Rolling Bearing Fault Diagnosis Method Based on Principal Components Analysis and Probabilistic Neural Network

Yun Feng Ni¹, Dong Dong Ban¹

¹College of Communication and Information, Xi’an University of Science and Technology, Xi’an, Shaanxi, 710054, China

Abstract. In this paper, a rolling bearing fault diagnosis method based on PCA and improved PNN network is proposed to solve the problems of high dimension, high redundancy, nonlinearity and nonstationarity of rolling bearing data. Firstly, the principal components analysis (PCA) algorithm is used to extract the feature information of the original data and obtain the principal component information after dimension reduction. Then the principal component information is sent as a feature to the probabilistic neural network (PNN) for training and outputting diagnosis results. The method is verified using Case Western bearing datasets. Through simulation comparison of this method and BP neural network method, the experimental results show that the method proposed in this paper is more accurate in bearing fault diagnosis.

1. Introduction
Rolling bearings are widely used in rotary machinery and are one of its important components. Data show that about one third of the rotating mechanical failure is caused by bearing failure[1]. Once the rolling bearing is damaged, it will reduce the life of the equipment and cause huge economic losses. Therefore, it is of great significance to study the method of fault diagnosis of rolling bearings[2]. In this paper, the vibration detection method is adopted for fault diagnosis of rolling bearing. The vibration signal of the bearing is collected by the vibration sensor, and the condition and fault of the bearing are judged through the analysis and processing of the data.

The vibration signal of rolling bearing has the characteristics of high dimension, high redundancy, nonlinearity and nonstationarity. Traditional data processing methods such as Fourier transform and wavelet transform are difficult to analyze data effectively. Principal components analysis is a multivariate statistical analysis method that projects high-dimensional sample data information into low-dimensional subspaces. The low-dimensional shadows obtained in subspaces can represent the main information of sample data[3]. The combination of principal components analysis and deep belief network improves the accuracy of fault diagnosis of deep belief network[4]. The fault diagnosis method of steam turbine flow passage based on improved EMD and PNN ensures the accuracy of diagnosis and greatly shortens the training time[5]. Fault diagnosis method based on improved EMD and PSO-SVM can accurately identify and diagnose rolling bearing faults[6].

In this paper, a rolling bearing fault diagnosis method based on PCA and PNN is proposed. PCA is used to extract features and reduce dimensions of bearing vibration data, and the obtained feature information is sent to PNN model for training, so that the fault types of rolling bearings can be identified quickly and accurately.
2. Theoretical basis

2.1 Principal component analysis algorithm

PCA converts high-dimensional sample data into orthogonal uncorrelated random principal components by orthogonal transformation, and these principal components carry a large amount of feature information, which can reflect the running state of the rolling bearing[3]. PCA can eliminate redundancy and noise in fault data and effectively reduce dimensionality of large amount of data on the premise of ensuring accuracy[7]. The steps of PCA data dimension reduction are as follows:

- There are n groups of samples in the original data, and each sample has p variables, which form the data matrix X of the order \( p \times n \).

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1p} \\
  x_{21} & x_{22} & \cdots & x_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\]  

(1)

- Each data in the original sample data matrix X is standardized by z-score;

\[
x_{ej}^* = \frac{x_{ej} - \overline{x}_j}{\sigma_j}
\]

(2)

In this equation:

\[
\overline{x}_j = \frac{1}{N} \sum_{a} x_{aj} \quad j = (1, 2, \cdots, p)
\]

(3)

\[
\sigma_j^2 = \frac{1}{N} \sum_{a} (x_{aj} - \overline{x}_j)^2 \quad j = (1, 2, \cdots, p)
\]

(4)

- The standardized matrix is \( \tilde{X} \). The correlation coefficient \( r_{ij} \) between any two elements in the matrix can be calculated as follows:

\[
r_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)^2 \sum_{k=1}^{n} (x_{kj} - \overline{x}_j)^2}}
\]

(5)

From the equation (5), the correlation coefficient matrix R of the matrix can be obtained, namely the covariance matrix;

\[
R = \begin{bmatrix}
  r_{11} & r_{12} & \cdots & r_{1p} \\
  r_{21} & r_{22} & \cdots & r_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{p1} & r_{p2} & \cdots & r_{pp}
\end{bmatrix}
\]

(6)

- According to equation (7), the eigenvalues and corresponding eigenvectors of the correlation coefficient matrix R are calculated;

\[
[\Lambda \vec{E} - R] = 0
\]

(7)

- Sort the eigenvalues in the order from largest to smallest, which is: \( \lambda_1, \lambda_2, \cdots, \lambda_p \). The corresponding eigenvectors are recorded by \( \vec{a}_1, \vec{a}_2, \cdots, \vec{a}_p \). Calculating the contribution rate of principal component and cumulative contribution rate:

\[
\frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k} \quad (i = 1, 2, \cdots, p)
\]

(8)
\[
\sum_{k=1}^{p} \lambda_k = \sum_{i=1}^{p} \lambda_i = 1 \quad (i = 1, 2, \cdots, p)
\]

In practical application, the eigenvalues \( \lambda_1, \lambda_2, \cdots, \lambda_m \) which have a cumulative contribution rate of 85%, are taken to represent the principal components 1, 2, ..., and \( m(m \leq p) \).

2.2 Probabilistic neural network model

Probabilistic Neural Network is a simple structure and simple algorithm design proposed by Dr. D. F. Spechtin 1989. It is widely used in pattern classification. It is a feedforward neural network model developed from RBF. The criterion of probabilistic neural network is Bayesian minimum risk criterion [8]. The probabilistic neural network model has four layers of network continuation structure as shown in figure 1.

![PNN network continuation structure](image)

**Figure 1. PNN network continuation structure**

Design process of probabilistic neural network:

- After normalizing the trained sample matrix, it is divided into training samples and test samples;
- The normalized sample data is sent to the PNN network model to establish a basic discriminant network and classification identification network.
- The training samples and test samples are tested, and the test results and recognition rate are output to verify the accuracy of PNN model.

3. Fault diagnosis scheme pca and pnn

The rolling bearing fault diagnosis method based on PCA and PNN mainly includes PCA dimension reduction module and PNN diagnostic module. Firstly, the PCA module is used to extract the feature and reduce the dimension of the rolling bearing fault data, and then the most characteristic parameters after dimension reduction are trained and tested by probabilistic neural network model. The flow chart of fault diagnosis is shown in figure 2.
The specific diagnostic procedures are as follows:

- Vibration signals of rolling bearing under normal conditions and different faults (such as wear of inner ring, rolling body and outer ring) are collected as failure symptom signals.
- PCA was used to standardize the original sample data and calculate correlation coefficients, eigenvalues and eigenvectors.
- The contribution rate and cumulative contribution rate of principal components are calculated by eigenvalues, as shown in table 1. The eigenvalues with a cumulative contribution rate of more than 85% were selected to determine the number of principal components. From this, the eigenvector can be obtained.

### Table 1. Principal component contribution rate and cumulative contribution rate (%)

| Principal component | 1     | 2     | 3     | 4     | 5     | 6     | 7     |
|---------------------|-------|-------|-------|-------|-------|-------|-------|
| Contribution rate   | 23.67 | 21.12 | 14.04 | 13.36 | 12.11 | 9.29  | 6.42  |
| Cumulative contribution rate | 23.67 | 44.73 | 58.75 | 72.09 | 84.18 | 93.58 | 100   |

- Take the principal component eigenvector obtained as the input vector of PNN, the working state of the rolling bearing was coded, and the state code was set as the output category vector to establish the PNN network model.
- PNN network training is conducted on test samples and training samples, and the working state of the corresponding rolling bearing is judged according to the output result.

### 4. Experimental analysis

#### 4.1 Experimental settings

Benchmark data of Case Western Reserve University were used for experimental bearing data. The experimental bearing is SKF's 6205 deep groove ball bearing. Single point fault is arranged on the bearing by EDM technology, the spindle speed is 1797 r/min and the sampling frequency is 12KHz, the fault size is 0.007-inches and 0.014-inches respectively, it is divided into inner ring, outer ring, rolling body fault bearing and normal bearing. There are 7 categories of data. 150 samples of each category are taken as training data and 60 samples are taken as test data. Each category is normal, inner ring fault, outer ring fault, rolling body fault: Category 1, Category 2, Category 3, Category 4, Category 5, Category 6, Category 7.
4.2 Experimental result

According to the cumulative contribution rate calculated in the previous section, the number of principal elements with cumulative contribution rate of 85% is six. The six principal components are selected as the real components, and they are input into the PNN and BP neural network as eigenvector to train and diagnose the network. The simulation results of the rolling bearing fault data of PNN and BP neural network are shown in Figure 3, Figure 4, Figure 5, and Figure 6.

By comparing the above simulation results of PNN network and BP neural network, it can be seen that probabilistic neural network has advantages over BP neural network in fault diagnosis of rolling bearing and has higher accuracy. The fault recognition rate of PNN is 100%, and that of BP network is 98.33%. Under the same conditions, the fault diagnosis accuracy of PNN network is higher. Therefore, the rolling bearing fault diagnosis algorithm based on PCA and PNN proposed in this paper is superior to other algorithms, and can effectively and accurately identify and diagnose rolling bearing faults.

5. Conclusion

This paper presents a method for fault diagnosis of rolling bearings based on PCA and PNN. The principal component obtained by PCA is the input sample of PNN fault diagnosis model. By training and testing the PNN network, the recognition rate of fault diagnosis is counted.
The proposed method for fault diagnosis has been further improved, while reducing the dimension of eigenvectors, shortening the diagnosis time, and improving the accuracy, which verifies the effectiveness of the method.

References
[1] Gu DS, Kim JG, An YS, et al. (2011) Detection of faults in gearboxes using acoustic emission signal. Journal of Mechanical Science and Technology, 25(5): 1279-1286.
[2] ZHANG X.W, LÜ Y.Y, DING Y.Q, et al. (2010) Surface defect inspection based on wavelet statistical analysis. Control Theory & Applications, 27(10): 1313–1336.
[3] CHENG P. L. (2017) The Fault Diagnosis Research of Refrigeration System of High-low Temperature Shock chamber Based on The PCA and PNN, Guangdong University of Technology.
[4] JIANG L.Y, LI W.L, CUI J.G, TU M.Y LIN Z.L. (2019) Fault diagnosis of aero-engine gas path system based on PCA and DBN. Journal of Shenyang Aerospace University, 36(01): 57-62.
[5] LI W.Y, YANG F, ZHOU Y.X, SHEN J.C, YU Y.L, ZHAO L.J, DING Y.N, LI W, et al. (2019) Research on fault diagnosis of steam turbine flow passage based on improved EMD and PNN. Journal of Vibration, Measurements & Diagnosis, 38(06): 1138-1142+1289-1290.
[6] XU K, CHEN Z.H, ZHANG C.B, DONG G.Z. (2019) Rolling bearing fault diagnosis based on empirical mode decomposition and support vector machine. Control Theory & Applications, 44: 1-8.
[7] ZHENG J.D, LIU T, MENG R, LIU Qi.Y. (2018) Generalized composite multiscale permutation entropy and PCA based fault diagnosis of rolling bearings. Journal of Vibration and Shock, 37(20): 61-66.
[8] QU H. (2017) Research on Fault Diagnosis of Ventilator Bearing Based on Improved LMD and PNN. China University of Mining and Technology.