ORIGINAL ARTICLES

Research on Fault Diagnosis of Air Conditioner Based on Deep Learning

Zhiting Liu[1], Yuhua Wang*[1], Yuexia Zhou[2]

[1] College of Mechanical and Electrical Engineering, Foshan university, Foshan, China
[2] College of Electronics and Information Engineering, Foshan university, Foshan, China

Received: November 26, 2018 **Accepted:** December 7, 2018 **Online Published:** December 29, 2018
DOI: 10.5430/ijrc.v2n1p18 **URL:** https://doi.org/10.5430/ijrc.v2n1p18

ABSTRACT

The essence of intelligent fault diagnosis is to classify the feature of faults by machine learning. It is difficult and key to extract fault characteristics of signals efficiently. The general feature extraction methods include time frequency domain feature extraction, Empirical Mode Decomposition (EMD), Wavelet Transform and Variational Mode Decomposition (VMD). However, these methods require a certain prior experience and require reasonable analysis and processing of the signals. In this paper, in order to effectively extract the fault characteristics of the air conditioner's vibration signal, the stacked automatic encoder (SAE) is used to extract the feature of air conditioner's vibration signal, and the Softmax function is used to identify the air conditioner's working condition. The SAE performs unsupervised learning on the signal, and Softmax function performs supervised learning on the signal. The number of hidden layers and the number of hidden layer's nodes are determined through experiments. The effects of learning rate, learning rate decay, regularization, dropout, and batch size on the correct rate of the model in supervised learning and unsupervised learning are analyzed. Thereby realizing the fault diagnosis of the air conditioner. The recognition correct rate of deep learning model reached 99.92%. The deep learning fault diagnosis method proposed in this paper is compared with EMD and SVM, VMD and SVM two kind of fault diagnosis methods.

Key Words: Air conditioner, Fault diagnosis, Deep learning, Automatic encoder

1. INTRODUCTION

Air conditioner manufacturers need to test the products on the production line according to the vibration signal of the air conditioner. Intelligent and effective detection methods need to be used to improve manufacturer's productivity.

The key of intelligent fault diagnosis lies in fault feature extraction. Traditional feature extraction methods include time frequency domain feature extraction, Empirical Mode Decomposition (EMD), wavelet transform and Variational Mode Decomposition (VMD). Shen[1] used 29 time domain and frequency domain features for gear fault diagnosis. Zhang Jinmin[2] performed wavelet transform on the vibration signal of gearbox to obtain the nodes, and compose the energies of nodes into a feature vector. Yang Yu[3] performed EMD on the vibration signal of the rolling bearing, selecte several Intrinsic Mode Function (IMF) containing the main fault information, and extract the energies of the IMFs as the feature vector. An bang[4] performed VMD on the gearbox vibration signal, and use the energy percentages and information entropy of the IMFs as the feature vector, achieved fault diagnosis of the gearbox. However, there are some problems in traditional feature extraction methods. For example, the

*Correspondence: Yuhua Wang; Email: 794661261@qq.com; Address: College of Mechanical and Electrical Engineering, Foshan university, Foshan, 528000, China.

feature extraction of time-frequency requires active selection. Its difficult to determine the wavelet basis of the wavelet transform. The EMD have endpoint effect and fault modes at low frequencies. The number of IMFs decomposed by VMD is difficult to determine.

Since 2006, Hinton first proposed deep learning.[5] Deep learning has obvious advantages in speech recognition and image classification due to its strong learning representation ability.[6,7] In the field of mechanical fault diagnosis, Zhou Funa[8] used the deep learning and Principal Component Analysis (PCA) to realize early fault diagnosis and life prediction. Guo Liang[9] used the deep learning network composed of sparse autoencoder to to identify bearing states.

In this paper, the deep learning network composed of Automatic Encoder (AE)[10] is used to extract the feature of air conditioner's vibration signal, and Softmax function[11] is used to identify air conditioner's six working conditions. The fault diagnosis method base on deep learning is compared with the fault diagnosis methods of EMD+ support vector machine (SVM) and VMD+SVM.

2. Basic theory of deep learning

2.1 Automatic Encoder

The Automatic Encoder (AE) is divided into three layers: input layer, hidden layer and output layer. The network from the input layer to the hidden layer is the encoding network. The network from the hidden layer to the output layer is the decoding network. When the nodes of hidden layer are lesser than nodes of input layer, the encoding network maps the high dimensional signal to the low dimension space by the activation function. The decoding network decodes the hidden layer data to restore the input layer data. The gradient descent optimization algorithm is used to tune the weights and biases of the network until the loss function reaches a minimum value. The AE model is shown in Figure 1.

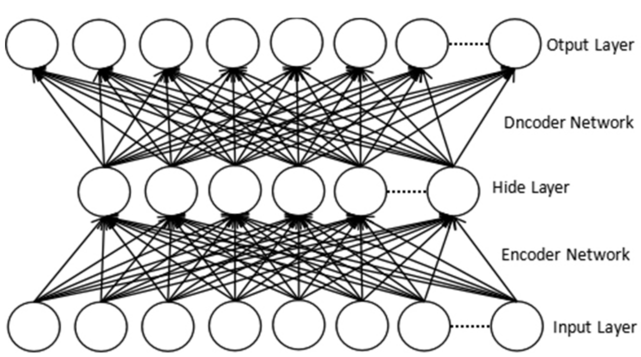

Figure 1. Automatic Encoder model

The tanh function is selected as activation function, and the input of the input layer is $\{x_1^m, x_2^m, \cdots, x_n^m\}$ (n input samples with dimension m), the hidden layer is $\{h_1^k, h_2^k, \cdots, h_n^k\}$ (n feature vectors with dimension k) and the output layer is $\{\hat{x}_1^m, \hat{x}_2^m, \cdots, \hat{x}_n^m\}$ (n reconstructed samples with dimension m), the formula of the tanh activation function is:

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{1}$$

The activation value of the hidden layer is:

$$h = f_{\theta_1}(z) = f(w_1 x + b_1) \tag{2}$$

Where: x is the input sample, θ_1 is the encoding network parameter $\{w_1, b_1\}$, w_1 is weight matrix, and b_1 is bias. The activation value of the output layer is:

$$\hat{x} = f_{\theta_2}(z) = f(w_2 h + b_2) \tag{3}$$

Where θ_2 is the decoding network parameter $\{w_2, b_2\}$, w_2 is weight matrix, and b_2 is bias.

The purpose of the AE is to find the optimal network parameter $\theta^* = \{w_1^*, b_1^*, w_2^*, b_2^*\}$, make the loss function reach the minimum, and the loss function $L(\theta)$ expression is:

$$L(\theta) = L(w_1, b_1, w_2, b_2) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\left(x^i - \hat{x}^i\right)^2 + \frac{\lambda}{2}\sum_{l=1}^{s}\sum_{i=1}^{m}\sum_{j=1}^{k}\left(w_{ij}^l\right)^2 \tag{4}$$

Where: The first item is the sum of the errors of n input samples and n reconstructed samples, x^i is ith sample, $\hat{x}^i$ is the ith reconstructed sample. The second term is L2 regularization that prevents overfitting.

2.2 Feature Extraction based on Stacked Automatic Encoder

Single autoencoder has a limited ability to construct nonlinear function. SAE is stacked by multiple AEs, and SAE has stronger nonlinear expression ability than AE. The hidden layer of the first AE is used as the input layer of the second AE, and the hidden layer of the second AE is used as the input layer of the third AE. An n-layer SAE can be composed by stacking n AEs in this way. A N-layer SAE model is shown in Figure 2.

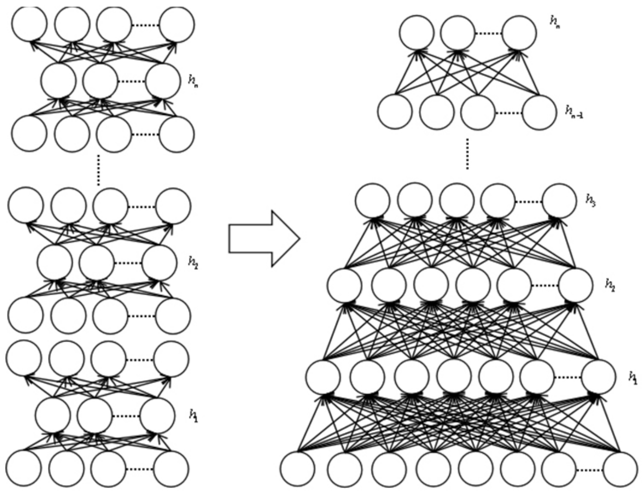

Figure 2. N-layers SAE model

2.3 Softmax Function

The Softmax[11] function is for multi-classification problem. the formula of Softmax function is:

$$y_i = \varsigma(x^i) = \frac{e^{x^i}}{\sum_{j=1}^{k} e^{x^j}} \tag{5}$$

Where $i = 1, 2, 3, \cdots, k$ is the output label and x is the input.

The training samples are $\{(x^1, y^1) \cdots (x^n, y^n)\}$, x is a vector with dimension m, and label y can be k different values, network parameter $\theta = \{w_{ij}, b_j\}, j - 1, 2, \cdots, K$. The probability $p = (y = j|x^j; \theta)$ represents the probability that the sample is discriminated as the j class when the input is x^i and the network parameter is θ. A k-classed Softmax function will output a k-dimensional vector, and the sum of the vector is 1. The output of the Softmax function when network parameter θ and input x^i is

$$y_\theta(x^i) = \begin{pmatrix} p(y^i = 1 \mid x^i; \theta) \\ p(y^i = 2 \mid x^i; \theta) \\ \vdots \\ p(y^i = k \mid x^i; \theta) \end{pmatrix} = \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T; x^i)} \begin{pmatrix} \exp(\theta_1^T; x^i) \\ \exp(\theta_2^T; x^i) \\ \vdots \\ \exp(\theta_k^T; x^i) \end{pmatrix} \tag{6}$$

Where θ_j is the weight vector connecting the jth output node, the sum of $y_\theta(x^i)$ is 1. The loss function of Softmax function is:

$$L(\theta) = \sum_{i=1}^{n} \sum_{j} y_i^j \log \frac{\exp(\theta_j^T x^i)}{\sum_{p=1}^{k} \exp(\theta_p^T x^i)} \tag{7}$$

Where y_i^j is the label of the jth sample.

The optimal parameter θ^* is solved by minimizing $L(\theta)$.

2.4 Deep Learning Network

The N-layer SAE and Softmax function are combined to form a deep learning network. The model is shown in Figure 3.

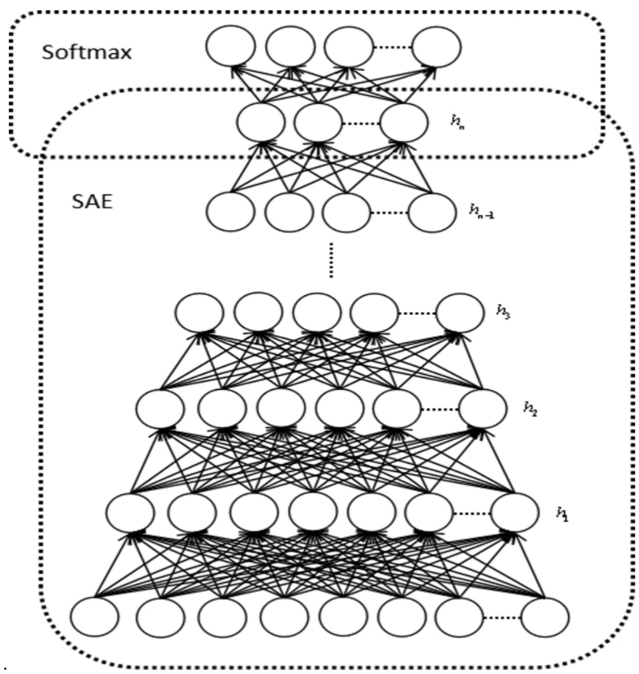

Figure 3. Deep Learning Network

The process of deep learning network identify signals is divided into two stages. In the first stage, the SAE is used to extract the feature layer by layer, the gradient descent algorithm was used to pre-training AEs one by one, this part is the unsupervised learning stage. In the second stage, the activation values of the final hidden layer are uesd as a feature vector input Softmax, the gradient descent algorithm[12] is used to fine tune the global network weight and bias value, this part is the supervised learning stage. The algorithm flow is shown in Figure 4.

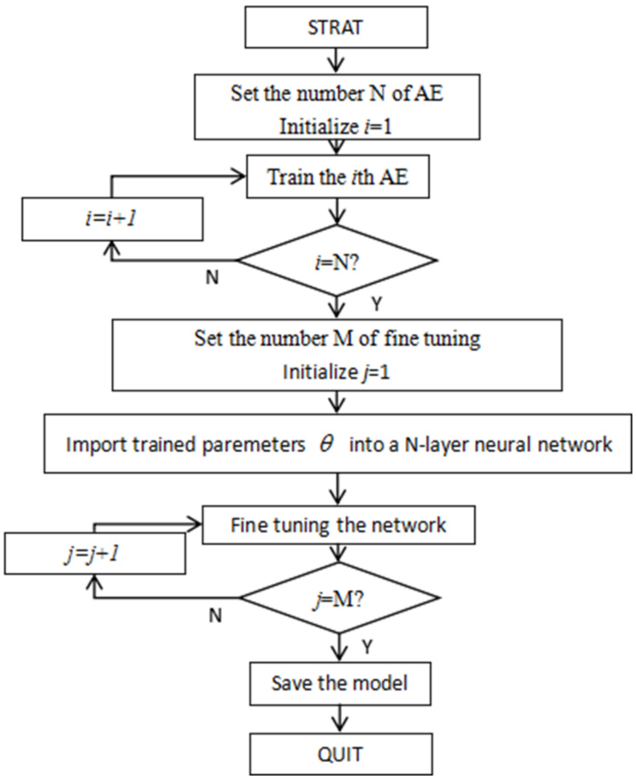

Figure 4. Flowchart for training deep learning network

3. Experimental data analysis

3.1 Experimental Data Acquisition

The main vibration sources of air conditioner are fans and compressors motors. According to previous research and analysis,[13] the best data collection point is shown in Figure 5.

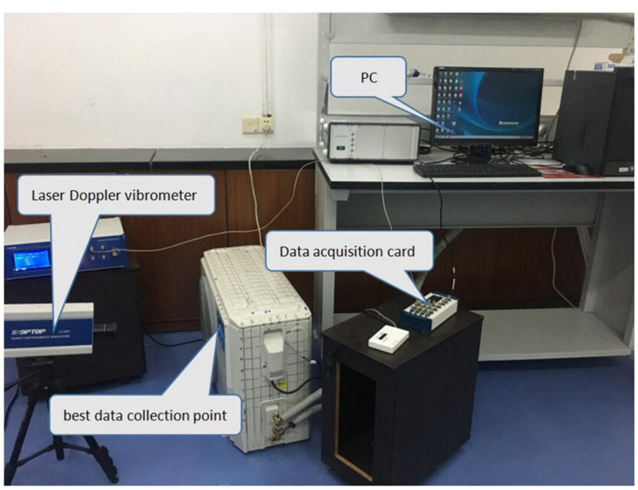

Figure 5. Air conditioning vibration acquisition platform

The vibration signals of the air conditioner are collected by the Laser Doppler Vibrometer (LDV) LV-S01, the sampling rate is 4000, and the sampling time is 2 second. Sampling 2000 sets of data for each condition, 80% of the samples are used as training sets and 20% are used as test sets, the six working conditions are shown in Table 1.

Table 1. Six working conditions of air conditioner

Label	Working condition	Number of samples
1	normal	2,000
2	Compressor lacks soundproof cotton	2,000
3	Fan blade cracks	2,000
4	fan blade weight imbalance	2,000
5	Compressor copper tube lacks damping block	2,000
6	Fan bracket loose	2,000

3.2 Analysis of Experimental Data

The spectrums of the six working conditions are shown in Figure 6.

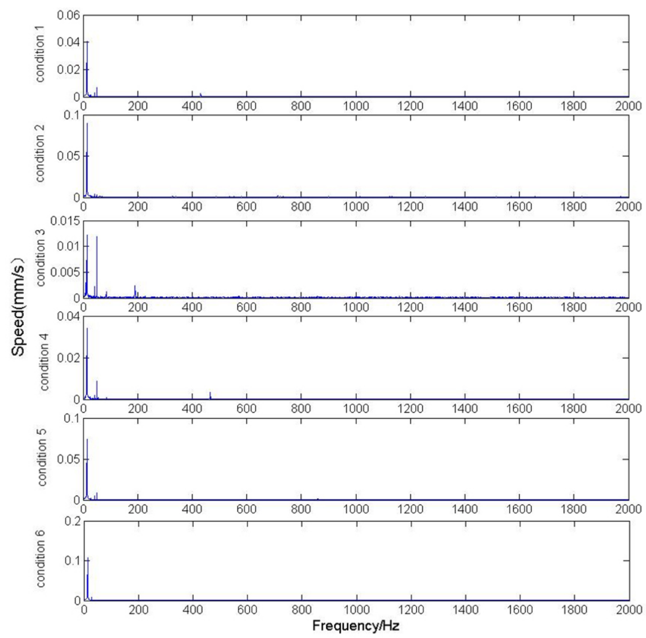

Figure 6. Spectrums of six working condition

Normalize the signals before they are entered the network.

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

The correct rates of different input signals are different. Time domain signals, frequency domain signals, time domain + frequency domain signals are inputted to deep learning network separately. After a lot of experiments, it is found that the classification accuracy of the frequency domain signal is the highest, the time domain + frequency domain is the second, and the time domain signal is the lowest.

Through a lot of experiments, it is found that the number of nodes of hidden layer and the number of hidden layers have great influence on the recognition correct rate. The deep learning network with hidden layer number N (N = 1, 2, 3, 4, 5) and hidden layer node number M (M = 100, 200, 300, 400) is constructed to find the best N and M. (learning rate is 0.5, Learning rate decay coefficient is 1, L2 Regularizati-on coefficient is 0, batch size are 500, Dropout coefficient is 0, the epochs are 100, and the epochs are 1,000). It can be seen from Fig. 7 that each curve is a convex function. When the M is 100, the correct rate curve is higher than the correct rate curve of other. The number of network nodes is selected as 100, and the number of hidden layers is selected as 3.

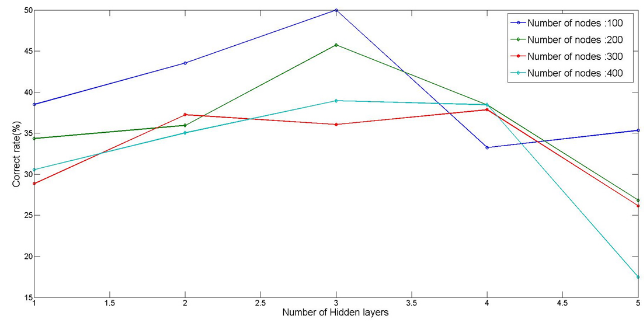

Figure 7. The correct rate of the model under different node numbers and hidden layer numbers

Table 2. Six working conditions of air conditioner

Parameter	**A0**	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**
SAE learning rate	0.1	0.05	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SAE learning rate decay	1	1	0.99	1	1	1	1	1	1	1
SAE L2 regularization	0	0	0	0.1	0	0	0	0	0	0
SAE batch size	800	800	800	800	800	9,600	600	100	10	1
SAE dropout	0	0	0	0	0.1	0	0	0	0	0
Softmax learning rate	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Softmax learning rate decay	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Softmax L2 regularization	0	0	0	0	0	0	0	0	0	0
Softmax batch size	800	800	800	800	800	800	800	800	800	800
Softmax dropout	0	0	0	0	0	0	0	0	0	0
Correct rate (%)	67.7	83.9	73.2	76.3	75.9	78.0	81.1	90.0	93.1	75.7
Parameter	**B0**	**B1**	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B9**	**b11**
SAE learning rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
SAE learning rate decay	1	1	1	1	1	1	1	1	1	1
SAE L2 regularization	0	0	0	0	0	0	0	0	0	0
SAE batch size	800	800	800	800	800	800	800	800	800	800
SAE dropout	0	0	0	0	0	0	0	0	0	0
Softmax learning rate	0.01	0.001	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Softmax learning rate decay	1	1	0.99	1	1	1	1	1	1	1
Softmax L2 regularization	0	0	0	0.01	0	0	0	0	0	0
Softmax batch size	800	800	800	800	800	9,600	600	100	10	1
Softmax dropout	0	0	0	0	0.1	0	0	0	0	0
Correct rate (%)	70.66	96.84	79.55	81.98	85.12	38.77	99.78	99.80	99.86	87.65

The learning rate, learning rate decay, regularization, Dropout, batch size and other parameters have certain influence on the correct rate of model. If the learning rate is too large, the model cannot converge to the optimal solution, conversely, the network converge slowly. Learning rate decay can change the learning rate after each iteration so that the learning rate changes dynamically.The model learning start learning rate is large, which makes the model converge toward the optimal solution quickly. The learning rate is small at the late training, so that the model converges to the optimal solution accurately to prevent the model from oscillating. regularization can prevent model's weight parameters from being too large, thus preventing over fitting. Batch size determines the direction of the decline. Batch size should not be chosen too small or too large. If the batch size is too small, the model can't easily converge, or the convergence needs to go through a large period. If the batch size is too large, the parameter correction speed may be slow due to the decrease of the number of iterations. Dropout[14] can block the update of some parameters of the model. It is a way to prevent overfitting, too.The deep learning network base on SAE is divided into two stages: unsupervised learning and supervi-sed learning. First, the parameters of the unsupervised learning phase are analyzed by single variable analysis, and then the parameters of the supervised learning phase are analyzed by single variable analysis. the settings of different parameters are shown in Table 2.

According to Table. 2, whether in the unsupervised training stage or the supervision training stage, the correct rate of the model is most affected by the learning rate, and the learning rate decay, L2 regularization and dropout can improve the correct rate slightly, the batch size is as small as possible, but it cannot be smaller than the number of labels. The correct rate of supervised learning is higher than that of unsupervised learning. This shows that supervised learning has a greater impact on the correctness of the network when adjusting network parameters. The network parameters are adjusted to the optimal value to train the network (In unsupervised learning stage, the learning rate is 0.1, the learning rate decay coefficient is 0.99, the L2 regularization coefficient is 0.1, the dropout coefficient is 0.1, the batch size is 10. In the supervised learning stage, the rate is 0.01, the learning rate decay coefficient is 0.99, the L2 regularization coefficient is 0.01, the dropout coefficient is 0.1, and the batch size is 10). The final result shows the correct rate of deep learning model reached 99.92%.

3.3 Compared with other fault diagnosis methods

The fault diagnosis method based deep learning is compared with the fault diagnosis methods of EMD+SVM and VMD+SVM.

EMD is performed on six kind of signals, the noise component IMF 1[15] and the fault modes are removed. The Root Mean Square (RMS) of IMF 2-IMF 6 are input as a feature vector to the SVM.

In order to determine the number of IMFs decomposed by VMD, the leaked energy is used to determine the number of IMFs.[16] The signals are decomposed into 7 IMFs, the correlations between IMFs and the original signal are used to select the appropriate IMFs, the sample entropy of IMF 1-IMF 5 are input as a feature vector to the SVM. The fault diagnosis method proposed in this paper is compared with the above two fault diagnosis methods. The comparison result is shown in Table 3.

Table 3. Comparison of different diagnostic methods

Classification method	Correct rate (%)
EMD+SVM	91.92
VMD+SVM	93.58
Deep learning	99.92

4. Conclusion

This paper provided a new idea for the fault diagnosis of air conditioners. The deep learning network composed of SAE is used to extract the feature of air conditioner's vibration signals, softmax function is uesd to deal with classification problems, which overcomes the shortcomings of traditional fault diagnosis methods: It requires a certain prior experience and reasonable analysis and processing of signals. The influence of deep learning network parameters on the correct rate of the network in the stage of supervised learning and unsupervised learning is analyzed, which provides a reference for the parameter setting of the deep learning network. The final correct rate of the experimental model is 99.92%.

Acknowledgements

This work is supported by the Special Fund for Public Welfare Research and Capacity Building in Guangdong Province No. 2015A010103017 and No. 2015B010101014.

References

[1] Shen Z, Chen X, Zhang X, He Z. A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM. Measurement. 2012; 45(1): 30-40. https://doi.org/10.1016/j.measurement.2011.10.008

[2] Zhang JM, Pei YQ, Wang S. Fault Diagnosis of Fan Gearbox Based on Wavelet Decomposition and Least Squares Support Vector Machine. Sensors and Microsystems. 2011; 30(1): 41-43.

[3] Yang Y, Yu DJ, Cheng JS. Fault Diagnosis Method of Rolling Bearing Based on EMD and Neural Network. Vibration and Shock. 2005; 24(1): 85-90.

[4] An B, Pan HX, Zhang Y, Zhang YX, Zhao XP. Fault diagnosis of gearbox using VMD and multi-parameter fusion. Combined Machine and Automation Technology. 2017; (4): 92-95.

[5] HINTON GE, SIMON O, Yee-Whye T. A fast learning algorithm for deep belief nets. Neural Computation. 2006; 18(7): 1527-1554. PMid:16764513. https://doi.org/10.1162/neco.2006.18.7.1527

[6] Dahl GE, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. Audio, Speech, and Language Processing. IEEE Transactions on: 2012; 20(1): 30-42. https://doi.org/10.1109/TASL.2011.213409

0

[7] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(8): 1798-1828. PMid:23787338. https://doi.org/10.1109/TPAMI.2013.50

[8] Zhou F, Gao YL, Wang JY, Wen CL. Early diagnosis and life prediction of slowly changing faults based on deep learning. Journal of Shandong University. 2017; 47(5): 30-37.

[9] Guo L, Gao HL, Zhang YW, Huang HF. Research on Bearing State Recognition Based on Deep Learning Theory. Vibration and Shock. 2016; 35(12): 166-170, 195.

[10] BENGIO Y. Learning deep architectures for AI. Foundations and Trends® in Machine Learning. 2009; 2(1): 1-127. https://doi.org/10.1561/2200000006

[11] Wang HB, Chen YX, Li YQ. Face recognition method based on principal component analysis and Softmax function regression model. Journal of Hefei University of Technology. 2015; 38(6): 759-763.

[12] Hou YB, Du JY, Wang M. Neural Network. Xian: Xi an University of Electronic Science and Technology Press; 2007: 35-39.

[13] Zheng WW. Study on Six-dimensional Vibration Detection System for Air-conditioning. Matser, foshan university, foshan, china. 2018.

[14] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. A Simple Way to Prevent Neural Networks from Overfitting. 2014; (15): 1929-1958.

[15] Wu ZT, Yang SX. A new method for fault feature extraction and pattern classification of rotating machinery. Beijing: Science Press; 2012: 108-115.

[16] Jiang WL, Wang ZW, Zhu Y, Dong K, Zhang S. Early fault identification of rolling bearings based on VMD denoising. Hydraulics & Pneumatics. 2017; (5): 13-20.