Study on the oil particle contamination forecasting Using LSTM network

Liangliang Zhai
REI, School of Energy and Power Engineering
Wuhan University of Technology
Wuhan 430063, Hubei, China
lishuai9368@163.com

Kun Yang
REI, School of Energy and Power Engineering
Wuhan University of Technology
Wuhan 430063, Hubei, China
kunyangwhut@163.com

Biao Hu
REI, School of Energy and Power Engineering
Wuhan University of Technology
Wuhan 430063, Hubei, China
137497191@qq.com

Shuai Li
REI, School of Energy and Power Engineering
Wuhan University of Technology
Wuhan 430063, Hubei, China
lishuai9368@163.com

Abstract*—As one of the main techniques of equipment condition monitoring, oil monitoring technology plays an extremely important role in evaluating the current state of equipment and predicting the development trend of equipment. In this paper, the LSTM neural networks was established by the historical data collected by a power plant. Using the cross validation method, and compared whit the popular time series prediction algorithm LSM, ARIMA, BPNN, SVR and RFR in the same test set, LSTM got the lowest RMSE value 42.26, which validates the applicability and accuracy of the LSTM neural network in the prediction of oil particle contamination.

Keywords—Oil-online monitoring; Oil particle; Deep Learning; LSTM neural networks

I. INTRODUCTION

Steam turbine plays an important role in generating capacity of power plants. Its reliability and safety are one of the most concerned problems in power plants. With the increasing complexity of power generation systems, the request for reliability of generators is also getting higher. Oil monitoring technology is a subject for the development of fault diagnosis. Its principle is to monitor the tribology state of the engine by monitoring the particles and tribology in the lubricant oil of the engine [1-2]. At present, many scholars have used traditional statistical methods and machine learning methods to model the historical gear box oil monitoring data collected by the power plant, and have achieved a great deal of achievements.

The traditional data driven methods, such as the least square methodthe autoregressive moving average(ARMA) [3], the back propagation neural network [4], the support vector regression[5],the random forest regression(RFR)[6] have been used in the field of oil monitoring. However, traditional statistical methods or machine learning methods are difficult to extract effective features from complex data produced under actual production conditions [7], the deep neural network model can abstract and combine low-level features of input signals and dig out deeper potential laws [8].

With the continuous development of deep learning methods, many deep learning models begin to be used in timing series forecasting. A special convolution neural network architecture, called LSTM RNN, has been proposed by Hochreiter and Schmidhuber [9]. Unlike traditional RNN, the LSTM model can learn long time sequences and overcome the effects of gradient disappearance, gradient explosion, and long term memory insufficiency [10], which makes it unique in memory of long and short time sequence information. In recent years, LSTM neural network has been successfully applied to many studies involving time series prediction: Such as the prediction of atmospheric pollutant concentration[11], traffic flow prediction[12]short term prediction of wind power [13], fault time series prediction [14], early prediction of rice blast in agriculture[15], pollutant emission forecast of generator set [16] and so on.

Combined with the advantages of deep learning technology, the LSTM model is applied to predict the number of oil particles in steam turbine gearboxes. Through modeling training, parameter optimization, RMSE as evaluation index, the accuracy rate of the test set is better than the traditional method in the test set by cross validation, which shows the superior performance of LSTM.

II. RECURRENT NEURAL NETWORK (RNN)

A. Simple recurrent neural network

The difference between simple recurrent neural network and ordinary deep neural network (DNN) is that the architecture of neuron connection is different and has the idea of circular feedback. As shown in Figure 1, recurrent neural network (RNN) is an improved multilayer perceptron network, including 1 input layer, 1 hidden layer, and 1 output layer. Unlike the DNN, there is a connection in the hidden layer that is input to the next hidden layer in the hidden layer. Expanding the hidden layer according to time, the input of each node contains the input of current input layer and the last output of the hidden layer.

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

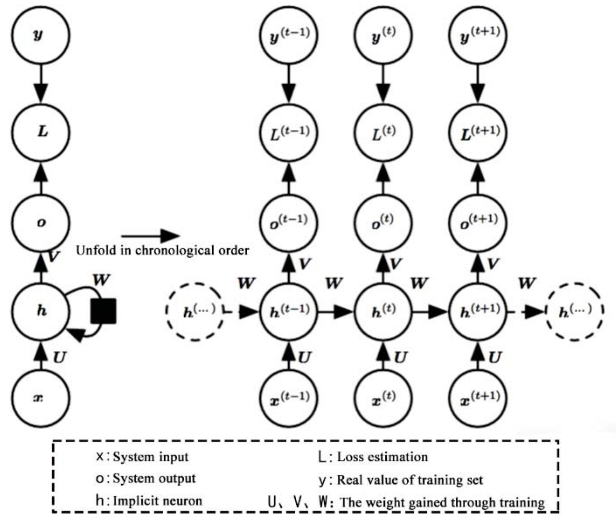

Fig.1 Recurrent neural network

RNN is a forward propagation algorithm. the reverse propagation principle is similar and the formula is slightly different. The current layer calculates the input data and outputs the data to the next layer according to the network connection and the weight value. And the calculation process is:

$$S_j(\mathrm{t}) = f(\sum_j^1 x_i(\mathrm{t})\,\mathrm{v}_{ji} + \sum_h^m s_h(\mathrm{t}-1)u_{jh} + \mathrm{b}_j) \qquad (1)$$

Where $\sum_j^1 x_i(\mathrm{t})\,\mathrm{v}_{ji}$ is the input of the input layer at the time t, $\sum_h^m s_h(\mathrm{t}-1)u_{jh}$ is the input of the hidden layer at the time t-1, b_j is the bias, $f(\bullet)$ is map function, $S_j(\mathrm{t})$ is the output of the input layer at the time t.

B. Long short term memory neural network

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

In view of the failure of RNN, it is used to describe the long-term correlation of data. The long and short term memory neural network has changed the structure of the control storage state on the basis of the common RNN network structure [9 17]. The computing node of LSTM is shown in Figure 2:

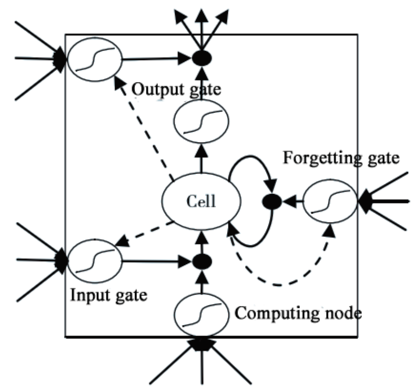

Fig.2 Computing nodes of LSTM

The computing node is composed of 4 parts: the forgetting gate, the input gate, the output gate and the Cell. The gate structure is a neural network with sigmoid as the activation function and a set of operations by bit multiplication, and the sigmoid function range is [0,1], so the output value of the neural network which is used as the activation function is also between 0 and 1. This reflects whether the structure allows input to pass. The core idea behind LSTM is a storage unit that can maintain its status information for a long time. The gate structure alleviates the problem of gradient disappearance. The LSTM model can be described by the following formulas:

Input gate i_t:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}\,\mathrm{c}_{t-1} + \mathrm{b}_i) \qquad (2)$$

Forgetting gate f_t:

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}\,\mathrm{c}_{t-1} + \mathrm{b}_f) \qquad (3)$$

Cell is the core of the node, and the calculation formula is as follows:

$$c(\mathrm{t}) = f_t \otimes c_{t-1} + i_t \otimes g(W_{cx}x_t + W_{cm}m_{t-1} + \mathrm{b}_c) \qquad (4)$$

Output gate o_t:

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}\,\mathrm{c}_{t-1} + \mathrm{b}_o) \qquad (5)$$

The output of the hidden state h_t is updated based on the state of computing unit c_t finally:

$$h_t = o_t * \tanh(\mathrm{c}_t) \qquad (6)$$

In the formula (2) (3) (4) (5) (6), $\sigma(\bullet)$ and $g(\bullet)$ are the sigmoid activation function and the hyperbolic tangent function respectively, W is the weight matrix of the cyclic connection. x_t, m_t represent the input data and memory information in the t time period respectively, b is the bias vector.

III. Experiment Verification and Result Analysis

A. Experimental preparation

The data of this paper came from a VG46 steam turbine of a power plant in Luoyang. There are 11895 data, and the data distribution is shown in Figure 3. It can be seen from the diagram that when the power is greater than 500MW, the functional relationship between the output power and the number of the generated 4μm particles is not obvious. Therefore, in the process of establishing LSTM model, raw data are used as input and output features. The power generation time series is used as the input data of LSTM, and the predicted value of 4μm particle number is LSTM output data, and the test set accounts for 10% of the total data.

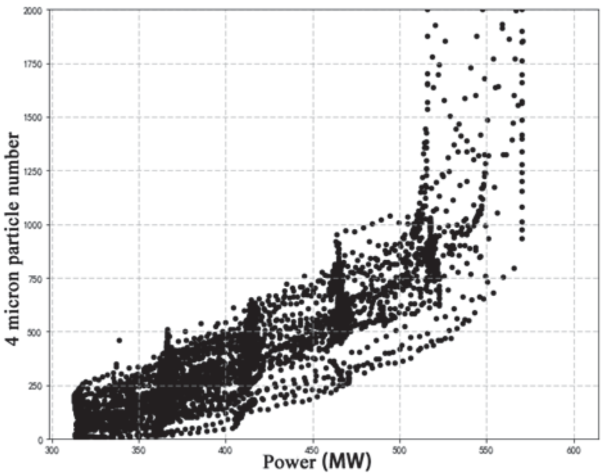

Fig.3 The scatter plot of the relationship between the power of steam turbine and 4μm particle

The standard root mean square error(RMSE) was chosen for the evaluation of experimental results. This is a widely used evaluation in regression analysis, and the standard root mean square error can describe the discreteness between the real value and the predicted value [18]. The RMSE formula is as follows:

$$\sigma = \sqrt{\frac{\left(\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2\right)}{n}} = \sqrt{\frac{\sum \varepsilon_i^2}{n}} \tag{7}$$

ε_i^2 represents the square of the difference between the predicted number of 4μm particle and the actual value.

B. Comparative experimental model

(1) LSM model

The least square method is a mathematical optimization technique. By minimizing the square of the error and matching the best function of the data, the matching function can be easily obtained by the least square method and used for prediction. As for the choice of function power, it can be judged according to the distribution map of actual data. Generally, the higher the power is, the more accurate the fitting is, and the risk of over fitting may also be brought. Therefore, cross validation can be used to prevent over fitting.

(2) ARIMA model

The autoregressive moving average model is the classical theory and method of time series analysis. The model can be expressed as ARIMA (p, d, q), of which p, d and q are the number of autoregressive items, the number of differences and the number of moving average items respectively [19]. The differential times d is mainly to transform the nonstationary time series into a stationary time series, and then use the Statsmodels packet of Python to find its autocorrelation coefficient and partial autocorrelation coefficient, and the best p and q are obtained by analysis. The model is tested with AIC (Akaike information criterion), BIC (Bayesian information criterion), and HQIC (Hannan-quinn information criterion).

(3) BPNN model

BP neural network is a feedforward neural network. It is one of the most widely used models in artificial neural networks. It can not only be classified but also can be regressive. It has a perfect theoretical system and learning mechanism [4]. Its topology consists of 3 layers: input layer, hidden layer and output layer, and the hidden layer can have more than 1 layers. The main parameters are the number of hidden layers, the number of nodes in each layer, the learning rate and the activation function. In this paper, we find the best parameters according to the open source Python grid optimization algorithm in GitHub.

(4) SVR model

Support vector regression (SVR) is a branch of support vector machine of machine learning method, which can be used to model time series. It maps the multidimensional input to the higher dimension of the feature space through a nonlinear kernel function, then performs the regression operation [5], and then obtains the nonlinear mapping relation with the output index. Generally, the Gaussian radial basis function (RBF) is selected as the kernel function. The main parameters are the penalty factor C and the sliding window length L, which are usually combined with optimization algorithm such as PSO and GA.

(5) RFR model

As a statistical theory, random forest regression integrates the idea of ensemble learning and stochastic subspace method [6]. A single decision tree is set up by random sampling in the sample space, and each tree will give a prediction result when it is used to predict. Finally, all decision trees are obtained, and the mean value is the final prediction result. The main parameters are the number of trees, the maximum depth, the single tree algorithm type. The advantage is that the better results can be obtained without the precise adjustment of the parameters, and the performance is stable.

C. Experiment result and analysis

This paper used Python3.6,TensorFlow1.2.0,Scikit-learn0.18.1 and Keras2.1.5 to implement LSTM model. The larger the hidden layer of LSTM is, the stronger the nonlinear fitting ability. At the same time, the complexity of the model and the computation time increase exponentially. Therefore, this paper set the hidden layer number to 2, then the LSTM is composed of 5 consecutive layers, which are 1 input layer, 2 hidden layers, 1 dense layer and 1 output layer, and the number of nodes in each layer is 64. The hidden layer is used to establish a model relationship between past and future time series. The dense layer is used to change the dimension of the output vector from the previous LSTM layer to match the output to the final prediction time series. Many optimization algorithms can be selected when choosing the adjustment model to update weights and parameters. The Adam (Adaptive Moment Estimation) algorithm is used here [20], which not only stores the exponential decay average of the previous square gradient of the AdaDelta algorithm, but also maintains the exponential decay average of the previous gradient M (T). Compared with the other adaptive learning rate algorithms, the convergence speed is faster, and the learning effect is more effective, and The problems in other optimization techniques can be corrected, such as the loss of learning rate, slow convergence or the updating of high variance parameters, which results in greater loss function and so on [20]. The initial learning rate of the model is set to 0.001, and the number of training iterations is set to 1000 epochs.Since the working state of generator set has continuity in time, the value of time steps should be determined by the actual situation in the production process. When other conditions such as ambient temperature, oil temperature, dielectric constant, moisture content and acidity change, the correlation between data will also change. Therefore, it is necessary to test the selection of time step. In the design experiment, the time step is trained from the 10-120 steps with 10 intervals to train the LSTM model, and the RMSE value obtained after each training is recorded. The result is shown in Figure 4.

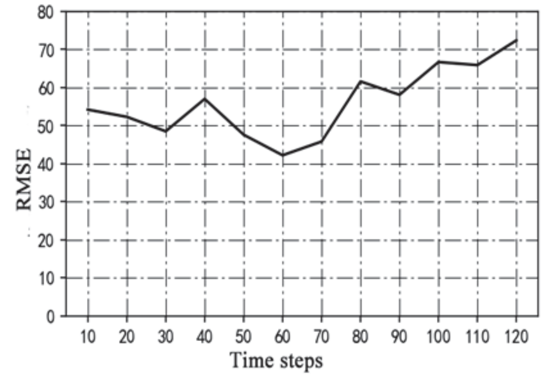

Fig.4 The change diagram of RMSE with timesteps

It can be seen from the diagram that when the time step increases from 0, the RMSE tends to decrease, and when the time step is equal to 60, the RMSE value is the smallest, which is 42.26. When the time step is greater than 60, the RMSE begins to increase. Because the input time series is too short, the intrinsic correlation can not be fully captured. When the time series of input is too long, the correlation of time series will decrease and the precision of the model begins to decline. Therefore, the time step is set to 60 according to the above analysis. At the same time, it also shows that LSTM needs to adjust the data collected from different working environments in order to achieve better prediction results. Therefore, the constructed LSTM network structure is shown inFigure5

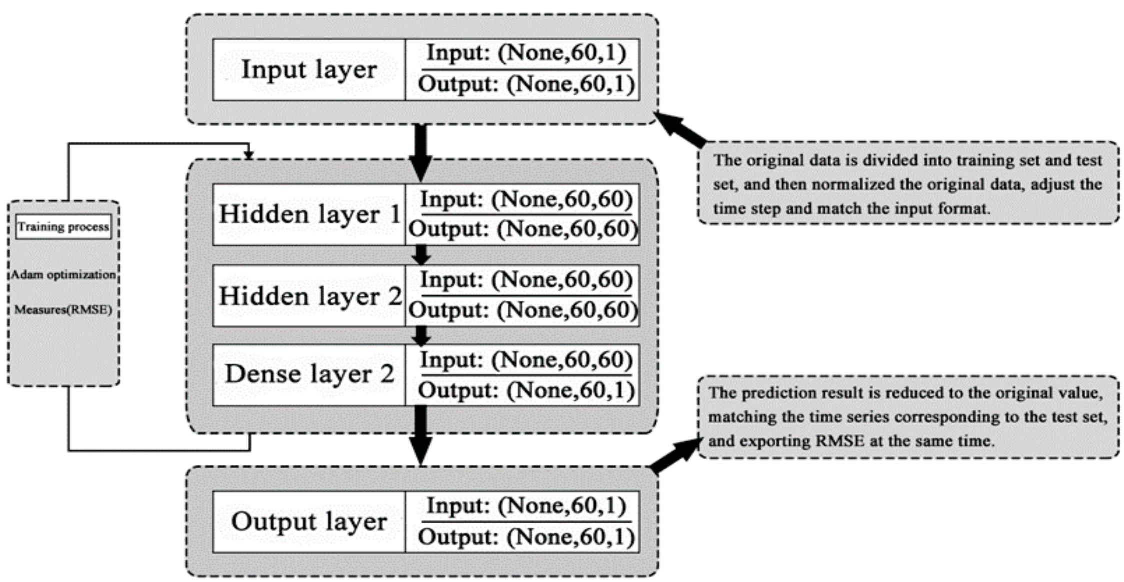

Fig.5 Illustration of the proposed LSTM network for time series prediction

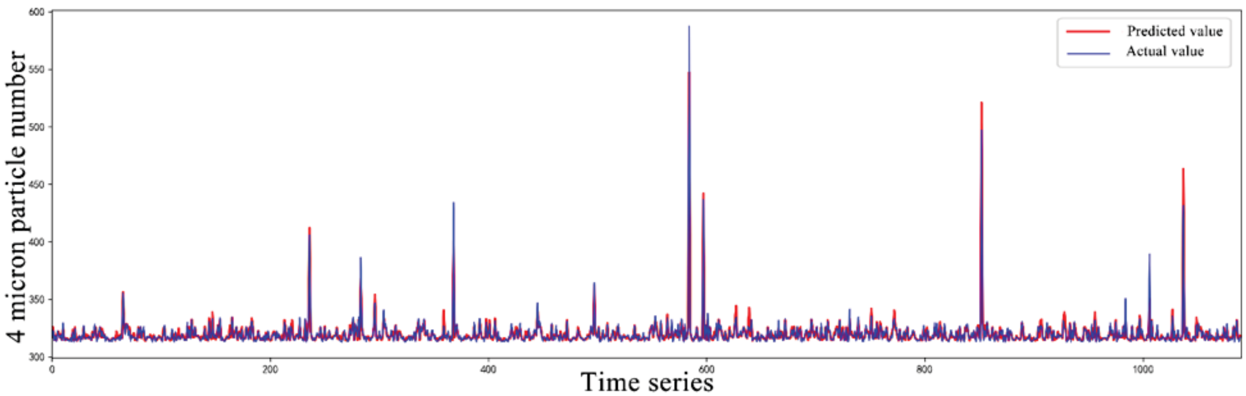

Fig.6 The curve of the predicted value and the actual value of LSTM on the test set

Figure 6 shows the prediction effect of the above LSTM model on the test set. It can be seen clearly that the predicted value curve can capture the actual change of the

number of 4μm particles, and can match the real value curve accurately

D. Comparison of experimental results

In order to verify the predictive effect of LSTM, a variety of time series prediction algorithms mentioned in the 2.2 section are modeled on the same data set. The average RMSE of each model is obtained through 5 cross validation, and the result is shown in Table 1.

Table.1 Comparison of results between LSTM and LSM, ARIMA, BPNN, SVR and RFR algorithms

Model name	RMSE	Parameter description
LSTM	42.26	The input layer is set to 1, the hidden layer is set to 2, the dense layer is set to 1, the output layer is set to 1, the number of nodes in each layer is set to 64, the time step is set to 60, the learning rate is set to 0.001.
LSM	160.25	Fitting function set to quadratic function.
ARIMA	135.93	p=3，d=1，q=1
BPNN	68.71	The number of input nodes is set to 1, the number of hidden layers is set to 2, the number of input nodes is set to 2, the number of hidden layers is set to 2, the number of hidden nodes is set to 64, and the number of output nodes is set to 1.
SVR	109.54	The basis function is the RBF function, the penalty factor C=1000, and the sliding window L=3.
RFR	71.76	It contains 100 trees, and CART, C4.5, ID3 decision tree types each account for 1/3, and the maximum depth is 8.

From the above table, we can see that the RMSE value of the LSTM model on the test set is the smallest, which shows that the method has the highest prediction accuracy for the 4μm particle number of the turbine unit. The other 5 kinds of regression algorithms are commonly used algorithms for predictive subjects. Compared with them, when LSTM is

IV. Conclusion

In this paper, a LSTM model based on the prediction of the number of 4μm particles in the history of turbine lubricant is presented. The LSTM model can model the time series according to the relevance of time, and search the best parameters according to the set optimization algorithm. Compared with other prediction algorithms using the same dataset, LSTM is superior to them. The following is the conclusion of this study.

1）Compared with the traditional shallow models, such as SVR, LSM and ARIMA models, the deep learning model has better prediction performance.

2）Compared with the single neural network DNN model and the BP neural network model, the RMSE value shows that the LSTM neural network can capture the time correlation more effectively and have higher predictive performance.

The next work is to incorporate all the parameters of LSTM into the scope of optimization, and incorporate other important parameters of oil into the prediction results. A comprehensive multiple input multiple output LSTM model for oil information prediction will be established.

REFERENCE

[1] YANG Q M, YAN X P, HE S Z. Field practical technology for oil monitoring and analysis[M]. Beijing: China Machine Press, 2006.

[2] LI Y J，ZUO H F，WU Z F. Intelligent diagnostics for engine wear failure based on debris analysis［J］. Journal of Nanjing University of Aeronautics and Astronautics, 2001, 33 (3): 221-226.

[3] ZHANG G P. Time series forecasting using a hybrid ARIMA and neural network model [J]. Neurocomputing, 2003 (50):159-175.

[4] KUNG S Y. An algebraic projection analysis for optimal hidden units size and learning rates in back- propagation learning[C]// International Conference on Neural Network, 1988:363-370.

[5] Theodore B T, Huseyin I. Support vector machine for regression and application to financial forecasting [J]. Proceeding of the IEEE-INNSENNS International Joint Conference on Neural Network，2000，(6): 348-353.

[6] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1):5-32.

[7] SMOLA A J，SCHOLKOPF B. A tutorial on support vector regression[J]．Statistics and Computing, 2004，14(3): 199-222.

[8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature,2015,521(7553): 436-444.

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

[10] GRAVES A. Long short-term memory[M]. Springer Berlin Heidelberg,2012: 1735-1780.

[11] XIANG L, LING P. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation[J]. Environmental Pollution 231, 2017: 997-1004.

[12] MA X, TAO Z, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transp. Res. Part C Emerg. Technol. 2015(54): 187-197.

[13] ZHU Q M, LI H Y, et al. Short-term wind power forecasting based on LSTM[J]. Power System Technology, 2017, 41(12): 3797-3802.

[14] WANG X, WU J, et al. Fault time series prediction based on LSTM r ecurrent neural network[J/OL].Journal of Beijing University of Aeron autics and Astronautics:1-12[2018-03-16].https://doi.org/10.13700/j.b h.1001-5965. 2017.0285.

[15] Kim Y, Roh JH, Kim HY. Early forecasting of rice blast disease using long short-term memory recurrent neural networks[J]. Sustainability, 2018, 10(1), 34.

[16] YANG X Z, KE Y Y, et al. A study on the prediction of generator set pollutant emissions based on LSTM[J]. Electrical Automation, 2016,38(05): 22-25.

[17] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. Berlin: Springer, 2012.

[18] Hyndman RJ, Koehler Anne B. Another look at measures of forecast accuracy[J]. International journal of forecasting, 2006, 22(4): 679-688.

[19] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research,2011,12(7): 257-269.

[20] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. Eeprint Arxiv:1412.6980,2014.