2019 IEEE International Conference on Energy Internet (ICEI)

A Fault Prediction Of Equipment Based On CNN-LSTM Network*

Ling Zheng
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
zhengling@ncepu.edu.cn

Wenhao Xue
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
1172227209@ncepu.edu.cn

Fei Chen
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
chenfei@ncepu.edu.cn

Pengtian Guo
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
1172227005@ncepu.edu.cn

Jiaqi Chen
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
1172227198@ncepu.edu.cn

Biying Chen
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
1172227001@ncepu.edu.cn

Hongbiao Gao
School of Control and Computer Engineering
North China Electric Power University
Beijing, China
gaoh@ncepu.edu.cn

***Abstract*—Unit equipment is the key to industrial production, and predicting unit failures is the focus of improving equipment productivity. In order to improve the accuracy and reliability of predicting, and consider the use of multiple related influencing factors for prediction. This paper presents a fault prediction method based on CNN-LSTM. Firstly, the data of multiple variables that affects the predicted value over a period of time are formed into a large matrix. Then, the Long short- term memory network is trained by the feature information extracted from the convolutional neural network. This can predict device data at future time points and build models that use large amounts of data for predicting. Finally, the rationality and effectiveness of the proposed method are verified by the comparison between the evaluation example and the LSTM network.**

Keywords-fault prediction; multiple factors; convolutional neural network; Long short- term memory network

I. INTRODUCTION

As unit equipment are the key to the power, chemical, petroleum, aviation and other industries, Its operating status will directly affect its output value. Large failures of them will directly affect their productivity, and more serious ones will bring huge losses to economic assets and even threaten the lives of workers. Therefore, how to maintain the normal operation of the unit equipment has become the heated interest of scholars' research.

Aiming at the problem of fault prediction, multiple ideas have been proposed by experts. The document [1-5] uses envelope spectrum analysis, cepstrum analysis, Gabor transform, wavelet transform, Intrinsic Time-scale Decomposition method and other traditional methods to research the fault prediction. However, the traditional research method is based on the analysis of the collected signals, and researchers who analyze the signals must have a good understanding of the structure and composition of the unit. The system established in this method can only be applied to a single device, which has poor versatility and high limitations.

With the advent of artificial neural networks, its excellent abilities of parallel processing and handling large amounts of data have made it gradually replaced traditional methods. Considering that the device usually judges whether the fault is caused by comparing the output data with the threshold value, the fault prediction problem for the device can be converted into a prediction problem for the device output data, and the neural network has excellent processing ability for time series prediction. Literature [6] used BP neural network to predict the trend of DNA coding which set off a research trend of

*Funding Project : Beijing Municipal Commission of Science and Technology, Technology Collaborative Innovation Project in Energy Field （Z181100005118016）

978-1-7281-1493-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICEI.2019.00101

neural network. After that, the literatures [7-10] proposed a recurrent neural network prediction model based on neural network multi-step time series prediction method, based on dynamic wavelet neural network model; Elman neural network is used to predict multi-parameters. For complex equipment and data, these methods have greatly improved the prediction effect. However, the training of these networks is prone to over-fitting problems and does not have a good generalization effect.

With the continuous development of machine learning, the concept of deep learning has been proposed. Deep learning is to approximate complex nonlinear functions with small errors through multi-layer information processing and feature extraction. Convolutional neural networks and recurrent neural networks, which have been proposed in the literatures [11-12], have practical significance in the field of deep learning. Recurrent neural networks have higher accuracy in time series prediction. But deep training can cause the gradient of the neural network to disappear. The long short-time memory network proposed in [13] adds three 'gate' structures on the basis of the recurrent neural network, and has a new improvement in the prediction of time series, with relatively high accuracy.

The output prediction of power plant equipment often has a large number of multi-dimensional influence factors. For a large amount of multi-dimensional data, the long short-term memory network are less efficient, and the accuracy is often not high. Although the convolutional neural network can usually process multi-dimensional data, the timing of the data can not be well reflected. The introduction of traditional particle swarms, attention mechanisms and other algorithms can process multidimensional data, reducing data redundancy and data loss, but these methods often ignore the relationship between input data. In order to find an effective solution, this paper proposes a model based on CNN-LSTM applied to the fault prediction stage of the equipment. Among them, the convolutional neural network part is good at data processing. It can reduce the large amount of data required for experiments without losing the connection between data. The extracted feature vector is put into the long short-term memory network, which is good at predicting time series data. This model not only ensures the characteristics of the input data and its relationship, but also ensures the timing of the model. Finally, the model prediction value, LSTM network prediction value and practical value are compared. The rationality and effectiveness of the method are verified by an example. It is also proved that the method has higher improvement effect than LSTM network prediction.

II. ALGORITHM MODEL ANALYSIS

A. Convolutional Neural Networks

Convolutional neural network (CNN)[14-15] is a feedforward neural network. As shown in Figure 1, it uses local connection and shared weights to reduce the dimensionality of data information to generate eigenvalues that are easy to train. It consists of a data input layer, a convolution calculation layer, a pooling layer, and a fully connected layer.

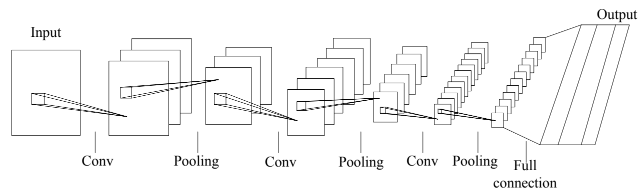

Figure 1. Convolutional neural network model

First, the large matrix data is put to the convolutional layer. The convolutional layer uses a filter matrix for feature extraction. As shown in Fig. 2, this filter matrix is also called a convolution kernel. The data in the matrix is extracted by the step size and the convolution kernel for the inner product extraction feature. The filtering operations of different convolution kernels can extract different features. Each convolutional layer can be set with a different number of convolution kernels, and the number of convolution kernels per layer determines the number of characteristic matrices output by the convolutional layer.

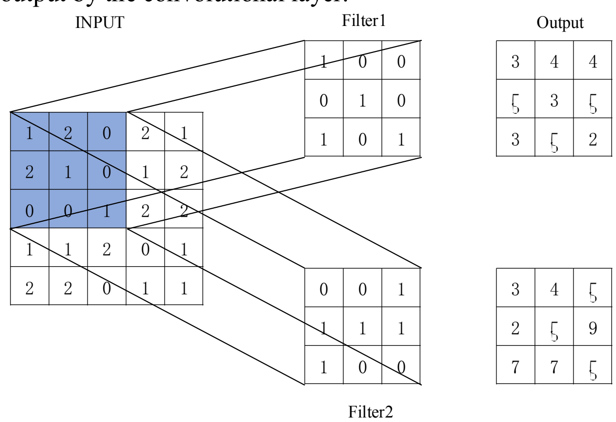

Figure 2. Convolutional layer calculation process

The feature matrix processed by the convolution layer enters the pooling layer. The pooling layer is mainly used for feature dimensionality reduction, compressing the number of data and parameters, reducing over-fitting, and improving the fault tolerance of the model. The pooling method mainly includes the max pooling method and the average pooling method. Both methods need to first define a filter matrix size, and divide the feature matrix into multiple matrices of the same size according to the filter matrix size. The max pooling method needs to extract the maximum output in the divided matrix as the element of the pooled feature matrix, while the average pooling rule is to input the average of all the values in each matrix divided as the element input of the feature matrix. In this paper, the max pooling method is used for pooling. The pooling process is shown in Figure 3.

After multiple operations of the convolution layer and the pooling layer, the obtained data feature matrix is input into the fully connected layer, and the function of the fully connected layer is to map the generated plurality of feature matrices into one dimension including all the feature information of the matrix. In the array, it is convenient for the next step of processing and application.

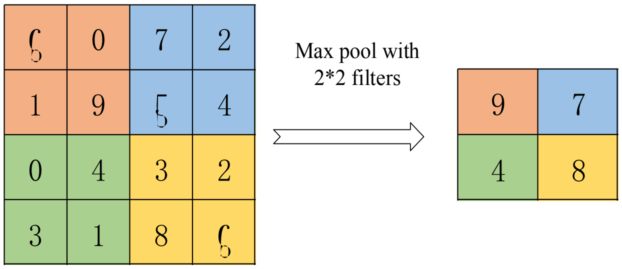

Figure 3. Pooling layer calculation process

B. Long Short-Term Memory Network

Recurrent Neural Networks (RNN) is the most popular neural network for time series prediction. Compared with other feedforward neural networks, RNN has a feedback connection, which allows the state to circulate itself, so it is possible to establish time series data. The dependencies, which make the RNN network popular in the processing of time series data. However, the training of long-term deep neural networks leads to the disappearance and explosion of the gradient of the circulating neural network. The emergence of the Long short-term memory network (LSTM) proposed by Grves improves this problem [16].

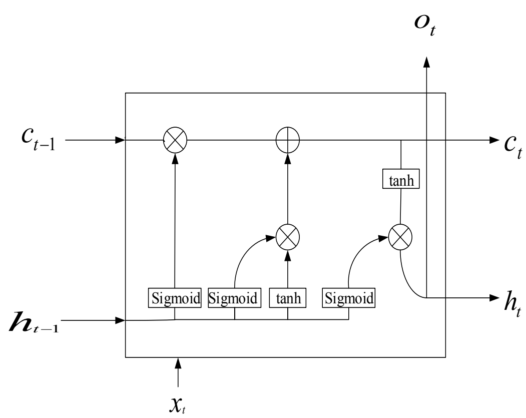

Figure 4. Basic unit of LSTM network

As shown in Figure 4, LSTM has a stronger processing power for long time sequences than RNN mainly because it adds memory cells to hidden layer nodes. The memory cells contain three "gates" Structure: input gate, forgetting gate and output gate. The forgotten gate is mainly used to "forget" the information that was not used before. It is judged by the gate value f_t calculated by the output h_{t-1} of the hidden layer at the previous moment and the current input x_t. When the gate value is close to 0, The state at the last moment is forgotten. When the gate value is close to 1, the state at the last moment is retained. The input gate is mainly to supply a new memory by inputting the gate value i_t after the previous state is forgotten. Calculate which states are added to the previous moment by h_{t-1} and x_t to become the new state, Where the sigmoid function determines which values are updated, and the tanh function creates a new candidate value vector, multiplies the values calculated by the two functions, and then linearly superimpose the state values output by the forgetting gate to obtain a new state c_t. The output gate value o_t is also affected by h_{t-1} and x_t. It calculates the output vector through the sigmoid function. Finally, multiply the tanh function calculation by c_t and multiply the output vector to get h_t to be output by this hidden layer. The calculation formula for the specific LSTM cell is as follows:

$$f_t = sigmoid(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{1}$$

$$i_t = sigmoid(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{2}$$

$$g_t = \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \tag{3}$$

$$o_t = sigmoid(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{4}$$

$$c_t = g_t \cdot i_t + c_{t-1} \cdot f_t \tag{5}$$

$$h_t = \tanh(c_t) \cdot o_t \tag{6}$$

Among them, f_t i_t o_t g_t h_t are the forgetting gate, the input gate, the candidate vector, the output gate, and the hidden layer output. W_{fx} W_{fh} W_{ix} W_{ih} W_{gx} W_{gh} W_{ox} W_{oh} are the matrix weight of each gate multiplied by the current time input x_t and the previous time hidden layer output h_t respectively. b_f b_i b_g b_o are the corresponding gates respectively Offset. The memory function of the three gate-controlled memory cells realizes its special memory function and avoids the gradient explosion problem of deep-level circulating neural networks.

C. CNN-LSTM Hybrid Model

In the production process, the unit equipment often has an expert experience threshold, and the predicted value is compared with the equipment threshold to determine whether the equipment is faulty. Therefore, the prediction problem of the fault can be converted into a prediction of the output value of the equipment. The block diagram of the CNN-LSTM model designed in this paper is shown in Figure 5. Since the equipment data information of the power plant is usually related to various influencing factors, such as temperature, humidity, multiple input values of the equipment, input values and output values of related equipment. At the same time, the values of the first few moments of the data will also interfere with the current state of the device. If a large amount of data is directly input into the LSTM, the LSTM network directly processes a large amount of data, which increases the complexity of the model and reduces the efficiency of the model. And it will have a greater impact on the accuracy of the prediction. In order to solve this problem, this paper introduces a convolutional neural network for data processing. Firstly making a continuous large amount of data into a data matrix. Then through the CNN network to extract the matrix. Finally the feature vector is input into the LSTM network for prediction, which increases the weight of the influence of various factors on the prediction result. This makes the prediction model has higher accuracy than the traditional LSTM network, and greatly reduces the complexity of the model.

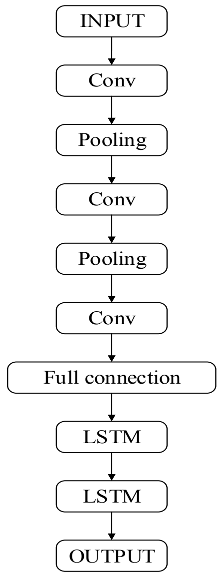

Figure 5. Hybrid model of CNN-LSTM network

III. EXPERIMENTAL ANALYSIS

A. Data Sources

To test and verify the reliability of the model, the study selects the data from the steam temperature model of the superheated primary desuperheating system of Unit 2 of a power plant. Select first-stage desuperheating water flow of the superheater and 19 factors to predict the steam temperature before the secondary desuperheating of the superheater. Those factors include the first-stage desuperheating water flow of the superheater, the average value of the wall temperature of 12 boiler drums, the smoke temperature of the furnace side AB, the fuel quantity, the main steam flow, the flue gas flow, the total air volume, and the main steam pressure. Finally the reliability of the algorithm is verified by comparing the predicted value with the practical value. For coupling the characteristic information that affects the output of the device, referring to the word vector representation method of natural language processing. Combine the possible influencing factors data at a certain moment into a vector representation to form a data sequence of length 20. Due to the delay problem, the field device is usually not affected by the current time input. This paper selects various factors as the input in the first 20 moments of a certain moment. The data sequence of 20 time points is longitudinally sorted by time sequence to construct a 20×20 size data matrix.

B. Model Training And Parameter Setting

Data matrix is normalized before model training. The processed 20×20 data matrix is input into the model, where the CNN part is responsible for feature extraction and the LSTM is responsible for state prediction. Experiment sets up a 3-layer convolution layer with the volume number of cores to 4, 8, and 16 respectively. The size of the convolution kernel is 3×3, the convolution step size is set to 1, and the pool size in the pooled layer is set to 2. The data features are extracted through three-layer convolution and two-layer pooling, after being processed by the fully connected layer. The convolutional neural network outputs one-dimensional feature data of length 512 into the LSTM layer. The LSTM layer is set to 2 layers with the number of neurons of 16 and 32 respectively. Dropout is performed in the LSTM network to prevent overfitting and the state value of the device is predicted by the model.

Since the field data collected by the sensor is one point per second, the created data matrix also intercepts the data sample in seconds according to the sliding time window. For example, The first input matrix needs to arrange the sequence data from the 1st second to the 20th second in chronological order to form a data matrix as input. The second input matrix requires sequence data from the 2nd second to the 21st second to be arranged in chronological order to form a data matrix as input. In order to ensure the accuracy and universality of the experiment, the experiment used a continuous 600,000 sets of data from a week in September 2018 to train the model. In addition, to ensure the validity and accuracy of the experiment, the prediction of equipment failure was set data points within 1200 minutes. In order to verify the feasibility of the experimental model, the experiment uses the LSTM network for comparative analysis, using the same period of data to train the LSTM network, and also predicts 1200 data points within 20 minutes of the device.

C. Evaluation Index

This paper introduces formula (7-9) to calculate the mean absolute percentage error (MAPE), root mean square error (RMSE) and accuracy (ACC) as the evaluation index formula. The specific calculation formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{true}(i) - y_{pred}(i)}{y_{true}(i)}\right| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{true}(i) - y_{pred}(i))^2} \tag{8}$$

$$\mathrm{ACC} = \left|\frac{\mathrm{y_{true}} - \mathrm{y_{pred}}}{\mathrm{y_{true}}}\right| \times 100\% \tag{9}$$

Which y_{pred} represents the experimental prediction value, y_{true} representing the practical value, and n represents the total prediction duration.

D. Model Verification

In the experimental test phase, the device data of October 3, 2018 and November 3, 2018 were selected as test data and input into the experimental model and the LSTM model. The prediction results of the two models are compared with the measured data. Figure 6 shows the results of comparing the predicted values of the two models to the practical values of the equipment at a certain time on October 3. Since the model considers multiple influencing factors in the prediction, the prediction processing at the corner of the curve is more optimized in the prediction graph. Also, the prediction result is closer to the practical value. Table 1 and Table 2 shows that the two models represent the error between the predicted value and the practical value in the predictions of October 3 and November 3 respectively. It can be seen from the table that the root mean square error and accuracy of the predicted values of the experimental model are significantly improved compared with the LSTM network, indicating that the prediction of the model has more accurate prediction results than the traditional LSTM network. The mean absolute percentage error of the model indicates that the error of prediction is slight. That is it has good universality with better robustness and stability, and is more suitable for practical applications.

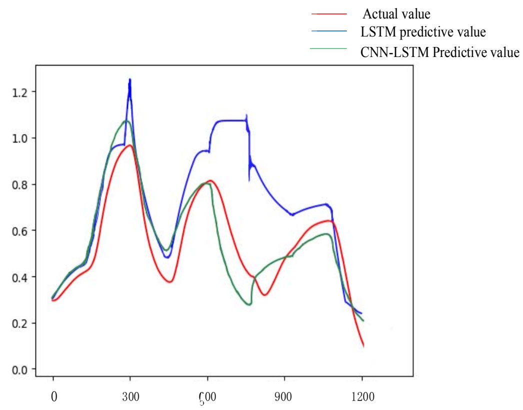

Figure 6. Comparison diagram of prediction and practical curve

TABLE I. COMPARISON OF MODEL PREDICTION VALUE ON OCT.3

	MAPE	RMSE	ACC
LSTM	0.152	0.177	76.738
CNN-LSTM	0.124	0.126	82.064

TABLE II. COMPARISON OF MODEL PREDICTION VALUE ON NOV.3

	MAPE	RMSE	ACC
LSTM	0.147	0.169	78.082
CNN-LSTM	0.127	0.133	83.227

IV. CONCLUSION

In this paper, based on the existing fault predicting technology, a short-term prediction method based on the hybrid model of CNN-LSTM network is proposed. Compared with the LSTM method, this method improves the accuracy of fault detection.

The advantage of this method compared with the LSTM network method is that a large amount of data is extracted by the convolutional neural network through the data matrix method, which not only considers the influence of various factors on the state of the device, but also greatly reduces the calculation amount of directly using the traditional LSTM model. Then sent the extracted feature vector to the LSTM for processing, which makes predicted data more comprehensive and more accurate. This experimental method can be improved gradually when it is put into actual work, then it can have a more accurate prediction of equipment failure.

REFERENCES

[1] Baydar, Naim, and A. Ball. Detection of Gear Failures in Helical Gears by Using Wavelet Transforms. Proceedings of the 33rd International MATADOR Conference. Springer London, 2000.

[2] Shangkun Liu, Guiji Tang, Bing Pang. Gearbox fault diagnosis method combining wavelet denoising and Hilbert demodulation. Mechanical Engineers, 2014(4): 103-105

[3] Gabor, D. . "Theory of communication." Iee Proc London 93.73(1946):58.

[4] Ning, L. I., et al. "Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine." Mechanical Systems & Signal Processing 28.2(2012):608-621.

[5] Frei, Mark G. , and I. Osorio . "Intrinsic time-scale decomposition: time–frequency–energy analysis and real-time filtering of non-stationary signals." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 463.2078(2007):321-342.

[6] Lapedes, Alan, and R. Farber. "How Neural Nets Work." Neural Information Processing Systems, Denver, Colorado, Usa 1987.

[7] Yam R C, Tse P W. Intelligent predictive decision support system for condition based maintenance. International Journal of Advanced Manufacturing Technology, 2001, 17(5): 383-391

[8] Xiang Z, Deyun X. Fault diagnosis of nonlinear systems using multistep prediction of time series based on neural network. Control Theory & Applications, 2000, 17(6): 803-808

[9] Wang P, Vachtsevanos G. Fault prognostics using dynamic wavelet neural networks. IEEE Proceedings of Systems Readiness Technology Conference. Valley Forge, PA, 2001: 857-870

[10] Hao Sun, Zhichun Yang, Lingling Zhang. Research on Vibration Response Trend Forecast Based on Neural Network. Mechanical Science and Technology, 2006, 25(12): 1454-1457+1518

[11] Graves, Alex , A. R. Mohamed , and G. Hinton . "Speech Recognition with Deep Recurrent Neural Networks." (2013).

[12] Meiying Zhang,Jie He.Introduction to Time Series Prediction Model Research. Jiangxi Science. 2009(05)

[13] Renze Luo,Peng Cao,Yunzhong Dai.Rotating Machinery Fault Diagnosis Theory and Implementation. Instrument Technology and Sensors, 2014(3): 107-110

[14] Feiyan Zhou,Linpeng Jin,Jun Dong.Review of Convolutional Neural Network. Journal of Computer Science，2017，40（6）：1229-1251.

[15] Feiyan Zhou, Linpeng, Jin,Jun Dong. Review of Convolutional Neural Network. Computer Science, 2017, 40(6): 1229-1251

[16] Kim Y. Convolutional neural networks for sentence clas-sification. Eprint Arxiv，2014.

[17] Hochreiter S， Schmidhuber J. Long short-term memory. Neural computation，1997，9(8)：1735-1780.