GRU-Based Fault Diagnosis Method for Ball Mill

QU Xingyu[1,2,3], ZENG Peng[1], LI Junpeng[3]

1. *Key Laboratory of Industrial Informatics*, *Shenyang Institute of Automation*, *Chinese Academy of Sciences* 110016, *China*;

2. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*;

3. *Northern Heavy Industries Group Co. ltd*, *Shenyang* 110860, *China*

Abstract: Recently, the fault diagnosis of the ball mill mostly depends on the experience of workers, which brings about a lot of uncertainty for fault diagnosis. In addition, the cost of labor is getting higher, so that the research of ball mill fault diagnosis based on machine learning has become increasingly valuable. The current fault diagnosis methods are mostly judging based on instantaneous data, which makes it difficult to reflect the ball mill indicators and the occurrence of time-related correlation (such as hysteresis effect). This paper presents a ball mill fault diagnosis method based on Gate Recursion Unit (GRU), which analyzes the fault data in the form of time series and compares with other common methods such as neural network, Autoencoder and Long Short-Term Memory (LSTM). After comparison, it is concluded that the fault diagnosis method based on GRU ball mill has the lowest error rate as 4.85%.

Key words: Fault Diagnosis; Deep Learning; RNN; GRU

1 Introduction

The grinding system of ore-dressing process is a typical data-driven industrial process optimization control system, ball mill is the key of grinding system, that is core equipment. Due to the characteristic of large time delay, large inertia, multi-variable strong coupling and complex working mechanism, control of ball mill water flow is very complicated. With the large-scale and intelligent development of ball mill, advanced control method can ensure smooth operation of equipment, improve production efficiency, reduce production costs, while the traditional control method cannot meet the existing demand. The running data of ball mill in the process of running contained the related information. At present, the use of the information is mainly real-time view and historical data query, which is "information waste"[1]. How to mine and analyze big-data during the operation have become the research focus of complex industrial process optimization control, and using data analysis method to diagnose the fault in the operation of equipment has become one of the difficulties in the research.

The fault diagnosis methods based on the traditional mathematical model are divided into quantitative knowledge methodology and mathematical analytic model.

Reference[2] proposes an improved envelope spectral fault diagnosis method based on Empirical Mode Decomposition (EMD) and Spectrum Kurtosis (SK).

In reference[3], a method to detect the fault of the rotor bar of the asynchronous motor based on the square function of the Park vector model is proposed. The simulation results show that, in the case of short data, compared to FFT analysis technique, the method has higher frequency resolution, accuracy rate, and lower calculation complexity, and is conducive to real-time monitoring of motor failure. Experiments show that this method can be applied to the rotor fault detection of asynchronous motor, which can accurately detect the fault characteristic components of the asynchronous motor in the stator current when the rotor bar breaks.

With the large-scale, complex and nonlinear of modern equipment, it is difficult to establish an accurate mathematical model of the system. Therefore, the application of fault diagnosis method based on the traditional mathematical analysis model has great limitations. For the fault diagnosis of complex industrial process optimization control system, a method based on data-driven artificial intelligence optimization model is adopted to compensate for this defect.

A fault diagnosis method based on data-driven artificial intelligence optimization model includes signal processing (spectral analysis and wavelet transform), statistical analysis (single variable method and multivariate method), information fusion method and intelligent machine learning method (such as neural networks, support vector machines and rough sets).

Inreference[4], adopting signal processing, wavelet transform is used to analyze signals in different scales to extract the features of signals at different scales for fault diagnosis.

In reference[5], using principal component statistical analysis (PCA), a matrix of high-dimensional historical data is proposed. A number of orthogonal vectors are determined after a series of matrix operations.

In reference[6], a new method of sensor fault detection based on association rules is proposed. According to the characteristics of the multi - sensor system and the different operating conditions of the thermal power plant, the algorithm of the association rules in the data mining is improved by making full use of the large amount of redundant information in the operation of the thermal power plant. Through the evaluation and simulation of the generated association rules, it shows that the association rules can be used for sensor fault detection and improve the calculation speed.

Each of the above methods has certain limitations. The signal processing method relies heavily on the data difference of signal acquisition, and has certain requirements for data processing method. The precondition of statistical analysis is that historical data is needed as training data, and it is necessary to classify all kinds of fault data, which can't be satisfied for complex systems of multi levels, scales and ranges. The information fusion method utilizes the complementary and redundant relations among different sensor information; after analyzing and integrating multiple information sources, the accurate location of the fault source is calculated, and it is difficult to give accurate fusion parameters for the strong coupling system.

To solve this problem, the artificial intelligence optimization model diagnosis method based on neural network is adopted, which takes implicit expression and represents some knowledge of a problem in the same network, of high universality, easy to achieve the overall knowledge acquisition and associative reasoning.

In reference[7], aiming at the problem that the neural network is easy to fall into local minimum, an improved BP algorithm with momentum and chaotic mapping is introduced, the modeling idea and algorithm realization of which are discussed, and a chaotic neural network model for ball mill fault diagnosis is established.

Reference[8], adopting a method of combination between multilayer perceptron neural networks and system identification, using Matlab system identification toolbox and neural network toolbox, proposed a method for predicting the concentrate grade with hybrid model structure composed of linear model and nonlinear compensation model. A prediction model of ore concentrate grade has been established.

The traditional neural network model, has the advantages of high classification accuracy, strong robustness and fault tolerance, and can deal with complex nonlinearproblem and fully approximate complex nonlinear relations, and can find dependencies among different inputs. The disadvantage is that a large number of parameters are required, the learning process cannot be observed and the learning time is long.

Reference[9] aims at problems such as small fault samples, strong nonlinearity, multi-fault processing and deficiency of traditional intelligent diagnosis methods, an intelligent fault diagnosis method is proposed based on decision tree (DT) and relevance vector machine (RVM).

By constructing a binary decision tree, two categories are classified using multipleRVM, so as to realize the multi-class classification of RVM.

Fault-tree-based methods have the advantage of processing both categorical and numerical data simultaneously, which is easy to handle interactionsamong variables, suitable for small scale data. The main shortcoming of this method is not good at predicting numerical results.

In the recurrent neural network, it can be used to deal with time series data from time to time, and the application of back propagation training. But gradient or disappearance of the gradient[10,11], in practice, encounters many difficulties. The reference[12] proposed Long Short-Term Memory (LSTM) with a special hidden unit, achieving long and short-term memory at the same time. LSTM has also made significant improvements after achieved significant results, such as LSTM[13], Gate Recursion Unit (GRU)[14] Depth Gate RNN (short for recurrent neural networks)[15], RNN with a connection that can track long-range effects[16-18] and so on. LSTM has a better effect about time series data, but the structure is a little more complicated than GRU, GRU also has a better effect than LSTM. Therefore this article will propose a fault diagnosis method for ball mill based on GRU.

The chapter two describes the analysis and preprocess of the data, including determining the input vector dimension, filling missing data, cleaning noise data, defining fault characterization methods and analyzing data association, then establishing the mapping relation between feature vectors and tag data in the sample data. The data set collects data at the acquisition nodes every other hour, that is to say, it has a very strong temporal correlation, the whole data set can also be regarded as a high-dimensional time series, and the problem of fault diagnosis is transformed into a prediction of time series.

The chapterthree describes formulas, structures and parameters of traditional BP neural networks, autoencoder softmax classifier, LSTM and GRU improved on LSTM.

The chapter four compares the results of different algorithms. The traditional neural networks do not perform well for the diagnosis of time series. In recent years, RNN has a high status in the field of time series prediction. In this paper, a GRU neural network is used as a time RNN to diagnose the data set of this paper, and achieve satisfied results.

2 Data Analysis and Preprocessing

Based on the post record data set of Vseries grinding system, the data sets are analyzed and studied, so as to realize the classification and prediction of the faults occurring in the operation of the grinding system. The data of each sampling point (temperature, pressure, sound, vibration, etc.) is recorded every hour, total 1000 hours, 30 kinds of measurement index.

Through observation, some of the measurement indicators have not changed, thus, before the data analysis it can be screened and eliminated manually, reducing the vertical dimension of a dataset from 30 to 24 dimensions. In this paper, the missing record is filled with the mean value of 5 hours data before or after this correspond point.

For the faults occurred in the process of grinding production, it may be caused by the equipment itself, such as the abnormal temperature of the bearing, on the other hand, it may be caused by technological aspects, such as fluctuations in the amount of material to be supplied. In this paper, the faults in the actual data set are analyzed, and the final purpose is to diagnose the faults, so take the faults as labels. For the characterization of these data, this paper uses the " one-hot vector" representation, that is, for a few types of faults, the corresponding position is 1, and the remaining position is 0. This paper is divided into 8 types of faults, as shown in Table 1.

Table 1 Causes of Faults and Methods of Representation

Fault Phenomenon	Failure Cause	Fault Characterization
Normal	No	100 000 00
Low Feed Pressure	Feed End Ore Box Leakage	010 000 00
High Feed Pressure	Cyclone Perforation	001 000 00
The Main Motor Current Is Low and the Clutch Cylinder Pressure Is Low	Cylinder Failure	000 100 00
The Temperature of Main Motor Bearing Is High, and temperature of oil is high	The Cooling Water Valve of Oil Station Is Damaged	000 010 00
The Temperature of Feed-End Trunnion Bearingis High and Oil Pressure Is Low	Low Oil Pressure, High Bearing Temperature	000 001 00
The Temperature of Discharge-End Trunnion Bearing Is High and Oil Flow Rate Is Low	Low Flow, Bearing Temperature Rise	000 000 10
The Main Motor Temperature Is High	High Oil Temperature, Low Oil Pressure	000 000 01

As it can be seen from the table, there are eight kinds of faults, such as low oil pressure, low flow rate and cooling water failure, and some of these indexes are relatively strong and difficult to identify. Through the analysis of the data, it found that the mapping between feature vector and target: In the 1000 sets of tags corresponding to feature vectors, there are 865 categories for no fault, the 135 for the fault, which is the serious imbalance classification. Here it has chosen two dimensional features for analysis, as shown in Figure 1.

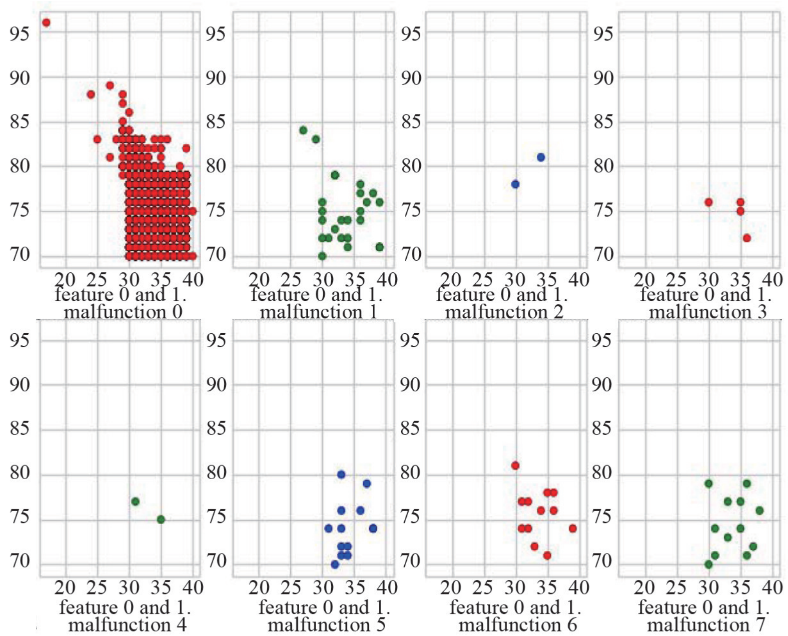


Fig. 1 Classification of 0,1 Dimension

In Figure.1, the horizontal ordinates represent zeroth dimensional features (feature contents) and first dimensional features (feature contents) respectively, and each graph represents the distribution characteristics of each fault (or no fault).

Asit can be seen from Figure 1, the first kind of fault (fault-free) accounts for a large proportion of the 8 types of failures, while other failures account for a much smaller proportion, which is a serious imbalanced classification problem. At the same time, the fault is dispersive, which is unfavorable for classification.

3 Proposed GRU Algorithm

3.1 Artificial Neural Networks

First of all, artificial neural network has strong learning ability for knowledge. It can improve the speed of data processing because it can process data in parallel. So it applies the neural network technology to the fault diagnosis in the grinding process to obtain the ideal diagnosis effect.

At present, BP networks are widely used in many feedforward networks, which is famous for its fast learning speed, good approximation ability and strong classification ability. In this paper, BP neural network is used to diagnose the faults in grinding, and the function structure of the diagnosis system is shown in Figure 2.

Traditional BP neural networks are easy to fall into local minima, so the " batch processing" learning method is used to improve this phenomenon. In the process of neural network training, this method is not affected by the order of the learning samples, and accelerates the convergence speed by self-adjusting learning rate. The model of BP neural network constructed in this paper is three layer neural networks, namely input layer, hidden layer and output layer. Since there is no large correlation among the data, and there is a big difference among the magnitudes, the data is normalized before the sample data is input into the network.

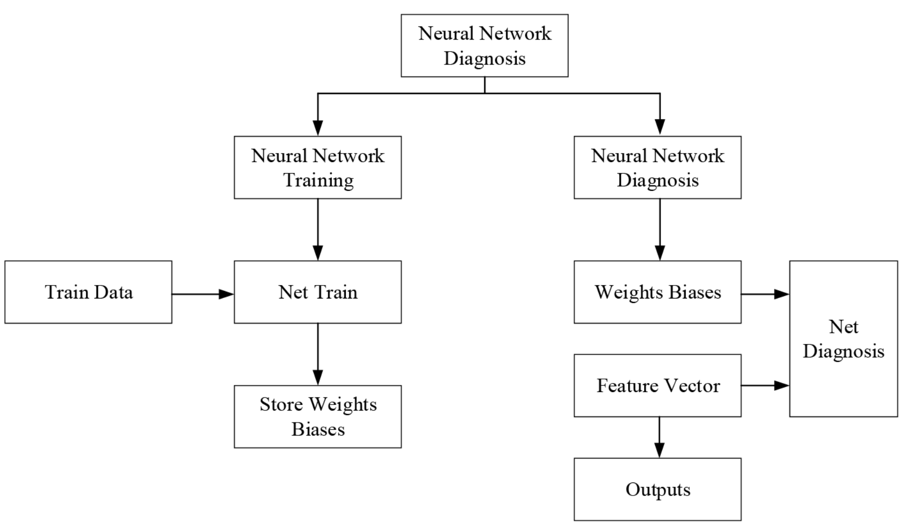


Fig. 2 Neural Network Fault Diagnosis Function Diagram

3.2 Autoencoder Softmax Classifier

In this paper, the input dimension of training samples is 24 dimensions. For neural networks, the higher the dimension, the slower the convergence. Further feature extraction can be performed on training sample inputs, and the Autoencoder neural network is a good choice.

In the experiment, two networks are established, one is Autoencoder self-encoded neural network, which can be regarded as a three layer neural network, and the number of inputs is the same as the number of outputs, the number of neurons in the hidden layer is 10, less than the number of inputs, and the activation function of the neurons in each layer is chosen as the Sigmoid function. The purpose is similar to that of PCA, which is to extract the input features further. Another is that the front end is the same as the first two layers of the Autoencoder network, and the output layer is connected with a Softmax classifier to form a Autoencoder Softmax classifier network. The output of the hidden layer of the network is exactly like the characteristic value after pooling, but it contains almost all the information of the input vector, and the output layer is 8 neurons, which is exactly the 8 fault categories of the output. Unlike traditional neural networks, the output layer becomes a Softmax classifier, the detailed structure of the network is shown in Figure 3.

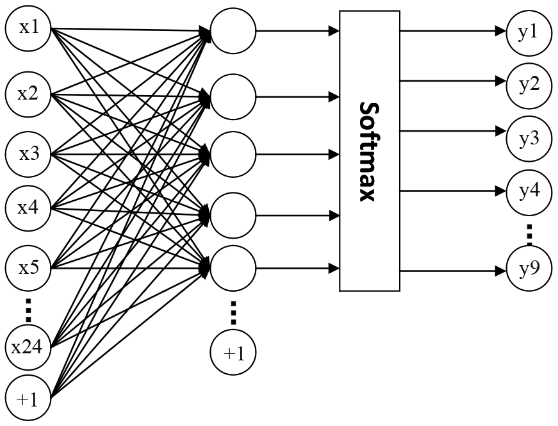


Fig. 3 Autoencoder Softmax Learning Network Architecture

Network training is divided into two parts, first of all, training Autoencoder network, input all the training samples of the 1000 * 24 dimension, the gradient descent algorithm is trained according to the BP network to minimize the mean variance of the output as far as possible. After the Autoencoder network is trained, the final weights and thresholds are preserved.

The next step is to train the Autoencoder Softmax classifier network, the weights and thresholds preserved by the previously trained network are used to initialize the weights and thresholds of the newly constructed neural network, re-training, at this time the training algorithm is L-BFGS, the objective function is to make the cross entropy minimum. When experiments are conducted on Autoencoder Softmax deep learning network, 900 sets of data are selected as training samples, and the remaining 100 sets of data are used as test set for cross validation. To verify the performance of a self-encoded network, it chooses eight dimensional features to observe the reduction effect, as shown in Figure 4.

Seen From Figure 4., most of the feature points after encoding and decoding are still close to the feature points before encoding, after many steps training, the self-encoded network can reproduce the input better, that is to say, the features extracted from the hidden layer contain most information of the input features, namely more obvious feature vectors of the hidden layer output can be used instead of the high dimensional input vectors of the original network, to reduce the dimensionality of the data.

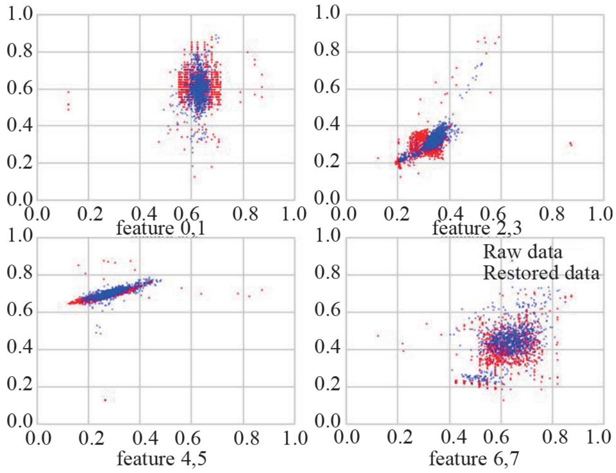


Fig. 4 Restoring Effect Graph of Autoencoder

After completing training, the output characteristics of the hidden layer are obtained after encoding, as shown in Figure 5:

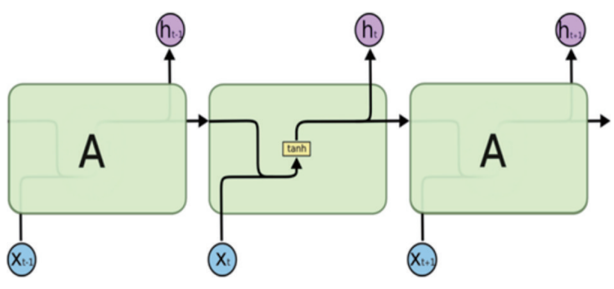


Fig. 5 Output Features of Autoencoder Network

It can be seen from Figure 5 that the number of characters after encoding is greatly reduced compared with the number of characters before encoding, which is the result of self coding for lossless compression of high dimensional redundancy, so that the effective features can be reproduced. Not only the number of features is reduced, but also the distribution of the fault points is more concentrated, so that the classifier can be used to achieve accurate classification.

3.3 Recurrent Neural Network (RNN)

Described by the previous data set, the data set used in this paper is a dataset of 1000 sets of 24 dimensional features, and the 1000 sets of data are collected at every point of the hour on the point of collection. Therefore, the 1000 sets of data can be considered as a time series with a time interval of 1 hour. There is a certain relation between each set of data, that is, the data collected in the last hour and the data collected after one hour will affect the status of the current moment.

When each set of data is entered as a sample, the network analyzes and predicts a state value. This is different from the traditional neural network. The traditional neural network can predict and analyze the current samples by the prior knowledge obtained by training, and the prior information does not include the relation information between front and back samples, so that there is no good prediction result for the data set in this paper. In this paper, there is a certain time relationamong data sets, so the classification of data samples can be transformed into a prediction problem of time series.

RNN is called recurrent neural networks, that is, in a sequence the previous output affects the current output. The details are as follows: the information ahead will be remembered by the current network and affect the current output, that is, the nodes between the hidden layers are no longer connected, moreover, the input of the hidden layer includes the output of the input layer, also consider the output of the hidden layer at the previous moment. Theconcrete structure is shown in Figure 6.

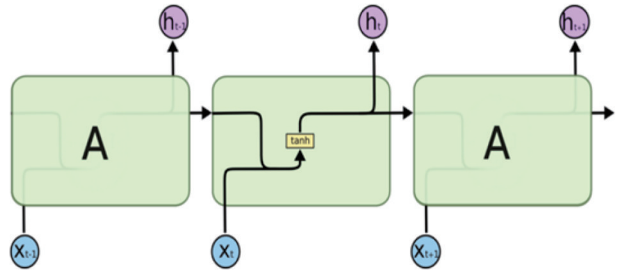


Fig. 6 Structure of RNN

Artificial neural network algorithm for long and short term memory is a special RNN model, in traditional RNN, the BPTT algorithm is usually used to train data, if the time is too long, the residualthat are needed to return will decline exponentially, and the network weight update will be slower, the function of long-term memory can't be realized for RNN. To solve this problem, a memory cell (cell state) is proposed to store memory and to control cell state changes through the " gate " structure. There are three gates in the LSTM, namely, the memory gate, the input gate and the output gate.

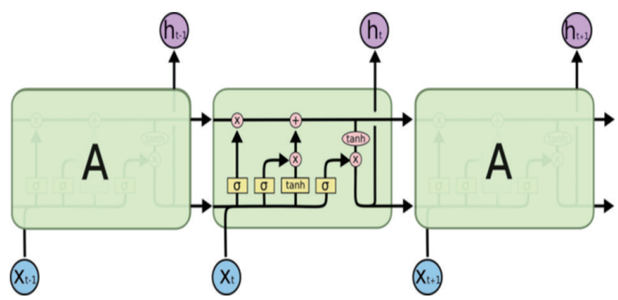


Fig. 7 Structure of LSTM

Cell state (also known as an information carousel): the conveyor itself is unable to decide which part of the information to remember, but determined by the control gate. Among them, the "forget gate" determines the information removed from the information carousel. The output of the last state of the forgetting gate and the input of the current state are entered into a sigmoid function to produce a value between 0 and 1, which is then multiplied by the information carousel to determine the degree of retention of the information. 0 means " giving up completely" , and 1 means " completely reserved" " . Formula for

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

Input gate: determining what new information is to be saved in the carousel. Through the " forget gate" to determine whether to store new information, at the same time, by combining the output of the previous state and the current status input, a tanh layer calculates a candidate new information which is added to the information carousel. Then, the newly obtained values are multiplied with the candidate new information, and the update information that added to the information carousel is obtained. That is

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \widetilde{C}_t &= \tanh(W_C[h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \widetilde{C}_t \end{aligned} \quad (2)$$

Output gate: determining what information is output from the information carousel. It uses a Sigmoid function to get a value between 0 and 1, to measure the extent to which it outputs the information in the information carousel. The information of the information carousel is first activated by a tanh layer (nonlinear transformation), then multiply to get the output information of LSTM. That is

$$\begin{aligned} O_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \quad (3)$$

The above three gates form the LSTM algorithm.

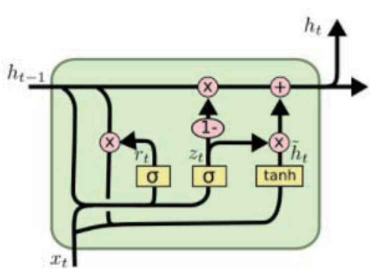


Fig. 8 Structure of GRU

GRU is an improvement of LSTM, which combines the "forget gate" with the input gate into an " update gate" and merges the cell state with the implied state. The GRU model is simpler and easier to train than the standard LSTM model, and is becoming more popular. The GRU expression is as follows:

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \widetilde{h}_t &= \tanh(W[r_t \cdot h_{t-1}, x_t] + b_C) \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \end{aligned} \quad (4)$$

This paper builds the traditional BP neural network by tensor flow deep learning framework with three layers. There are 24 neurons in the input layer, 200 in the hidden layer, 100 in the one next to the hidden layer and 2 in the final output layer. Xvair is used to initialize the neural network, mini-batch gradient descent algorithm is used when training the network. L2 regularization is added to the loss function in order to increase the robustness of the model. the learning rate is first adjusted from low to high, until the oscillation phenomenon occurs in loss. In this way, the number of nodes in the middle layer is adjusted, the learning rate is reduced until loss no longer decreases, and the final parameters are determined.

Tensor flow is also used to achieve the autoencoder softmax network. The autoencoder model uses the classic three layers neural network, that is input layer, hidden layer and output layer, the dimensions of input layer and output layer are consistent. The loss function is the least square loss function. The evaluation index of model is RMSE. The method in adjusting parameters is similar to BP, and the training is to minimize the RMSE, that is to say, the output and input are very close.

Next, the output layer ofautoencoder is removed, followed by softmax structure, which is to realize the effective classification of faults. In other words, the features of the input samples disposed by dimension reduction of autoencoder are input into softmax structure for training, and the parameters are adjusted on the training set to obtain the optimal network parameters.

As for LSTM and GRU, tensor flow deep learning framework is used as well. The input dimension is 24 for data set, step is 1 because of the collection per hour, batch-size is 30, so the input dimension is [batch-size, step, rnn-size], where rnn-size is 24, which is input node number, then followed by a series of hidden layer and finally the output layer. The initial learning rate is 0.001, and the training method is the same as above. Finally save the network parameters and apply to the test data set. The experimental results are analyzed in the following chapter.

4 Experimental Results and Analysis

4.1 Traditional BP Networks

In this paper, the 1000 * 24 dimensional data sets are divided into 900 * 24 dimensional data as training samples, and the cross validation of the remaining 100 * 24 dimension data as the test set can prevent the network overfitting. Batch processing is performed when the training sample set is input to the network. Each time entering 100 * 24 dimensional data, it can save the weights and thresholds after training the network, the test set is then fed into a trained network to verify network reliability and diagnostic accuracy.

After inputting training sample set, the output of the network is divided into two types: one is fault, and the other is opposite; for faulty types, further analysis is carried out on which type of fault.

In the data analysis of the third part, the classification of sample data sets has serious classification imbalance problem, and the traditional classifier performance evaluation index is notapplicable in this problem. In the non-equilibrium classification, confusion matrix is used to define new evaluation index. In this paper, three kinds of indexes are defined, the error rate, the prediction accuracy of the fault and the recall rate.

The traditional BP neural network is verified, the structure of which is 20 * 200 * 8 and activation function is RELU, and the experimental results are as follows:

Through the experimental results, it can be calculated that the BP neural network without any feature transformation has a accuracy rate of 46.67%, recall rate of 46.67% and error rate of 38.71%.

Table 2 Testing Results of BP

Reality \ Forecast	Faulty	Fault-Free
Faulty	21	24
Fault-Free	24	55

It can be seen that if the characteristic item is input to the neural network without any characteristic changes, the network performs poorly in fault identification accuracy, and the error rate is high. It also shows that the traditional BP neural network is not suitable for imbalanced classification problem with high dimension input.

4.2 Autoencoder Softmax

The AutoencoderSoftmax neural network is experimentally verified in this part. The network structure is an Autoencoder network. The network is dimensioned by the activation function of RELU and the number of neurons is 24 * 100 * 8 * 100 * 24. The parameters of the first three layers are saved and it gets a classifier with a structure of 24 * 100 * 8 * 8 with the Softmax layer. The result is as follows:

Table 3 Testing Results of Autoencoder Softmax

Reality \ Forecast	Faulty	Fault-Free
Faulty	37	8
Fault-Free	8	55

With the data in Table3, as for the Autoencoder Softmax depth learning network, the accuracy rate is 82.22%, the recall rate is 82.22%, and the error rate is reduced to 14.81%.

4.3 LSTM

It is same as the previous experiment, three indicators are also used to evaluate the performance, the experimental results are as follows:

Table 4 Testing Results of LSTM

Reality \ Forecast	Faulty	Fault-Free
Faulty	42	3
Fault-Free	3	55

With the data in Table4, as for the LSTM, the accuracy rate is 93.33%, the recall rate is 93.33% and the error rate is reduced to 5.83%.

4.4 GRU Experimental Results and Comparative Analysis

It is same as the previous experiment, three indicators are also used to evaluate the performance, the experimental result is as follows:

Table 5 Testing Results of GRU

Reality \ Forecast	Faulty	Fault-Free
Faulty	43	2
Fault-Free	3	55

The data in Table 5 shows that the accuracy rate is 95.56%, the recall rate is 93.48%, and the error rate is reduced to 4.85%. It is easy to see that the GRU network accuracy rate has exceeded 90% and-best of all, the rate of error for fault prediction has dropped to 4.85%, which stabilizes the operation for whole system.

The abovefor algorithms are compared and analyzed, the results are shown in Figure 9.

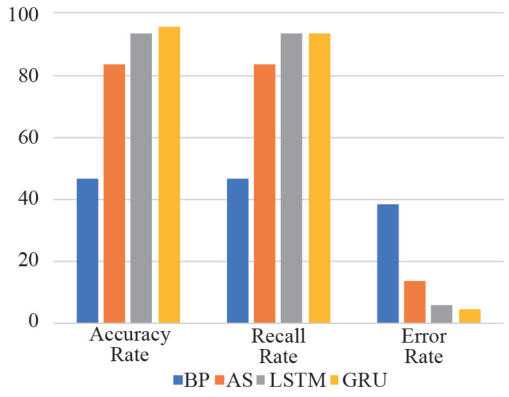


Fig. 9 Identification Effects Comparison of Four Methods

As it can be seen from Figure 9, LSTM and GRU network are obviously superior to the previous two networks in diagnosing failures, they greatly reduce the error rate. In addition, LSTM and GRU networks do not perform extract further feature before inputting training samples, compared with the network after the feature extraction by the Autoencoder networks, which achieves a great degree of retention of the input sample information and avoids the loss of useful information. Moreover, the structure of GRU is simpler and the training is easier to converge to a better solution, so the effect is better than standard LSTM. It can be seen that the recursive network has a good classification effect on the balanced classification problem.

5 Conclusion

The Autoencoder Softmax deep learning network proposed in this paper, can well extract the effective features from the complex feature vector data set, and the dimensionality of the input data is losslessly compressed, which greatly reduces the computational complexity of the network. But in the actual training process, it is hoped that the training sample can retain the useful information of the data as much as possible, so that the training network will be more robust and fault-tolerant. In dealing with high-dimensional and time correlated data samples, the convergence speed and identification effect of GRU networks are better than those of LSTM and Autoencoder Softmax deep learning networks. Moreover, the GRU network also obtains satisfactory results in the imbalanced classification problem. In the fault diagnosis of this paper, the prediction accuracy for the fault of GRU network is 95.56%, The recall rate is 93.48%, and the error rate is 4.85%. This method will provide a theoretical basis for the future non-traditional classification problems, and will have far-reaching significance in depth learning and fault identification.

References

[1] Liu J. (2014) *Research on Outlier Detection Method of Complex Data Type* (Doctoral dissertation, South China University of Technology).

[2] Cai, Y. P., Ai-Hua, L. I., Shi, L. S., Bai, X. F., & Shen, J. W. (2011). *Roller bearing fault detection using improved envelope spectrum analysis based on emd and spectrum kurtosis.* Journal of Vibration & Shock, 30(2), 168-172.

[3] Hou, X. G., Wu, Z. G., & Xia, L. (2003). *Method for detecting rotor faults in asynchronous motors based on the square of the park´s vector modulus.* Proceedings of the Csee, 23(9), 137-140.

[4] Lopez, J. E., Tenney, R. R., & Deckert, J. C. (1994). *Fault detection and identification using real-time wavelet feature extraction.* IEEE-Sp International Symposium on Time-Frequency and Time-Scale Analysis (Vol.23, pp.217-220). IEEE.

[5] Jolliffe, I. T. (2011). *Principal component analysis.* Journal of Marketing Research, 87(100), 513.

[6] Bao, W., Yu, D. R., Wang, W., & Xu, Z. Q. (2003). *Sensor fault detection in thermal power plants based on association rule.* Proceedings of the Csee.

[7] Scattolini, R. (2009). *Architectures for distributed and hierarchical model predictive control – a review.* Journal of Process Control, 19(5), 723-731.

[8] Ranaee, V., Ebrahimzadeh, A., & Ghaderi, R. (2010). *Application of the pso – svm model for recognition of control chart patterns.* Isa Transactions, 49(4), 577-586.

[9] Garrido, J., Vázquez, F., & Morilla, F. (2011). *An extended approach of inverted decoupling.* Journal of Process Control, 21(1), 55-68.

[10] Hochreiter, S. (1991). *Untersuchungen zu dynamischen neuronalen Netzen*. Master´s Thesis, Institut Fur Informatik, Technische Universitat, Munchen.

[11] Bengio, Y., Simard, P., & Frasconi, P. (2002). *Learning long-term dependencies with gradient descent is difficult.* IEEE Transactions on Neural Networks, 5(2), 157-166.

[12] Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory.* Neural Computation, 9(8), 1735-1780.

[13] Deng, D. (2011). *Research on coal mine electromechanical equipment closed-loop management system based on IOT and information technology.* International Conference on Artificial Intelligence, Management Science and Electronic Commerce (pp. 5101-5104). IEEE.

[14] Cho, K., Merrienboer, B. V., Gulcehre, C., Bahdanau, D., Bougares, F., & Schwenk, H., et al. (2014). *Learning phrase representations using rnn encoder-decoder for statistical machine translation.* Computer Science.

[15] Yao, K., Cohn, T., Vylomova, K., Duh, K., & Dyer, C. (2015). *Depth-gated lstm.* Computer Science.

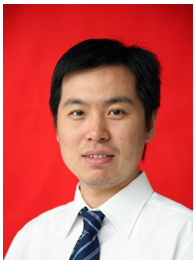
[16] Koutník, J., Greff, K., Gomez, F., & Schmidhuber, J. (2014). *A clockwork rnn.* Computer Science, 1863-1871.

[17] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). *Lstm: a search space odyssey.* IEEE Transactions on Neural Networks & Learning Systems, 28(10), 2222-2232.

[18] Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). *An empirical exploration of recurrent network architectures.* International Conference on International Conference on Machine Learning (pp.2342-2350). JMLR.org.

Author Biographies

QU Xingyu received his PhD degree in 2018 from Shenyang Institute of Automation, Chinese Academy of Sciences. He is Senior engineer and the director of the design institute in Northern Heavy Industries Group Co. Ltd. His main research interests include automatic control and network control system of large industrial equipment.



ZENG Peng received PhD degree from Chinese Academy of Sciences in 2005. Now he is a PhD in CUST.His main research interest is spacecraft ground calibration and measurement.

LI Junpeng received bachelor's degree from Northeastern University in 2014. Now he is a assistant engineer in Northern Heavy Industries Group Co. Ltd.. His main research interests include automatic control and network control system of large industrial equipment.

E-mail:15940425217@163.com