

PAPER • OPEN ACCESS

Fault Identification of Vehicle Automatic Transmission based on Sparse Autoencoder and Support Vector Machine

To cite this article: Canyi Du *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **490** 072050

View the [article online](#) for updates and enhancements.



IOP | ebooksTM

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

Fault Identification of Vehicle Automatic Transmission based on Sparse Autoencoder and Support Vector Machine

Canyi Du¹, Shaohui Zhang², Zusheng Lin² and Feifei Yu^{3,*}

¹School of Automobile and Transportation Engineering, Guangdong Polytechnic Normal University, Guangzhou, China

²School of Mechanical and Automotive Engineering, Xiamen University of Technology, Guangzhou, China

³School of Mechatronic Engineering, Guangdong Polytechnic Normal University, Guangzhou, China

*Corresponding author's E-mail address: 83111503@qq.com

Abstract: Support vector machine(SVM) got a good classification ability, but the recognition accuracy was easily affected by the value of the kernel parameters. Aiming at this problem, sparse autoencoder(SAE) has its unique advantages in dealing with complex structured data, so the combination of sparse autoencoder and support vector machine(SAE+SVM) was proposed on the fault identification of vehical automatic transmission. Firstly, eight indicators such as engine speed, throttle opening, water temperature and so on are collected from acquisition automobile automatic transmission under 3 running conditions. The data was used as input dataset of the sparse autoencoding model to extract the features. Then the features was used for the fault classification and identification based on support vector machine. Compared with using support vector machine only, the experiment results showed that the recognition accuracy based on the combination of sparse autoencoder and support vector machine(SAE+SVM) was less affected by the value of the kernel parameters and got better recognition accuracy. So the combination of sparse autoencoder and support vector machine can be better used in the real-time fault identification and diagnosis of automatic transmission.

1. Introduction

Automatic transmission of automobile is a precise electromechanical device which integrates mechanical, hydraulic and electric control systems. Its structure and working principle are complex, and its fault diagnosis is difficult. Although modern automotive self-diagnostic systems can detect whether automatic transmission sensors and actuators work properly. However, mechanical fault, hydraulic system fault and comprehensive fault often go beyond the monitoring scope of self-diagnosis system, and the traditional technical means have relatively large limitations. Therefore, it is of great significance to study the intelligent fault diagnosis of automatic transmission. In reference [1]-[3], the paper studies the use of artificial neural network technology to diagnose the faults of automobile transmission. However, the shortcomings of artificial neural network, such as slow convergence speed, local minimum point, over-learning or under-learning and requirement to a large number of learning samples, restricts the extensive application of neural network in fault diagnosis. The author used Support Vector Machine (SVM) to diagnose the faults of automotive automatic



transmission and got a good recognition accuracy in the early stage (literature [4]). However, when the number of training samples was small, the recognition accuracy of SVM was not ideal, only 96.1%. Moreover, the recognition accuracy of SVM was easily affected by the value of the kernel parameters, and different kernel parameters lead to large variable section of recognition accuracy. As a hot topic of data mining and feature extraction, depth learning theory has its unique advantages in dealing with complex structured data. However, at present, the application of deep learning in mechanical equipment mainly takes mechanical vibration data as input information [5-10], while there are few studies on non-vibration signals. Combining the advantages of support vector machine and depth learning, the depth learning model is used to extract the features of the collected data in advance, and then the obtained features are input into the support vector machine to further improve the recognition accuracy. At this point, support vector machines enhance the robustness of kernel parameters.

By setting three automatic transmission states and collecting characteristic signals of eight indicators such as engine speed, throttle opening, water temperature, speed, spindle speed, solenoid valve A for gear shifting, solenoid valve B for gear shifting and solenoid valve C for gear shifting, the sparse self-coding model was applied to fault recognition to obtain better recognition accuracy.

2. Principle of sparse autoencoder

Sparse autoencoder is an unsupervised algorithm. The hidden layer features of input data are obtained by encoding and decoding the characterization of input data, so as to reduce the dimension and improve the classification effect. The model example is shown in Figure 1.

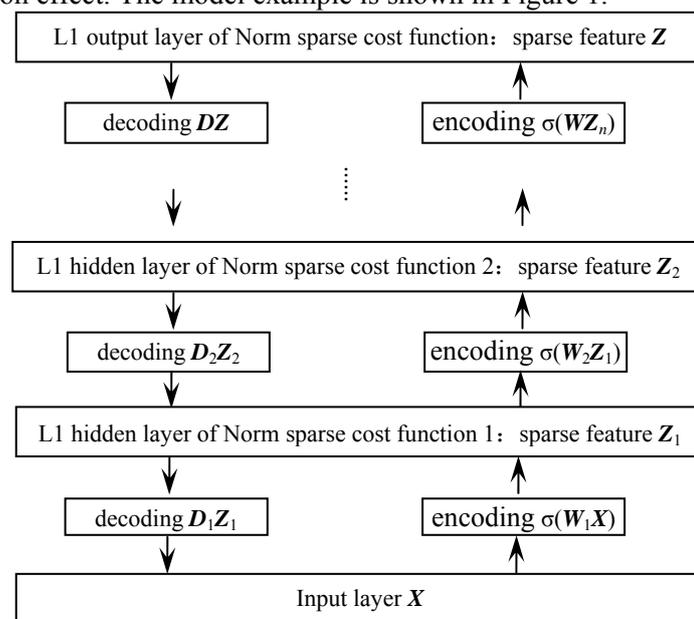


Figure 1. The principle flow of sparse autoencoder algorithm

In order to make the output in each hidden layer represent the input pattern as much as possible, the weighting parameters D , W are constantly revised to minimize the cost function:

$$\min_{D_{i+1}, W_{i+1}, z_{i+1}} \|D_{i+1}z_{i+1} - z_i\|_2^2 + \lambda \|z_{i+1}\|_1 + \|\sigma(W_{i+1}z_i) - z_{i+1}\|_2^2 \quad (1)$$

The first and second terms of the formula are the encoding process, the third is the decoding process. W_{i+1} , D_{i+1} , z_i , z_{i+1} represent the encoding weight of the i th hidden layer, the decoding weight, the input of the i th hidden layer and the output of the i th hidden layer, respectively. When $i=0$, $z_0=X$, λ is used to control the relative importance of sparse penalties; $\|z_{i+1}\|_1$ is the L_1 -norm, used to control the sparse degree of output; σ is the Sigmoid function, the formula is as follows:

$$z_{i+1} = \sigma(z_i) = \frac{1}{1 + \exp(-W_{i+1}z_i)} \quad (2)$$

Sparse autoencoder makes only a small number of points in the hidden layer active by sparsifying the hidden layer nodes, avoiding the problem of homogenization of hidden layer node features, and

has good robustness. As can be seen from figure 1, in theory, better recognition effect can be obtained through the hierarchical abstract transformation of the input data $X (Z_1, Z_2, \dots, Z_n, Z)$. However, the number of different features directly affects the final recognition result. When the number of hidden layers is determined, the number of features of each hidden layer needs to be the best matched to make the final output features have better classification performance. There is no definite method to determine the number of features of each hidden layer. For this problem, the method of parameter optimization is used to determine the number of features of each hidden layer.

3. Diagnosis process of automatic transmission based on sparse autoencoder

The research object was a 4-speed automatic transmission of a Honda car. According to the structure and control principle of automatic transmission and on the basis of analyzing the failure law and working characteristics of transmission, the characteristic parameters reflecting the working state of transmission are selected as the low-level input of sparse autoencoder, including eight signal characteristics, which were speed signal of engine, throttle opening signal, speed signal, spindle speed signal, coolant temperature signal, switching signal of the electromagnetic valve and so on.

The transmission is an automatic transmission controlled by electro-hydraulic system. According to the input signal of each sensor, the ECU of the electronic control unit changes the oil circuit of the hydraulic control system by controlling the on and off of the solenoid valve, thus realizing automatic shift control.

The electronic control system of transmission consists of control module PCM (Power Control Model), sensor and actuator(three shift control solenoid valves, two pressure control solenoid valves for A/T clutch, one lock control solenoid valve, etc.) According to the position of shift lever, throttle position, vehicle speed, engine speed, speed of transmission spindle and intermediate shaft, and water temperature, the control module realizes shift time controlling, torque converter locking, slope mode controlling and other functions. When the PCM receives the input signal of each sensor and needs to shift, the PCM will control the shifting solenoid valves A, B, C and pressure control solenoid valves for A/T clutch A, B and other actuators for corresponding action. The first three will change the position of the shift valve to determine the required gear, the latter two will adjust the corresponding pressure, so as to achieve accurate and smooth transmission between high and low gear.

3.1. Test methods and data collection

Due to the limitations of actual test conditions, only two kinds of faults are set, which are: fault 1 (stop the control signal of the shift solenoid valve B) and fault 2 (stop the control signal of the speed sensor of the intermediate shaft). In these two failure modes, the shift and operation of the automatic transmission will have certain fault characteristics. Considering people's usual driving habits, two driving modes were chosen to test: (1) keep the throttle opening around 5° , that is, the car start accelerating from static when the throttle is slightly opened; (2) keep the throttle opening around 10° , that is, the car start accelerating from static when the throttle is slightly larger until the speed reaches a higher speed.

1) In the normal state of the automatic transmission without failure, the accelerator should be quickly pressed to the position of about 5° and kept unchanged, enabling the car to accelerate from the static state until the speed reaches about 45km/h, and recording relevant data of the whole process.

2) In the state of fault 1, when the fault indication lamp of automatic transmission is on, similarly, the accelerator should be quickly pressed to the position of about 5° and kept unchanged, enabling the car to accelerate from the static state until the speed reaches about 45km/h, and recording relevant data of the whole process.

3) In the state of fault 2, when the fault indication lamp of automatic transmission is on, similarly, the accelerator should be quickly pressed to the position of about 5° and kept unchanged, enabling the car to accelerate from the static state until the speed reaches about 45km/h, and recording relevant data of the whole process.

4) In the above 3 cases, the same method is used to change the throttle opening to about 10° , and then redo.

The above data collection is completed through a specially designed data collection system based on virtual instrument technology[1]. In order to achieve real-time collection, the system program sets a

high sampling rate, and 5 sets of data can be collected synchronously per second. At the time of analysis, only one set of data is taken per second for analysis. Because in this one second, the change rate of data is not large, and the trend is the same.

3.2. Data curve analysis

In order to eliminate the influence of feature dimension, the eight data features collected are normalized, and the sample data is quantized in the range of [0,1]. The specific eight characteristic indicator curves are shown in figure 2.

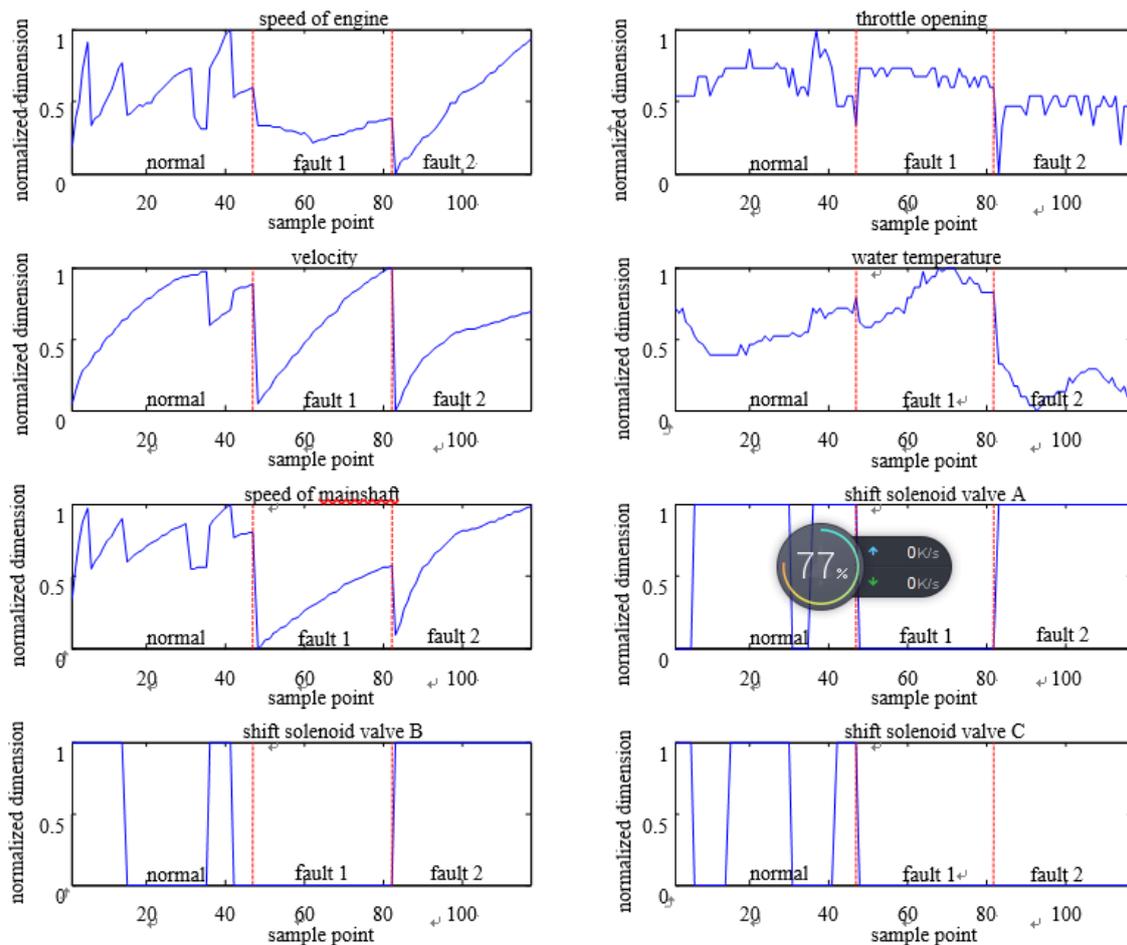


Figure 2. The specific eight characteristic indicator curves

From the characteristic curve diagram, there is a large overlap between the three States, and no specific index can be better divided into three categories. The engine speed signal, throttle opening signal, speed signal, spindle speed signal and coolant temperature signal have strong nonlinearity. It is difficult for human experience to judge the state of transmission from these indexes. Therefore, artificial intelligence is used to fuse these features and extract more effective features to improve diagnostic accuracy.

4. Analysis of experimental results

In order to facilitate the analysis of the recognition accuracy of the diagnostic model under different training samples, the training samples and test samples are set up according to reference [4]. mode 1: training samples 80, test samples 37; Mode 2, training sample 60, test sample 57; Mode 3: training sample 40, test sample 77. All the training samples were randomly taken from 117 groups of sample data, repeated 50 times, and the average of the classification accuracy of the test samples was evaluated. At the same time, the influence of nuclear parameters on recognition's results was compared. As can be seen from the literature [4], the support vector machine can obtain better recognition accuracy than the BP neural network model, so the sparse self-coding is compared with the SVM. It is

proved that using sparse self-coding can get better accuracy of recognition. Table 1 is the corresponding result of recognition, in which the accuracy of recognition is the maximum value of the range of nuclear parameters.

Table1. The comparison of recognition accuracy with the SVM、SAE and SAE+SVM

Training samples	Test samples	diagnostic model		recognition accuracy
80	37	SVM	$\sigma=0.06$	0.9951
		SAE	8-15-3-3	0.9951
			8-5-3-3	0.9832
		SAE+SVM	8-15-3-3	0.9951
8-5-3-3	0.9854			
60	57	SVM	$\sigma=0.06$	0.9814
		SAE	8-15-3-3	0.9849
			8-5-3-3	0.9895
		SAE+SVM	8-15-3-3	0.9888
8-5-3-3	0.9919			
40	77	SVM	$\sigma=0.06$	0.9517
		SAE	8-15-3-3	0.9725
			8-5-3-3	0.9491
		SAE+SVM	8-15-3-3	0.9745
8-5-3-3	0.9655			

As shown in table 1, with the increase of training samples, the accuracy rate of the model recognition increased significantly. The sparse self-coding algorithm obtains that the accuracy rate of recognition is higher than that of SVM when the number of training samples is 40 and 60, and equal to that of SVM when the training samples are 80. The results show that SAE can extract the effective features of data set better and get better results of recognition. When using SAE + SVM, the accuracy rate of recognition in three training samples was higher than that of using SAE and SVM methods alone (The three are equal when the training sample are 80).

Fig.3 is the recognition accuracy curve of the SVM and SAE+SVM methods under different kernel parameters, as seen from the graph, when using SVM alone to classify, the recognition accuracy intervals of the three training samples are respectively: [0.8055,0.9517], [0.8982,0.9814], [0.8957,0.9951], the curve appears in the form of rising and falling first, and when the parameter is less than 0.06, as the parameter increases, the accuracy of recognition rises rapidly. When the parameter is greater than 0.06, the accuracy rate of recognition decreases with the increase of the parameter, and the range fluctuation amplitudes are 14.62%, 8.32% and 9.94% respectively. It can be seen that the recognition accuracy of SVM is greatly affected by the kernel parameters, and the accuracy of recognition is closely related to the value of the kernel parameters. This is because the SVM is only a classification means and does not have the function of feature mining for input data, and its performance is closely related to the quality of data features. When the depth structure of SAE+SVM is 8-15-3, the accuracy interval of the three training samples is: [0.9634,0.9745], [0.9709,0.9888], [0.9827,0.9951], and the range fluctuation amplitudes are 1.11%, 1.79% and 1.24% respectively; When SAE+SVM method and depth structure is 8-5-3, the recognition under the three training samples is carried out. The correct intervals were [0.9558,0.9655], [0.9818,0.9919], [0.9746,0.9854], and the range ranges were 0.67%, 1.01% and 1.08%, respectively. This shows that the method combined with SAE+SVM can get more stable accuracy of recognition. At this time, the robustness of SVM is enhanced and the influence of kernel parameters is reduced.

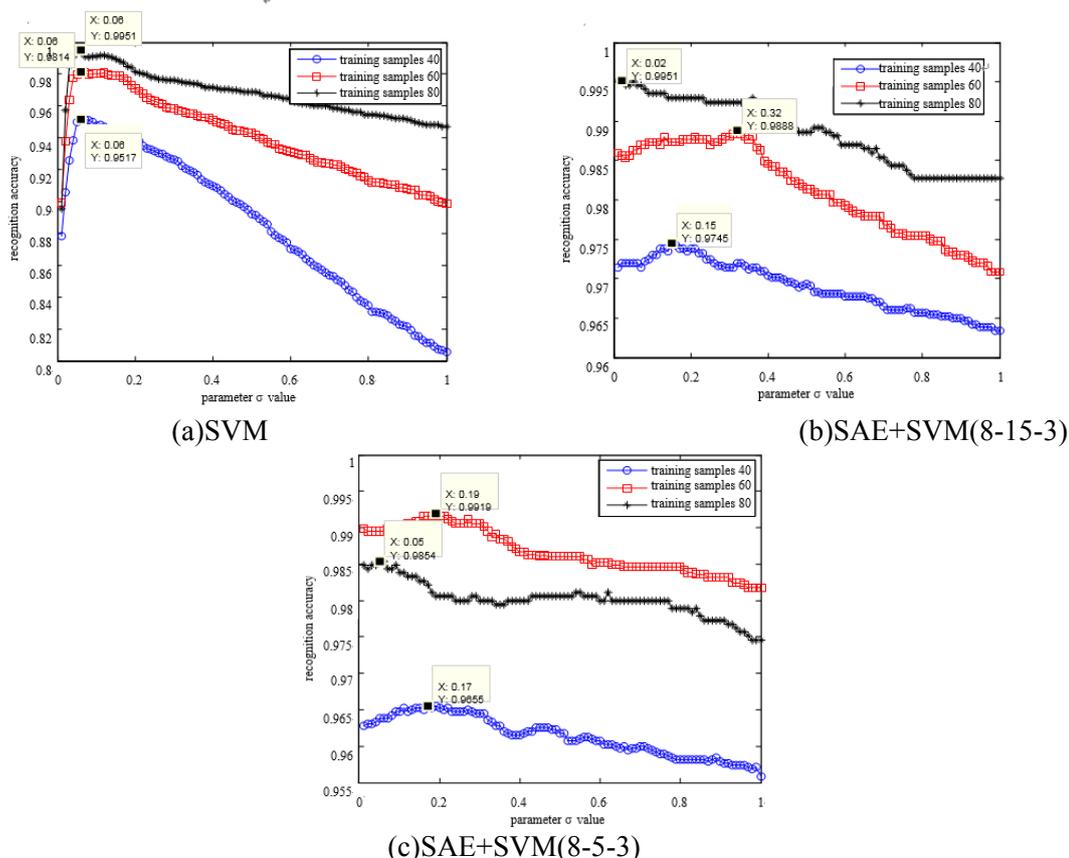


Figure 3. The recognition accuracy curve of the SVM and SAE+SVM methods under different kernel parameters

Fig.4 is the first three feature clustering graphs extracted by sparse self-code under two depth structures. It can be seen that after feature extraction by SAE, similar samples are clustered and different samples are separated distinctly. Input these features into the SVM model can effectively suppress the effect of kernel parameter selection on recognition.

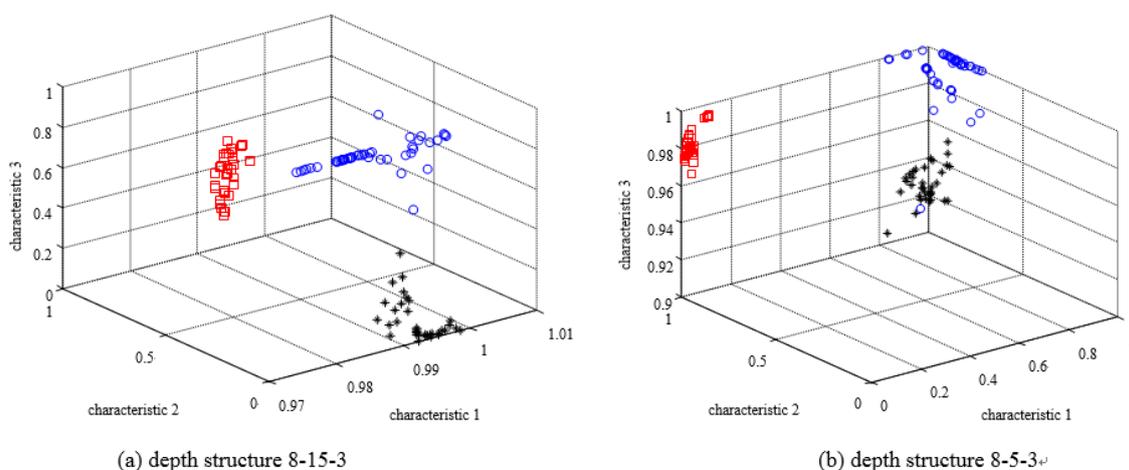


Figure 4. The first three feature clustering graphs extracted by sparse self-code under two depth structures

5. Conclusion

Based on the merits of sparse self-coding algorithm in mining complex structure data, it is combined with support vector machine (SVM) and applied to the fault diagnosis of automatic transmission (AT). The test results of data acquisition under three states of AT show that the integrated sparse self-coding

diagnosis and support vector machine method is more robust than the SVM. In three kinds of training samples, we can get higher classification accuracy, get better recognition clustering effect, and provide a new idea for condition monitoring and fault diagnosis of automobile automatic transmission.

Acknowledgements

This work was financially supported by the Education Research Innovative Project of Guangdong University(2017GXJK102) & Natural Science Foundation of Guangdong Province(2018A030313947).

References

- [1] Du Canyi, Lu Huazhong, Yu Feifei, etc. Research on fault diagnosis method of automatic transmission based on virtual instrument and neural network [J]. Journal of Hubei auto industry Institute, 2006, 20(1): 1-5.
- [2] Du Canyi, Yang Cuili, Pan Wei. Application of support vector machine in auto automatic transmission fault recognition [J]. Automotive engineering, 2012,34(3): 241-244.
- [3] Qinghua Hu, Rujia Zhang, Yucan Zhou. Transfer learning for short-term wind speed prediction with deep neural networks[J]. Renewable Energy, 2016, 85: 83-95.
- [4] Meng Gan, CongWang, Chang'an Zhu. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings[J]. Mechanical Systems and Signal Processing,2016, (72-73), 92-104.
- [5] Li Weihua, Shan Waiping, Zeng Xueqiong. Bearing fault classification based on depth belief network recognition [J]. Journal of vibration engineering, 2016, 29(2): 340-347.
- [6] Sun Wenyi, Shao Siyu, Yan Ruqiang. Fault diagnosis of induction motors based on sparse automatic coding depth neural network [J]. Journal of mechanical engineering, 2016, 52(9): 65-71.
- [7] Lei Yaguo, Jia Feng, Zhou Xin, etc.Methods of health monitoring for large data of mechanical equipment based on deep learning theory [J]. Journal of mechanical engineering, 2015, 51(21): 49-56.