Acta Technica 62 (2017), No. 5B, 665–10

Rolling bearing fault diagnosis based on GA-BP and H-ELM

Li Rongyuan[1], Zhang Guoyin[1], Wang Hairui[1], Wang Xue[1], Song Yiran[1], Qi Lei[1]

Abstract. The methods of deep learning are widely used in the fault diagnosis of rolling bearings. Rolling bearings play a key role in mechanical failure components. At present, many fault diagnosis methods, such as Back-Propagation (BP) neural network, genetic algorithm (GA) has been proven to have a good diagnostic effect, but there are still defects in processing large amounts of data: (1) the training process is too long. (2) It is easy to trap into local minima value. (3) It requires feature extraction, no general feature extraction method. Extreme learning machine (ELM) has a high training speed, but low classification ability, and proposes a layered learning machine (h-elm) framework based on multi-layer perceptron. Aimed on these a hierarchical extreme learning machine (H-ELM) framework based on multi-layer perceptron is proposed.The first part of the unsupervised multi-layer encoding part and the second part is the supervised feature classification part. This method is compared with the traditional diagnosis method. The simulation results show that the method has good generalization performance and can improve the accuracy and speed of rolling bearing fault diagnosis.

Key words. GA-BP neural network, H-ELM. Rolling bearing, Fault diagnosis.

1. Introduction

Rolling bearings in mechanical failure are one of the key components to ensure the safe operation of high-speed rail trains. It is usually composed of four parts: inner race, outer race, ball and cage [1-2].Therefore, the fault diagnosis of rolling bearing is critical, and its operational safety directly affects the overall performance of the track train[3-4].Among the many intelligent fault diagnosis, feature extraction is a key step in fault diagnosis. Wavelet packets decompositions (WPD) and empirical mode decomposition (EMD) [5] are widely used in feature extraction. EMD-LM-BP presented in literature[6], firstly, the wavelet energy extraction is carried out in the time-frequency domain of the bearing. Then, based on the LM-BP fault diagnosis, although the network convergence rate is improved, the number of iterations is re-

[1]Workshop 1 - School of Information Engineering and Automation, Kunming University of Science and Technology, Yunan, 650504, China

duced, but the number of experimental samples is small and only applicable to small sample data. The weight and threshold of BP algorithm based on GA are proposed to improve the efficiency and accuracy of bearing fault diagnosis[7]. The sample data is less, high efficiency, but not representative. The literature[8] unsupervised automatic encoder network proposed to lay the beginning of its development, in the depth of the network, DL framework of all hidden parameters needs to be adjusted several times. Therefore, the training of DL for the bearing fault diagnosis and time-consuming. Therefore, DL training is difficult and time-consuming for bearing fault diagnosis.

The extreme learning machine (ELM) is a new algorithm of single-hidden layer feedforward neural networks (SLFNs). Compared with the shortcomings of the traditional feedforward neural network. Singapore Nanyang Technological University Huang Guangbin in 2006 proposed ELM[9], and prove the learning speed and generalization ability of this algorithm. There is no need to adjust the random input weights and hidden layer bias, the application of the algorithm has been widely proved that the method not only learns fast, but also has a good generalization ability[10]. However, ELM in some practical applications also face some problems, such as voice recognition, image classification. Prior to the classification, feature learning is usually required. Therefore, the need for multi-layer solution. Huang Guangbin et al. [11,14] have demonstrated that ELM's general approximation capability cannot guarantee that no input of random projection. ELM multilayer is not fully exploited. ELM-based multi-layer perceptron (MLP) research more and more, proposed hierarchical overrun learning architecture (H-ELM), to further improve the original ELM learning ability. H-ELM consists of two parts: the first part is unsupervised multilayer coding; the second part is the supervisory feature classification. This method is compared with the traditional diagnostic method. The simulation results show that the method has good generalization performance and can improve the accuracy and speed of rolling bearing fault diagnosis.

2. Fault Feature Extraction of Bearing

In this paper, the design of the bearing fault diagnosis process shown in Figure 1, the first is the original signal acquisition, preferred to use in different directions of the rolling bearing acceleration sensor to collect vibration signals, amplified by the amplifier, through the A / D converter to convert the computer to identify the signal; Followed by the collection of information processing analysis; the final diagnosis of rolling bearing failure type.

The steps of extracting the feature information by wavelet transform are as follows:

a.The three-layer wavelet decomposition of the sampled vibration signal is carried out, and the low frequency to high frequency characteristic signals of each layer is extracted respectively.

b.Reconstruct the wavelet packet decomposition coefficient, extract the frequency band signal.

c.Find the signal energy of each band.

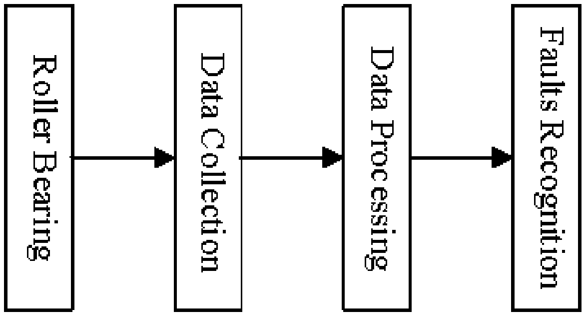

Fig. 1. Rolling Bearing Fault Diagnosis Process Diagram

d.Construct the eigenvector. T=[E30,E31,E32,E33,E34,E35,E36,E37], When the energy is large, the feature vector T can be normalized, Normalized vector T'=[E30/E ,E31/E,E32/E,E33/E,E34/E,E35/E,E36/E,E37/E].

3. Learning Algorithm

3.1. Back-Propagation Neural Network

BP neural network is a multilayer feedforward neural network, and its main training algorithm is error back propagation algorithm. It is composed of three parts, input layer, hidden layer and output layer[12]. BP neural network has good self - learning and adaptive ability, greater tolerance, easy to build. But the defects are as follows:

a. training speed is slow.The calculation cost is big;

b. easy to fall into a local minimum;

c. learning rate is difficult to control, prone to over-training, generalization performance is low.

Late use of LM to optimize BP, it only improved the speed, and the accuracy needed to be improved.

3.2. GA Optimization BP Neural Network Model

GA algorithm is used to optimize the initial weights and bias of BP network. It mainly includes population initialization, definite fitness function, selection, crossover and mutation operator. GA-BP algorithm flow charts Figure 2.

3.3. Extreme Learning Machine

The model of extreme learning machine adopts traditional single hidden layer feedforward neural network structure, including input layer, hidden layer and output layer. Each neuron corresponds to each variable, where the input layer has n input variables; the hidden layer has l variables, the output layer has m output variables, G () is the hidden layer neuron activation function, b is the hidden layer Threshold

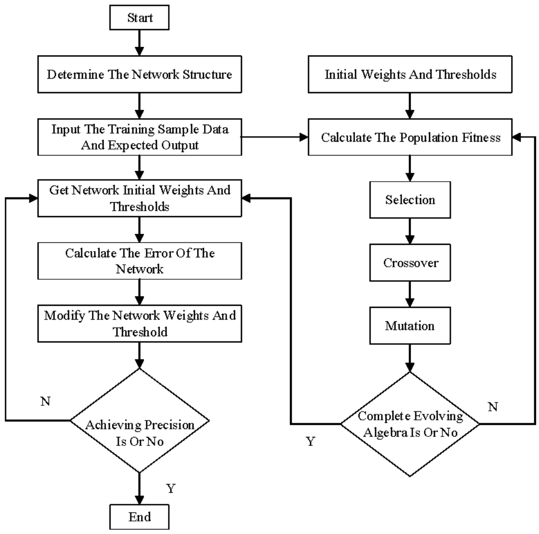

Fig. 2. GA-BP algorithm flow chart

of neurons. Xi input training samples, Ti output characteristics. The network structure is shown in Figure 3.

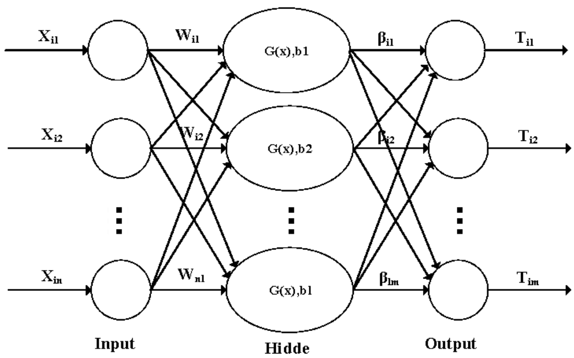

Fig. 3. Extreme learning machine network model

The extreme learning machine training model can be expressed by the following formula:

$$\sum_{1}^{L} \mathrm{g}(x_i \bullet w_i + b_i) \bullet \beta_i = t_j, j = 1, 2, ..., Q \tag{1}$$

Where $w_i = [w_{1i}, w_{2i}, ..., w_{li}]$ is the input weight; $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$is the output weight; $t_i = [t_{i1}, t_{i2}, ..., t_{im}]^T$is the network output. Q is the set of Q training samples.

The formula (1) can be expressed as

$$H\beta = T' \tag{2}$$

T'is the transpose of the network output vector T; H is the hidden layer output matrix. Huang guangbin and others suggested that any given Q of different samples (xi, ti),where $x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in R^n$,$t_i = [t_{i1}, t_{i2}, ..., t_{im}] \in R^m$, the activation function g (x) infinitely differentiable, w, b random assignment can be trained, any minimum errorε>0, and tends to 0, the formula is as follows

$$\varepsilon = \sum_{j=1}^{Q} ||t_j - y_j|| \tag{3}$$

Where $y_j = [y_{1j}, y_{2j}, ..., y_{mj}]^T (j = 1, 2, ..., Q)$. The hidden layer to the output layer connection weights β can be solved by the least squares method

$$\hat{\beta} = H^+T' = H^+(HH^T)^{-1}T' = H^T(\frac{I}{C} + HH^T)T' \tag{4}$$

Where H+ is the Moore-Penrose pseudo-inverse of matrix,where T is the network forecast output, according to ridge regression theory, the addition of the regular coefficient C, I is the identity matrix, and has a good generalization ability[13].

3.4. Hierarchical Extreme Learning Machine (H-ELM)

Hierarchical Extreme Learning Machine (H-ELM) [14] was presented by Guang-Bin Huang et al. It's a multi-layer perceptron (MLP) algorithm for training data. The data before learning should be converted to ELM random feature space, as shown in Figure 4 based on ELM layered learning framework, the framework is mainly two parts: Unsupervised feature extraction is shown in Figure 4 (a), The supervisory characteristics are classified as shown in Figure 4 (b)

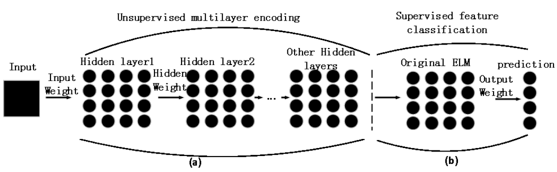

Fig. 4. Overall framework of H-ELM

a. The first part of the unsupervised learning feature stratification, the more layers, learn the characteristics of more sparse, each layer of the output formula is as follows:

$$H_i = g(H_{i-1} \cdot \beta) \tag{5}$$

Where Hi is the i-th layer random mapping output Hi-1 is the i-1h layer random mapping output , g() is an activation function, $\beta$is the hidden layer weight. In addition, each layer is extracted using ELM sparse automatic encoders, and the objective function is used as the input sample. In addition, Each layer uses ELM sparse autoencoder to extract input sample features , and the objective function is defined as follows:

$$O_\beta = argmin_\beta\{\parallel H\beta - X \parallel^2 + \parallel \beta \parallel_{\ell 1}\} \qquad (6)$$

Where X is the input data, H is the random mapping output matrix, $\beta$is the optimal solution of the hidden layer weight, In order to obtain β, the literature [15] through the introduction of $\ell 1$ constraints, so that the characteristics of learning more sparse.

b. The second part supervises the classification of learning characteristics, and the last layer of the network is based on ELM regression training to determine the final outcome of the network.

4. Analysis of Fault Diagnosis Experiment Results

4.1. Establish an experimental system

In order to verify the wpd-bp, WPD - ga-bp and the ml-elm, the h-elm method of the original signal. Experimental use 6203-2RS JEM SKF deep groove ball bearings, the use of acceleration sensors to collect drive and fan side of the data (experimental data from the US West University Laboratory). Sampling frequency 12kHz, speed 1797r/min, damage diameter 0.1778mm expressed as mild damage, damage diameter 0.5334 mm expressed as moderate damage, each state sampling point 100000, divided into 100 samples, each sample 1000 data points. Lightly damaged bearing outer ring point is three o'clock.Moderate damage bearing outer ring point is 6 o'clock directions. Which bearing the seven states (normal, mild inner race fault,mild ball fault,mild outer race fault,moderate inner race fault,moderate ball fault,moderate outer race fault) mild damage selection Drive side, moderate damage select fan end. (1, 2, 3, 4, 5, 6, 7), the matrix form is (1000000, 0100000, 0010000, 0001000, 0000100, 0000010, 0000001). A total of 700 samples, each selected 75% of each state as a training, training a total of 525 samples, the remaining 175 as a test sample.

The test operating system is Windows7, 64 bit, Intel (R) Core (TM) i3-3217U cpu @ 1.80GHz processor, memory (RAM) is 8G; MATLAB 2014a.

4.2. Feature Extraction

WPD decomposition of four kinds of vibration signal process: wavelet decomposition, signal reconstruction, structural feature vector.

Some training data of wavelet energy extraction are shown in table 1. The hidden layer is calculated from the formula a, with an approximate value of 5 to 14. Each time the assignment is repeated five times, the average of 12 is the optimal hidden layer node, as shown in Figure 5. The structure of the BP neural network is 8-12-4. The error of the test sample is used as the judgment ability of the network, and the fitness value is calculated by the error. The larger the fitness value, the better the individual. The number of BP training is 1000, the mean square error is 0.01, the learning rate is 0.01. Hidden layer transfer function using logsig, the output layer is purelin function. Where the training function of the L-M algorithm selects trainlm.

Table 1. Part training data

Sample	E_{30}/E	E_{31}/E	E_{32}/E	E_{33}/E	E_{34}/E	E_{35}/E	E_{36}/E	E_{37}/E	**Fault Type**
1	0.366318	0.589561	0.002324	0.040142	0.000020	0.000656	0.000688	0.000291	1
2	0.048045	0.098973	0.317956	0.063457	0.001147	0.006049	0.418271	0.046102	2
3	0.056012	0.027563	0.178941	0.011839	0.001908	0.002060	0.710544	0.011133	3
4	0.003034	0.004745	0.831164	0.010805	0.000226	0.000172	0.148877	0.000977	4
5	0.070555	0.145344	0.466924	0.093188	0.001685	0.008883	0.614238	0.067701	5
6	0.138769	0.068286	0.443322	0.029331	0.004727	0.005104	1.760355	0.027581	6
7	0.041325	0.064637	11.322189	0.147193	0.003081	0.002341	2.028013	0.013305	7

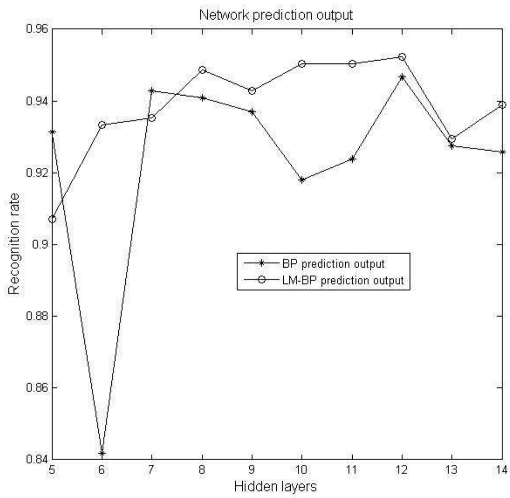

Fig. 5. The average recognition rate of each hidden layer

4.3. Network Training and Testing

The remaining 175 samples were used as test samples. According to the different hidden layer training effect is not the same, which gradient descent BP neural network training requires iteration to reach the preset condition 1000 times to stop, and L-M algorithm on average only 10 iterations to achieve the effect. It can be seen that the efficiency of the algorithm based on L?M is much higher than that of the gradient descent BP neural network, although the recognition rate averages 94.29%, but the diagnosis rate is not high. Finally, optimize the network with GA. Evolutionary algebra 20, training target: 0.01, population size 10, the crossover probability 0.5, mutation probability 0.01.

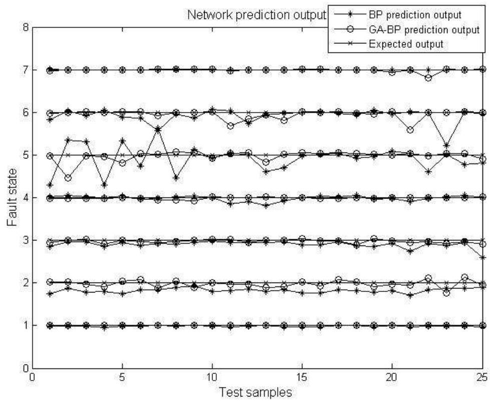

Fig. 6. Comparison of GA-BP and BP prediction

Comparison of the two identification errors As show in Figure 6, 175 test samples are divided into seven groups, that is, seven state features (1 to 7 digital representations) with the vertical axis to represent. Among the seven features, each state feature is taken for 25 samples, as the horizontal axis. It can be seen that the GA-BP error is significantly less than BP, more approaching the expected value. But GA-BP training time is longer. The following is the H-ELM simulation of the original signal. 525 primitive signal training samples, 175 test samples, test input neurons of 1000, hidden layer N1 = 50, N2 = 50, N3 = 200, the output neuron was 7, each method repeated 5 experiments.

Table 2. Comparison of experimental results

Method	**Accuracy(%)**	**Trainning time(s)**
BP	90.28	7.1604
LM-BP	94.29	4.9296
GA-BP	97.01	1361.91993
H-ELM	98.85	0.5518

Table 3 Accuracy of each fault state diagnosis

Fault state	**BP accuracy(%)**	**LM-BP accuracy(%)**	**GA-BP accuracy(%)**	**H-ELM accuracy(%)**
Normal	100	100	100	100
Mild Inner Race Fault	76	88	96	100
Mild Ball Fault	88	88	96	96
Mild Outer Race Fault	96	96	100	100
Moderate Inner Race Fault	88	88	96	96
Moderate Ball Fault	84	96	96	100
Moderate Outer Race Fault	100	100	100	100

From table 2, table 3 shows the result of the experiment, although the GA has a good accuracy, but the population initialization takes longer to find the optimal weights threshold, H - ELM has not only high accuracy, short operation time, but also has the very high stability.

5. Conclusion

Aiming at the limitation of traditional fault diagnosis method, this paper proposes a multi-level learning machine fault diagnosis method. The method is verified by using the bearing 7 states, and the simulation results prove that the method can obtain a good fault diagnosis without the feature extraction. Compared with the traditional fault diagnosis method, H-ELM avoids the feature extraction process and can get the classification effect very well. H-ELM has great improvement in accuracy, stability and efficiency, and has good application value. The future will further study the expansion of H-ELM.

References

[1] J. S. Tomar, D. C. Gupta, N. C. Jain: *Adaptive detrended fluctuation analysis as a feature extraction method for gear's vibration signal.* Journal of Vibration and Shock *15* (1984), No. 2, 211–220.

[2] J. S. Tomar, A. K. Gupta: *Fault Analysis of Engine Crankshaft Rolling Bearing.* Farm Machinery Using & Maintenance *98* (1985), No. 2, 257–262.

[3] R. H. Gutierrez, P. A. A. Laura: *Common Faults of Rolling Bearings and Their Vibration Signals.* Science & Technology Information *18* (1985), No. 3, 171–180.

[4] R. P. Singh, S. K. Jain: *Concealed Data Aggregation Scheme for Multiple Application in Database as a Service Model.* IJREAT International Journal of Research in Engineering & Advanced Technology *7* (2004), No. 1, 41–52.

[5] M. N. Gaikwad, K. C. Deshmukh: *Multivariate EMD and full spectrum based con-*

dition monitoring for rotating machinery. Mech Syst Sign Process *29* (2005), No. 9, 797–804.

[6] S. Chakraverty, R. Jindal, V. K. Agarwal: *Method for rolling bearing fault diagnosis based on wavelet packet and improved BP neural network.* Modern Electronics Technique *12* (2005) 521–528.

[7] N. L. Khobragade, K. C. Deshmukh: *Fault diagnosis for rolling bearing based on BP neural network of genetical gorithm.* Machine Designand Manufacturing Engineering *30* (2005), No. 4, 555–563.

[8] Y. F. Zhou, Z. M. Wang: *Reducing the dimensionality of data with neural networks.* Science *316* (2008), Nos. 1–5, 198–210.

[9] R. Lal: *Universal approximation using incremental constructive feedforward networks with random hidden nodes.* IEEE Trans. Neural Netw *34* (2003), No. 4, 587–606.

[10] R. Lal, Y. Kumar: *Real-time learning capability of neural networks, Technical Report ICIS/45/2003.* Mechanics of Advanced Materials and Structure *20*, (2013), No. 4, 264–275.

[11] Y. Kumar: *Representational learning with extreme learning machine for big data.* IEEE Intell.Sys *18* (2013), No. 2, 589–597.

[12] J. R. Kuttler, V. G. Sigillito: *Application of Teaching Data Mining Based on Cloud Computing in the Prediction of Learning Achievement.* International Journal of Database Theoryand Application *78* (1981), No. 4, 585–590.

[13] K. M. Liew, K. Y. Lam: *Deterministic neural classification.* Neural Computation *27* (1991), No. 2, 189–203.

[14] K. M. Liew: *Extreme Learning Machine for Multilayer Perceptron.* IEEE Transactions on Neural Networks and Learning Systems *29* (1992), No. 24, 3087–3097.

Received November 16, 2017