Application of BP Neural Network and Convolutional Neural Network (CNN) in Bearing Fault Diagnosis

Hai Deng¹, Wan-xuan Zhang² and Zheng-feng Liang¹

¹Junior engineer, Xi’an Modern Chemistry Research Institute, Xi’an, China
²Junior engineer, Xi’an Aerospace Power Research Institute, Xi’an, China

E-mail: hai179590@163.com

Abstract: The convolutional neural network (CNN) identification method and the BP neural network identification method were used to diagnose the bearing fault respectively. When using the CNN diagnostic method, first perform continuous wavelet transform (CWT) on the vibration signal of the rolling bearing to obtain a time-frequency map, and then compress the time-frequency map to an appropriate size; Then, the compressed time-frequency map is used as a feature map to input into the CNN classifier model established; finally, an experimental study is carried out based on the artificial bearing fault data set of Western Reserve University. The results show that the average accuracy rate of this method is greater than 99%. In the BP fault diagnosis method, based on the data set, nine parameters of average value, maximum value, minimum value, peak-to-peak value, root mean square value, standard deviation, variance, skewness and kurtosis are established as training input vectors. Using BP neural network with 10 hidden layer nodes for fault identification, the results show that the average accuracy of identification is 93.7839%. The comparative analysis of the two methods shows that the BP identification method has higher training efficiency and takes less time; the CNN identification method has higher recognition accuracy, but the training takes more time.

Keywords: fault diagnosis, ball bearings, wavelet packet decomposition, convolutional neural network, BP neural network

1. Introduction

As industrial machinery becomes larger, faster and more complicated. The most widely used rolling bearings often fail due to wear, corrosion, overload operation, lack of lubrication, overload and other problems, which will cause abnormal vibration and noise of the equipment, and affect the accuracy and reliability of the entire unit. According to statistics, 30% of mechanical failures are caused by rolling bearing failures. Therefore, it is very important to monitor and diagnose the rolling bearings of mechanical equipment.

The traditional rolling bearing fault diagnosis mainly relies on operators to test and analyze based on experience. However, due to the high complexity of bearing design, this has brought difficulties to the accurate implementation of rolling bearing fault diagnosis. With the development of modern computing detection technology, The intelligent detection technology based on wavelet packet feature extraction convolutional neural network(CNN) and BP neural network method has been widely used. Anjali Jawade [1] used the minimum value of CWT coefficient and back propagation BP neural network to study the intelligent diagnosis of induction motor faults; Chen Z Q [2] used the characteristic diagram of gearbox vibration signals as the input of CNN to diagnose and identify...
gearbox faults, achieving a higher recognition rate and efficiency. Song Chong-zhi [3] established a gearbox fault diagnosis model based on improved Elman neural network. This method has the advantages of fast convergence and avoidance of local minimum. Zhang Wen-jing [4] used time-frequency analysis technology to filter the vibration signal of the wind turbine fault, and used the energy change as the fault characteristic value of the wind turbine gearbox, and applied BP neural network method to carry out intelligent fault diagnosis.

In this paper, the convolutional neural network (CNN) identification method based on wavelet time-frequency graph and the BP neural network identification method based on fault feature are used to diagnose and identify bearing faults. The characteristics of the two methods are compared and analyzed, high-precision identification of ball bearing fault diagnosis is achieved.

2 Theoretical model

2.1 Wavelet packet decomposition model

The wavelet packet transform is a family of functions, which can divide the low-frequency and high-frequency parts of the signal into multiple layers to achieve leak-free decomposition and reconstruction. It can obtain the characteristic information of the signal at any frequency segment, improve the frequency resolution of the detail signal and the approximation signal, and facilitate the extraction of the rolling bearing fault characteristic vector[5]. The schematic diagram is shown in Figure 1. In Fig. 1, A represents a low frequency feature (Approximate), and D represents a high frequency detail (Detail).

![Wavelet packet decomposition structure diagram](image)

Fig.1 Wavelet packet decomposition structure diagram

2.1.1 Continuous wavelet transform

The definition of continuous wavelet transform is: for any \( L^2(R) \), the function \( f(t) \), the CWT is:

\[
X(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt
\]

Where \( \psi\left(\frac{t-b}{a}\right) \) is the wavelet basis function, the expression is as:

\[
\psi\left(\frac{t-b}{a}\right) = g\left(\frac{\tau-b}{a}\right)e^{i2\pi f_{\tau}}
\]

Where \( a \) is the scale parameter, \( b \) is the time shift parameter, and \( 1/\sqrt{a} \) is the coefficient introduced in order to normalize the transformation result. When the value of \( a \) is large, it is helpful to extract low-frequency features in the signal, and when the value of \( a \) is small, it is beneficial to extract high-frequency features in the signal.

2.1.2 Wavelet packet time-frequency phase diagram
Wavelet packet decomposition implements time-frequency decomposition through orthogonal mirror filters. Assuming the signal is $y(t)$, there is the following formula:

$$y_{2n}(t) = \sqrt{2} \sum h(k) y_n(2t - k) \quad (3)$$

$$y_{2n+1}(t) = \sqrt{2} \sum g(k) y_n(2t - k) \quad (4)$$

The function $\{y_n(t)\}$ is called an orthogonal wavelet packet, which is the result of all decompositions of all frequency bands of the original signal at various scales. Let $k = n - 2^j$, then $y_n(t) = y_{2+k}^j(t)$ be the decomposition result of the signal for the scale $j$ in the frequency band $k$.

The wavelet packet can constitute the decomposition results of many different orthogonal bases, and the combination with the smallest entropy is selected as the best base. After the signal is decomposed by the wavelet packet and the optimal basis is selected, the decomposition result on the optimal basis is expressed on the time-frequency plane, that is, the time-frequency phase plan of the wavelet packet [6].

### 2.2 BP neural network model

BP (Back Propagation) neural network is a multi-layer feed-forward neural network, which is composed of input layer, hidden layer, and output layer. Neurons in the same layer are not connected to each other, and different neuron layers are connected to each other. The BP algorithm consists of two stages: parameter forward propagation and error backward propagation. The data signal is input from the input layer, processed by the hidden layer, and then output to the output layer. The actual output and the expected output are subtracted to obtain the output error of the network. The error signal propagates in the reverse direction of the network and continuously corrects the connection weights and thresholds of each layer according to the negative gradient direction of the error function, so that the error signal finally meets the accuracy requirements, thereby achieving the Non-linear mapping of input and output [7]. The BP neural network model is shown in Figure 2.

![BP neural network model](image)

**Fig 2.** Topological structure of BP neural network

(1) Input layer: the output of each node $i$ is equal to $x_i$ ($i=1, 2, \ldots, n$), the input layer actually only plays a role of transmission, but only passes the value of the input variable to the hidden layer.

(2) Hidden layer: hidden layer node $j$

Input $h_j$ is:

$$h_j = \sum_{i=1}^{n} x_i \quad (5)$$

Output $O_j$ is:

$$O_j = f(h_j) = \frac{1}{1 + e^{-h_j}} \quad (6)$$

where $j = 1, 2 \ldots m$; $m$ is the number of hidden layer nodes, $f$ generally uses Sigmoid function.

(3) Output layer: output layer node $K$

Input $h_k$ is:

$$h_k = \sum_{j=1}^{m} w_{jk}x_j - \theta_k = \sum_{j=0}^{m} w_{jk}x_j \quad (7)$$
Output $y_k$ is:

$$y_k = f(h_k) = \frac{1}{1+e^{-h_k}}$$

(8)

(4) Forward propagation process: suppose the input vector of the kth sample is $X_k = (x_{k1}, x_{k2}, ..., x_{kj})$.

The output $y_j$ of the jth node of the hidden layer is:

$$y_j = f\left(\sum_{i=1}^{L} W_{ji}x_i + \theta_j\right)$$

(9)

where L is the number of neurons in the input layer, $W_{ji}$ is the weight of the input layer to the hidden layer, $x_{ij}$ is the input of the kth sample, $\theta_j$ is the threshold of the jth neuron in the input layer.

The output $O_r$ of the rth node of the output layer is:

$$O_r = f\left(\sum_{j=1}^{M} W_{rj}y_j + \theta_r\right)$$

(10)

where M is the number of neurons in the hidden layer, $W_{rj}$ is the weight of the connection between the hidden layer and the output layer, $\theta_r$ is the threshold of the jth neuron in the hidden layer.

2.3 Convolutional neural (CNN) model

Convolutional neural network (CNN) is a multi-layer perceptron (MLP) designed to recognize two-dimensional feature maps. It is a deep learning network model with multiple hidden layers. It can transfer low-level features to high-level features through layer-by-layer feature transfer to achieve feature learning and expression. Compared with shallow networks such as BP neural network and SVM, CNN has a stronger ability to learn and express complex features and faster calculation speed. It is a network model with excellent recognition performance. It has been initially applied in fault diagnosis, which improves the level and efficiency of fault diagnosis.

The CNN is composed of an input layer, a hidden layer, a fully connected layer, and an output layer. Among them, the hidden layer is composed of several convolutional layers and sampling layers alternately, the fully connected layer and the output layer constitute a classifier, this classifier can be logistic regression, softmax regression and SVM; The convolution layer is the convolution of a specific convolution kernel and the feature map of the input layer, plus an offset, and the output feature is obtained through an activation function; The sampling layer performs feature filtering on the output feature map of the convolution layer. The schematic diagram of CNN structure is shown in Figure 3.

Fig 3. Schematic diagram of convolutional neural network

The training method of CNN adopts the method of batch sample input, which includes two parts of forward propagation of data and backward propagation of error. First, set the training parameters of the network, initialize the weights and offsets of the network, input the feature map, process it through the convolutional layer, sampling layer, and fully connected layer, and then transmit it to the output layer. The output of each layer is the input of the next layer, Then, the error between the actual output and the expected output is back-propagated layer by layer through the BP algorithm, and the error is distributed to each layer to adjust the weights and offsets of the network. Until the convergence conditions are met, to achieve supervised training of the network[8].

In this paper, a rolling bearing intelligent fault diagnosis method based on wavelet time-
frequency map and CNN is used. First, the vibration signal of the rolling bearing is generated by CWT to generate a time-frequency map, and then it is input as a feature map to CNN to realize the diagnosis and recognition of rolling bearing faults.

3 Example simulation verification and analysis

3.1 Data source
The vibration data comes from the open data set of Case Western Reserve University (CWRU). In this experiment, the bearing at the motor drive end was used as a diagnostic object, and single-point damage was introduced on the inner ring, outer ring and rolling body of the test bearing to simulate three kinds of bearing faults. The damage sizes were 0.007 inch, 0.014 inch and 0.021 inch. Then under four different working conditions (different load and speed), the acceleration sensor on the upper side of the motor drive side collects the signal, and the sampling frequency is 12kHz.

Divide the collected vibration signal, each sample contains 1024 sampling points. There are 238 samples in normal state under each working condition. Under different fault types, there are 117 samples with different damage sizes and different working conditions. Finally, 1400 samples in normal state and three fault states are obtained respectively. Table 1 is the signal samples of different states of the bearing.

Table 1. List of signal samples in different states

| Damage size | load | Motor speed | Normal signal | Inner ring failure | Ball failure | Outer ring failure |
|-------------|------|-------------|---------------|-------------------|--------------|-------------------|
| 0           | 0    | 1797        | 238           | \                | \            | \                 |
|             | 1    | 1772        | 238           | \                | \            | \                 |
|             | 2    | 1750        | 238           | \                | \            | \                 |
|             | 3    | 1730        | 238           | \                | \            | \                 |
| 0.007       | 0    | 1797        | \             | 117              | 117          | 117               |
|             | 1    | 1772        | \             | 117              | 117          | 117               |
|             | 2    | 1750        | \             | 117              | 117          | 117               |
|             | 3    | 1730        | \             | 117              | 117          | 117               |
| 0.014       | 0    | 1797        | \             | 117              | 117          | 117               |
|             | 1    | 1772        | \             | 117              | 117          | 117               |
|             | 2    | 1750        | \             | 117              | 117          | 117               |
|             | 3    | 1730        | \             | 117              | 117          | 117               |
| 0.021       | 0    | 1797        | \             | 117              | 117          | 117               |
|             | 1    | 1772        | \             | 117              | 117          | 117               |
|             | 2    | 1750        | \             | 117              | 117          | 117               |
|             | 3    | 1730        | \             | 117              | 117          | 117               |
| total       | \    | \           | 952           | 1404             | 1404         | 1404              |

3.2 CNN simulation

3.2.1 Wavelet packet data decomposition
The “morlet” wavelet and “cmor3-3” wavelet base with a bandwidth parameter and a center frequency of 3 are selected to perform CWT on the bearing sample signal to generate a time-frequency graph. Then, use the “imresize” function of MATLAB to compress the time-frequency map, and set the size of the compressed time-frequency map to 28×28. The result of the time-frequency graph before graying is shown in Fig 4, and the result of the time-frequency graph after graying is shown in Fig 5.
It can be found from the above that the data quality is very high, and the fault characteristics are clearly reflected on the time-frequency graph. Finally, 1404 samples of feature maps in different states of the bearing were obtained, and 75% (1053) of the samples were selected as training samples, and the remaining 351 samples were used as test samples.

Table 2. Data set

| Signal type            | Training samples | Test sample | Sample label          |
|------------------------|------------------|-------------|-----------------------|
| Normal signal          | 714              | 238         | [1 0 0 0]T            |
| Inner ring failure     | 1053             | 351         | [0 1 0 0]T            |
| Ball failure           | 1053             | 351         | [0 0 1 0]T            |
| Outer ring failure     | 1053             | 351         | [0 0 0 1]T            |
| Total                  | 3873             | 1291        | \                     |

3.2.2 Experimental results

After the network structure is determined, the training samples are used to train the network, and the training parameters are set as follows: batch=3, epoch=1.

The structure of the convolutional neural network used is as follows:

Table 3. Convolutional Neural Network Structure

| name        | $N_{c1}$ | $C_{1x}$ | $N_{c2}$ | $C_{1x}$ |
|-------------|----------|----------|----------|----------|
| Numerical value | 6        | $5 \times 5$ | 12       | $5 \times 5$ |

The configuration of the experimental platform is as follows: the 64-bit operating system of Windows 10, the CPU is i7-7700HQ, the GPU is GTX1050 (2G), the memory is 8G, and the program running environment is MATLAB2018a. After 10 steps of training, the change curve of its correct rate is shown in Figure 6. It can be seen that after 6 training iterations, the error value tends to be a small value and remains stable, indicating that the network has trained to convergence, and the final correct recognition rate is 99.9225%.

After testing 1291 data, only one sample was finally checked for error, and the ball damage was misjudged as outer ring damage. The average time used for one iteration is 13s.
3.3 BP simulation

3.3.1 Data preprocessing and feature selection

The acceleration signal measured in real time is a random signal, and a variety of information is mixed, which cannot be directly used for BP network pattern recognition, and the signal needs to be processed. Therefore, the feature analysis in this paper adopts the time domain feature analysis method, and the averaging and normalization processes are performed before using the BP neural network for identification. The principle of the value normalization method is as follows:

\[
    x = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

(11)

Where \(x\) is the normalized data; \(x_i\) is the data before normalization; \(x_{\text{min}}, x_{\text{max}}\) is the minimum and maximum value of the data before normalization.

Then use MATLAB to calculate the time-domain characteristic parameters of the signal data, including peak value, peak-to-peak value, mean value, absolute mean value, mean square value, root mean square value, root mean square amplitude, variance, skewness, kurtosis, waveform index, Peak index, pulse index, margin index, skewness, steepness, etc.

In the end, this paper selects nine parameters of good stability, large difference between different modes, relatively sensitive to faults as the input of BP network, including average value, maximum value, minimum value, peak-to-peak value, root mean square value, standard deviation, variance, skewness, and kurtosis.

3.3.2 BP verification results

In this paper, the BP neural network is used for fault diagnosis. The structure of the neural network is unified as shown in Figure 7. The input layer is 9 nodes, the hidden layer is 10 nodes, and the output layer is 4 nodes.

The excitation function of the hidden layer uses the Sigmod function. The network structure is shown in Figure 7.
In order to compare with the previous CNN effect, 1024 sampling points are also taken as one sample, and iterative training takes 183 steps. The highest accurate recognition rate is 94.9%. The result is shown in Figure 8.

![BP neural network structure](image)

**Fig 7.** BP neural network structure

In order to compare with the previous CNN effect, 1024 sampling points are also taken as one sample, and iterative training takes 183 steps. The highest accurate recognition rate is 94.9%. The result is shown in Figure 8.

![Graphs showing error rate vs iteration steps](image)
a. Overall recognition error rate  
b. Test data recognition error rate

**Fig 8.** Test recognition results

### 3.4 Comparison of identification effects of CNN and BP neural networks

In order to compare the recognition performance of CNN and BP in this fault diagnosis. In this paper, 75% of the data set is used as the training data set, and 25% is used as the test data set. After training 20 times, each time the data set is randomly divided again to compare the identification effect of CNN and BP.

The results are shown in the following table. It can be found that the recognition accuracy of the CNN neural network is higher than that of the BP neural network. The average recognition accuracy of the CNN neural network is 99.8258%, and the average recognition accuracy of the BP neural network is 93.7839%.

| Num | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CNN | 99.9% | 100% | 99.5% | 99.7% | 99.9% | 99.8% | 99.9% | 99.8% | 99.7% | 99.9% |
| BP  | 93.6% | 93.7% | 93.3% | 93.8% | 93.1% | 94% | 93.9% | 94.9% | 94.1% | 93.9% |

| Num | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CNN | 99.9% | 100% | 99.4% | 99.7% | 99.9% | 99.8% | 99.9% | 99.8% | 99.8% | 99.9% |
| BP  | 93.7% | 94% | 94.5% | 94% | 93.6% | 94.1% | 92.8% | 93.8% | 92.6% | 94.2% |

| Mean | Std  | Min  | Max  | Median |
|------|------|------|------|--------|
| CNN  | 99.8528 % | 0.14632 | 99.3803 % | 100 %  | 99.8838 % |
| BP   | 93.7839 % | 0.53717 | 92.5639 % | 94.8877 % | 93.8420 % |
4 Conclusion
(1) The time-frequency map obtained by wavelet packet decomposition can effectively reflect the time-frequency characteristics of the vibration signal, and can accurately characterize the running state of the rolling bearing;
(2) Both CNN network and BP neural network can effectively diagnose and identify the running faults of rolling bearings, but CNN network has higher recognition accuracy than BP network, CNN network has stronger generalization ability and clustering ability, feature extraction and recognition ability is stronger than BP network;
(3) The BP method has high training efficiency, short time, and high CNN recognition accuracy, but the training takes more time.

Reference
[1] Jawade A, Raska R S and Jadhav S 2014 Artificial neural network based induction motor fault classifier using continuous wavelet transform. Systems Science & Control Engineering. 2(1) 684-90
[2] Chen Z Q, Li C and Sanchez R V 2015 Gearbox fault identification and classification with convolutional neural networks. Shock and Vibration 15(8) 18-28
[3] Song C Z, Wu Y G and Wang L 2008 Research on improved Elman network in engine gearbox fault diagnosis. Journal of Instrumentation 29(7) 1414-7
[4] Zhang W J, Zhang L and Liu M 2017 Fault diagnosis of the wheel box of wind turbines based on wavelet and BP neural network. Control Theory and Applications. 36(6) 11-5
[5] Wen X S and et al 1997 Pattern recognition and status monitoring. (Changsha: National University of Defense Technology Press)
[6] Zhang W and et al 2017 A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mechanical Systems & Signal Processing. 100(9) 439-53
[7] Wang G F and et al 2004 Fault diagnosis of bearings based on wavelet packet and radial basis neural network. Journal of Engineering Sciences. 26(2) 184-7
[8] Yuan J H and et al 2017 Intelligent fault diagnosis method of rolling bearing based on wavelet time-frequency graph and CNN. Mechanical Design and Research. 2 93-7