basis of symptom parameters of AE signals. Practical examples of fault diagnosis for rotating machinery have verified that the proposed method is effective.

2. Intelligence diagnosis of rolling bearing by AE signals

Traditional diagnosis methods and theory can play a better role of single process, single fault, and gradually developed fault of simple systems. However, they have larger limitations for complex processes and faults, abrupt faults, and highly automated equipment. Intelligence diagnosis methods especially neural network, which do not depend on the control object and mathematical model, have an advantage of solving these problems. Otherwise, inputting too many training sample parameters will cause slow convergent speed and low identification accuracy when the neural network is used alone. Therefore, attribute reduction usually works before the neural network is used to make pattern recognition to improve recognition accuracy and efficiency. There are many attribute reduction methods by now. Each method has its own advantages and disadvantages. To make better use of the advantages of various methods and avoid the limitations of single method at the same time, this paper presents a method for roller bearings, combination of the principal component analysis and the rough sets algorithm, to reduce details of time-domain symptom parameters for training the neural network, as shown in figure 1.

![Figure 1. Process of the proposed method.](image)

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 15 symptom parameters in the amplitude domain are calculated as follows, respectively.
$p_1 = \frac{1}{N} \sum_{i=1}^{N} x_i$ (Mean value)  

(1)

$p_2 = \left[ \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right]^{1/2}$ (Root mean square)  

(2)

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

(3)

$p_4 = \left[ \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \right]^{1/2}$ (Standard deviation)  

(4)

$p_5 = \max \{ |x_i| \}$ (Peak value)  

(5)

$p_6 = \frac{1}{N} \sum_{i=1}^{N} |x_i|$ (Absolute mean value)  

(6)

$p_7 = \left[ \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right]^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} |x_i| \right)^{-1}$ (Shape factor)  

(7)

$p_8 = \max \{ |x_i| \left( \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right)^{-1/2} \}$ (Crest factor)  

(8)

$p_9 = \max \{ |x_i| \left( \frac{1}{N} \sum_{i=1}^{N} |x_i| \right)^{-1} \}$ (Impulse factor)  

(9)

$p_{10} = \max \{ |x_i| \left( \frac{1}{N} \sum_{i=1}^{N} (|x_i|)^{1/2} \right)^{-1} \}$ (Clearance factor)  

(10)

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

(11)

$p_{12} = \max \{ x_i \}$ (Amplitude)  

(12)

where $p_{13}$ is the rise time (RT), crossing the threshold to the largest signal amplitude for the first time; $p_{14}$ is the ringing counting; and $p_{15}$ is the duration time (DT), time interval between the first and the last time crossing the threshold.

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. 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 is orthogonal to the preceding components.
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. The first step of PCA is data standardization of original feature sets using the formula as

$$ p_{ij}^* = \frac{p_{ij} - \bar{p}_j}{\sqrt{\text{var}(p_j)}} \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, p) $$

Then, making eigenvalue decomposition to the related matrix $P^T P$ of $P^*$. Let $C$ be orthogonal matrix and meet $CC^T = I$, where $I$ is an identity matrix, $CP^T C^T = A$. Defining the transformation matrix as $Y = CP^*$ and the contribution rate of the $i$ principal component as

$$ n_i = \sum_{k=1}^{p} \frac{\lambda_k}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \ldots, p) $$

the accumulation contribution rate of the first $m$ principal components as

$$ \sum_{i=1}^{m} n_i $$

Finally, the first $m$ ($m < p$) principal components are selected to multiply the primitive symptom matrix, when the accumulation contribution rate reaches the specified requirements, and getting a new matrix $Z = P*Y_m$ which can achieve dimension reduction of feature set. The original feature set can also be reduced as $Z = [z_1, z_2, z_3, \ldots, z_m]$.

2.3. Rough Sets (RS)

Rough set theory [7] is a new mathematical tool that processes the fuzzy and uncertainty knowledge. The main idea is to export 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. For any information system $K = (U, R)$, $U = \{x_1, x_2, x_3, \ldots, x_n\}$ is all of discussion called domain and $R = \{r_1, r_2, r_3, \ldots, r_s\}$ called attribute set. To any $P \subseteq R$ and $P \neq \emptyset$, the elements of $U$ about the equivalence class of $Z$ constitute an indiscernibility relation signed as $ind(Z)$. The equivalence class formed by $Z$ can be expressed as $U | ind(Z) = \{X_1, X_2, \ldots, X_n\}$. Set $X$ related to the relationship $Z$ can be precisely 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 $Z_+(X) = \cup (X | X \subseteq X)$ and $Z_-(X) = \cup (X | X \cap X) \neq \emptyset$. The lower approximation is also called the positive field of $Z$ and signed as $\text{pos}_z(X)$. Supposing that $Z$ and $S$ are equivalence relation in $U$ ($U | S = \{X_1, X_2, \ldots, X_n\}$), the $Z$ positive field of $S$ is signed as $\text{pos}_z(S)$, which meet $\text{pos}_z(S) = \bigcup_{i=1}^{n} Z_+(X_i)$. If there exists $r \in Z$, which meet $\text{pos}_z(S) = \text{pos}_{z \cup r}(S)$, $r$ is $S$ omitted in $Z$, and $Z \backslash \{r\}$ is $S$ relatively simplified of $Z$.

2.4. BP Neural Network (BPNN)

BP neural network is a multilayer feed forward neural network model. The transfer function of the neuron is $S$ function and the inputs are a continuously quantity between 0 and 1, which can realize a nonlinear mapping from the input to the output. It is called 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. The BP neural network model
consists of the input, hidden, and output layers. Each layer has several nodes, which are some neurons. The most important factor of the neural network is the number of nodes of each layer. The number of nodes of input and output layers 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 that reflects the essence of artificial network. The construction of normal BP neural network is shown in figure 2.

![Figure 2. Structure of BPNN.](image)

3. Experimental Verification

Figure 3 shows the rotating machine and the bearing for diagnosis. An AE sensor is used to measure AE signals for bearing diagnosis. Figure 4 shows bearing flaws artificially made for diagnosis. These types of bearing faults are an outer race flaw (O) and an inner race flaw (I). The AE signals are measured at a rotation speed of 600 rpm (10Hz). The sampling frequency is 1 MHz, and the sampling time is 2 s or 20 cycles.

![Figure 3. Experimental system of bearing flaw.](image)

![Figure 4. Bearing flaws.](image)

The experiments collected acoustic emission signals in three conditions: normal (N), outer (O), and inner (I) race faults. The time-domain data are divided into 20 parts. Each part contains 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 the data of the first and second groups form the sample and test
data, respectively. The first step of PCA is to normalize the original features data. The partial sample and test data after standardization process are shown in table 1.

| Sample Data | Test Data |
|-------------|-----------|
| 1 | 2 | 3 | ... | 29 | 30 | 1 | 2 | 3 | ... | 29 | 30 |
| Mean | 0.417 | -1.15 | 1.649 | ... | 0.789 | 0.074 | 0.097 | -0.800 | 2.126 | ... | -1.37 | 0.304 |
| RMS | 0.815 | 0.834 | 1.105 | ... | 0.517 | 0.137 | 0.711 | 0.643 | 1.308 | ... | 0.353 | 0.464 |
| Var. | 0.822 | 0.845 | 1.183 | ... | 0.470 | 0.047 | 0.697 | 0.615 | 1.453 | ... | 0.280 | 0.406 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| RT | 0.700 | 0.383 | 1.321 | ... | -0.84 | -0.12 | -0.86 | -0.772 | -0.60 | ... | 1.581 | -0.88 |
| DT | -0.34 | -2.28 | -0.28 | ... | -0.14 | -3.03 | -0.90 | -0.718 | 0.447 | ... | -0.79 | -0.75 |

After getting the standardization characteristic matrix, the second step is to make eigenvalue decomposition to its relevant matrix, to get the projection matrix, i.e., the transformation matrix. The first eight principal components are chosen as the transformation matrix based on accumulative contribution rate (99.9%). Finally, the original features matrix is multiplied with the transformation matrix to get the characteristic matrix after reduction, which are named as Z1 to Z8. The partial sample and test data after the PCA process are shown in table 2.

| Sample Data | Test Data |
|-------------|-----------|
| Z1 | Z2 | Z3 | ... | Z7 | Z8 | Z1 | Z2 | Z3 | ... | Z7 | Z8 |
| 1 | -11.57 | -0.192 | 1.294 | ... | 16.89 | -3.433 | -8.877 | -1.898 | 1.004 | ... | 10.95 | -3.80 |
| 2 | -10.85 | -0.616 | 1.216 | ... | 15.53 | -3.416 | -11.67 | -0.140 | 1.122 | ... | 16.89 | -3.50 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3 | -23.39 | 2.782 | 1.515 | ... | -32.3 | -8.115 | -36.46 | 11.95 | 2.043 | ... | 62.82 | -5.59 |

For RS, make continuous attributes discretization as the decision table, which consists of the sample and test data after reduction, using the method of equidistance. The section number is 10. Then, observe each condition attributes of the sample data on decision table through line. Delete the line whose attributes are completely the same as all the other lines. After that, delete the column that each attribute value is not all the same between two lines of the other column attributes when it is deleted. The partial training and state data after the RS process for NN are shown in table 3.

| Training Data | State Data |
|---------------|------------|
| Z1 | Z4 | Z6 | Z7 | Z8 | N | O | I |
| 1,2,3,6 | 1 | 10 | 1 | 1 | 3 | 1 | 0 | 0 |
| 4,9,10 | 1 | 9 | 1 | 1 | 2 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 29 | 2 | 8 | 2 | 2 | 1 | 0 | 0 | 1 |

In this paper, the numbers of input, hidden, and output layers for the BP neural network are 5, 10, and 3 based on the actual situation. When the precision is set to 0.03, training accuracy reaches the requirements at 1851 as shown in figure 5(a). Moreover, training accuracy reaches the requirements at 3026 using BPNN without reduction as shown in figure 5(b). Table 4 shows the partial diagnosis
results for each state where the recognition rate is 100%. According to the test results, the probability grade output by the BPNN show the correct judgment in each state. Therefore, the BPNN can precisely distinguish the type of bearing faults, more efficiency on the basis of the symptom parameters of AE signal after the PCA and RS processes.

![Figure 5. Training process of BPNN.](image)

**Table 4. Verification results.**

| Possibility grade in each state outputted by BPNN | State judgment |
|-----------------------------------------------|----------------|
| N    | O         | I            |                |
| 1    | 0.77271   | -0.0498      | 0.3471         | Normal        |
| ...  | ...       | ...          | ...            | ...           |
| 11   | 0.17869   | 1.127        | -0.09647       | Outer race flaw|
| ...  | ...       | ...          | ...            | ...           |
| 21   | 0.090933  | 0.019154     | 0.84119        | Inner race flaw|
| ...  | ...       | ...          | ...            | ...           |
| 30   | 0.007907  | 0.10804      | 0.95153        | Inner race flaw|

4. Conclusions

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 AE signals. The diagnosis approach is constructed on the basis of the principal component analysis, rough sets, and BPNN. The diagnosis knowledge used for BPNN learning can be acquired by the principal component analysis and rough sets. The BPNN can quickly converge during learning, and quickly and automatically distinguish fault types with high accuracy during diagnosing. This method is suitable for different types of rotating machinery and has been successfully applied to the condition diagnosis of a bearing experimental system.
Acknowledgements
This project study was supported by National Program on Key Basic Research Project (Grant No. 2012CB026000), National Natural Science Foundation of China (Grant No. 51075023, 51135001), and Fundamental Research Funds for the Central Universities (Grant No. ZZ1124) and SRF for ROCS, SEM.

References
[1] Wang H Q, Guo Y W, Yang J F, Song L Y, Pan J, Chen P and Yuan H F 2011 Adv. Mater. Res. 199-200 1020-23
[2] Ho D. and R.B. Randall 2000 Mech. Syst. signal Pr. 14 763-88
[3] Mba D. and Raj B. K. N. 2006 Shock Vib. Dig. 38 3-16
[4] Yuan H F, Wang P and Wang H Q 2011 Adv. Mater. Res. 199-200 895-8
[5] Samanta B. and Al-Balushi K. R. 2003 Mech. Syst. signal Pr. 17 317–28
[6] Fukunaga K 1972 Introduction to Statistical Pattern Recognition (London: Academic Press)
[7] Wang H Q and Chen P 2008 WSEAS Trans. Bus. Econ. 5 66-71