Short-term power generation load forecasting based on LSTM neural network

Bicheng Huang\textsuperscript{a*}, Langxing Tong\textsuperscript{b}, Yu Zuo\textsuperscript{c}

\textsuperscript{1}Guixi branch of state power investment group Jiangxi Power Engineering Co., Ltd, Yingtan335000, China
\textsuperscript{a*}huangbichengyx@163.com, \textsuperscript{b}13870002180@163.com, \textsuperscript{c}gxfdcjdbh@163.com

Abstract—With the gradual deepening of power market reform, power generation enterprises' prediction of their future short-term power generation load is conducive to affecting the load distribution of power grid dispatching to power generation enterprises, so as to achieve the purpose of power marketing. In this regard, this paper proposes a power generation load forecasting method based on long and short-term memory (LSTM) neural network, which takes the power generation load sequence data as the model input to predict the future short-term power generation load of power generation enterprises, and proves that compared with the traditional Neural network method, LSTM neural network has higher prediction accuracy.

1. Introduction
With the continuous increase of new energy installed capacity and the impact of hydropower’s high and low water periods on thermal power generation, coupled with the continuous deepening of the power market reform\cite{1}, the competition among power generation companies has become increasingly fierce, and power marketing has become an important means to enhance the competitiveness of power generation companies.

The power grid provides all kinds of users with electrical energy that meets the requirements. Due to the difficulty of storage of electrical energy\cite{2}, how to reasonably realize the distribution of power supply loads and ensure stability and economy has become an important issue for the power grid sector \cite{3}. By forecasting the short-term power generation load of power generation enterprises, it can effectively reflect the load bearing capacity of power generation enterprises, which is conducive to affecting the load distribution of power grid dispatching to power generation enterprises, so as to achieve the purpose of power marketing. At the same time, in view of the sharp rise in domestic coal prices and the difficulties in coal transportation, the short-term prediction of power generation load also plays a guiding role in the use and storage limit of power coal for power generation enterprises.

The power generation load has the characteristics of time series and non-linearity\cite{4}. Based on its characteristics, the forecasting model research is generally divided into two categories. One is the time series analysis method, such as regression analysis method, exponential smoothing model method, Kalman filter method, Multiple linear regression, Fourier expansion model and autoregressive moving average model. The basic idea is to predict the future generation load value from the past generation load and current generation load of a random time series. The advantage is that the time series relationship of the data is considered, the disadvantage is that the predictive ability of non-linear relationship data is limited; the second is machine learning analysis methods\cite{5-7}, such as back propagation neural network, grey projection and random forest algorithm, deep belief network...
prediction, multi-core support vector machine algorithm to regression prediction, expert system method prediction. The common problem of these algorithms is the lack of consideration of the time correlation of time series data, and it is necessary to artificially add time features to ensure the prediction results.

With the development of big data technology, big data has now become a research topic of common concern in academia and industry[8]. The in-depth application of data collection technology, network technology and computer technology in power generation companies, the increase in the number of various sensors and smart devices and the continuous improvement of their functions, all types of corporate data are becoming more transparent and easy to obtain[9]. The collection frequency and data accuracy of power generation and other information are also continuously improved, which provides a high-quality and massive data set for the research of power generation load forecasting, and also provides a data basis for the use of deep machine learning.

As an effective nonlinear cyclic neural network, Long Short-Term memory network is gradually used in the field of prediction because it takes into account the timing and nonlinear relationship of data. The Long Short-Term Memory network model can fully reflect the long-term historical process in the input time series data. Therefore, this paper proposes a load forecasting method based on the Long Short-Term Memory time series neural network to forecast the short-term load of power generation enterprises.

2. Establishment of LSTM neural network model
LSTM neural network is a special type of RNN neural network[10], a time recursive sequence neural network, which can solve the long-term dependence problem of general RNN, that is, LSTM is suitable for processing and prediction Events with very long intervals and delays in a time series. Different from the single neural network layer of the RNN, the LSTM neural network has four, that is, in addition to the external RNN loop, it also has an internal LSTM self-loop. As shown in Fig.1, interact in a very special way.

The LSTM algorithm adds a "processor" structure (cell) to judge whether the information is useful. There are three gates in the cell, which are called input gate, forgetting gate and output gate. When information enters LSTM, it is necessary to detect whether it is useful according to rules. Only the information that conforms to the rules can be left, and the information that does not conform to the rules will be abandoned through the forgetting door.

![Fig.1 structure diagram of LSTM hidden layer module](image)

It can be seen from Fig.1 that, in addition to the hidden state $H_t$ that is the same as the RNN, there is a hidden state $C_t$ that propagates forward at time $t$. This hidden state is called the cell state.

The first box called $\sigma$ in Fig.1 is the forgetting gate, which is the switch to control whether to forget, that is, to control whether to forget the hidden cell state of the upper layer with a certain probability. Among them, $\sigma$ represents the sigmoid activation function. The function expression is shown in formula(1):
\[ S(x) = \frac{1}{1+e^{-x}} \]

The hidden state \( H_{t-1} \) of the above sequence and the data \( X_t \) of this sequence are used as inputs, and the forgetting gate output \( F_t \) is obtained through the sigmoid activation function. Since the value of sigmoid function is between \([0,1]\), \( F_t \) represents the probability of forgetting the hidden cell state of the previous layer.

The output of the forgetting gate is shown in formula (2):
\[ F_t = \sigma(\omega_f \cdot [H_{t-1}, X_t] + b_f) \]

Where: \( \sigma \) represents sigmoid activation function, \( \omega_f \) represents the connection weight and \( b_f \) represents the threshold.

The second \( \sigma \) box and tanh box in Fig.1 (generally only tanh box in RNN network) are input gates, which are responsible for processing the input of the current sequence position, and tanh is also an activation function. The function expression is shown in formula (3):
\[ \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \]

It can be seen that the input gate is composed of two parts. The first part uses the sigmoid activation function and the output is \( I_t \); the second part uses the tanh activation function, and the output is \( C_t \). After multiplying the two results, the cell state is updated.

At this time, the memory unit status is updated to formula (4):
\[ C_t = F_t \bar{C}_{t-1} + I_t \bar{C}_t \]
\[ I_t = \sigma(\omega_i \cdot [H_{t-1}, X_t] + b_i) \]
\[ C_t = \tanh(\omega_t \cdot [H_{t-1}, X_t] + b_t) \]

The update of the hidden state \( H_t \) consists of two parts. The first part is \( O_t \), which is obtained by the hidden state \( H_{t-1} \) of the previous sequence and the data of this sequence \( X_t \), and the sigmoid activation function. The second part is composed of the hidden state \( C_t \) and the tanh activation function.

At this time, the output formula is shown in formula (5):
\[ O_t = \sigma(\omega_o \cdot [H_{t-1}, X_t] + b_o) \]
\[ H_t = O_t \tanh(C_t) \]

Therefore, when the data enters the LSTM time series neural network for processing, the “cell” can choose the information according to the preset parameters. Only after passing the discrimination can part of the information be saved, otherwise it will be abandoned through the forgetting gate. Therefore, after removing more interference information, the prediction accuracy of LSTM time series neural network will be higher.

3. Generation load forecasting model based on LSTM

3.1. Preprocessing of load data

In the process of power generation, power generation enterprises may have planned maintenance, temporary maintenance, load shedding, etc. according to power grid dispatching and their own factors. As well as improper operation and equipment aging in the process of power generation load data collection. At this time, the power generation load of the power generation company There will be outliers in the values, which will have a great impact on the prediction process, and outliers need to be cleaned and filled in the data.

For the situation where the power generation load value of multiple days is zero when the generator
set is scheduled to be overhauled, refer to the power generation load situation of the same type of generators running in the power generation company.

For the discrete and single outliers caused by improper data collection, the horizontal comparison method is used to identify and process the bad data, and the sample statistical index and the set threshold are used to judge whether there is abnormal data.

$$\bar{x}_n = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x}_n)^2}$$

(6)

Laida 3σ principle believes that the probability of a set of test data containing only random error values distributed in \((\mu-3\sigma, \mu+3\sigma)\) is 0.9973. Therefore, abnormal data can be detected and corrected according to the 3σ principle.

First, calculate the mean and variance of the numerical series of generation load according to formula (6):

Then use the formula (7) to judge the abnormal number of the 3σ principle, where \(\varepsilon\) is the threshold, usually 1~1.5.

$$|x_n - \bar{x}_n| > 3\sigma \varepsilon$$

(7)

Where: \(\varepsilon\) is the threshold, usually 1~1.5.

If the data satisfies formula (7), \(x_n\) is bad data, and formula (8) can be used for correction:

Where: \(x_n^*\) is the modified power generation load value, \(x_{n\pm1}\) is the value of the two power generation loads near \(x_n\), and \(x_n^{1,2}\) is the power generation load value that is closest to the value of \(x_n\).

$$x_n^* = \frac{\alpha}{2} \sum_{n=1}^{x_{n\pm1}} + \frac{\beta}{2} \sum_{n=1}^{x_{n\pm1}} + \frac{\gamma}{2} \bar{x}_n$$

$$\alpha + \beta + \gamma = 1$$

(8)

3.2. Model establishment

The generation load forecasting model is established by using LSTM time series neural network to predict the generation load \(X\) of power generation enterprises.

Suppose the sliding window size is \(d\). The definition of the sliding window is that LSTM uses the power generation load model data of the previous \(d\) (sliding window size) moments to predict the current data. When the current time is \(n\), the power generation load sequence value \([x(n-d+1) \ldots x(n-1) x(n)]\) is used to predict the power generation load forecasting model value \(x(n+1)\) of the power generation enterprise at the time \(n+1\).
Data preprocessing
Establish LSTM power generation load forecasting model and initialize parameters
Training LSTM generation load model
Meet the prediction accuracy or iterations
Input historical observation sequence $X_t$
Predicted value of generation load $X_{t+1}$

Fig. 2 flow chart of generation load forecasting based on LSTM

The overall forecasting process of power generation load using LSTM time series neural network is as follows:

1. Clean and fill the abnormal value of power generation load data of power generation enterprises, and normalize the data;
2. The LSTM time series neural network generation load forecasting model is established. Taking the generation load value of power generation enterprises as the input, the LSTM model parameters are initialized, and the first 90% of the generation load series values are selected to construct the training set;
3. The prediction model is trained with the training set. Adam function is selected for gradient descent. The gradient threshold is 1 and the initial learning rate is 0.005. After 350 generations of training, the learning rate decreases and the fading factor is 0.2;
4. Detect whether the root mean RMSE and LOSS are less than the threshold or whether the number of iterations is reached. If yes, continue to the next step; otherwise, return to step 3;
5. Input the first $d$ historical generation load values into the LSTM model to obtain the generation load values at the next time.

4. Test Results and Discussions
Before formal forecasting, it is necessary to de-trend the data to eliminate the influence of the offset generated in the data acquisition on the calculation, and the analysis can be concentrated on the fluctuation of the data trend itself.
In Fig.3, the blue solid line segment is the power generation load data curve of power generation enterprises obtained by data preprocessing. The purple dotted line is the data trend, that is, the least square fitting line of the load model data. The red dotted line is obtained after the trend of the original data is removed, and its mean value is 0 as shown by the orange dotted line below.

Using de-trend data to forecast can effectively reduce the forecast error, and the final forecast value can be obtained by adding the corresponding trend data to the predicted result.

The LSTM time series neural network is used to predict the generation load. 90% of the data is taken as the training set to train the LSTM neural network. The neural network contains one input layer, one output layer and 200 hidden layers. Due to the small amount of data, in order to improve the prediction accuracy, the batch_size is set to 1 and the number of iterations is set to 500 generations.

As can be seen from Fig.4, by the 350th generation, the root mean RMSE and LOSS have approached 0. From the 350th generation to 500th iterations, the two error curves have approximately coincided with the abscissa, and the training results of LSTM network are good.

Compare the prediction results of LSTM time series neural network in the test set with the prediction results of RNN cyclic neural network, and the results are shown in Fig.5.
Taking the power generation load predicted for 12 days as an example, the average error percentage MAPE value of RNN prediction is 4.40%, and the average error percentage MAPE value of LSTM prediction is 2.36%. Obviously, the prediction accuracy of LSTM neural network is higher than that of RNN neural network.

Fig.6 and Tab.1 show in detail the 12-day power generation load value and the LSTM predicted power generation value and prediction accuracy.

It can be clearly seen from Fig.6 and Tab.1 that the load forecasting method based on LSTM neural network has higher forecasting accuracy, which proves the superiority of this method for power generation load forecasting.

| Time serial number | Actual value | Predictive value | Accuracy | Time serial number | Actual value | Predictive value | Accuracy |
|--------------------|--------------|------------------|----------|--------------------|--------------|------------------|----------|
| 1                  | 0.825        | 0.871            | 94.69%   | 7                  | 0.850        | 0.877            | 96.97%   |
| 2                  | 0.832        | 0.822            | 98.73%   | 8                  | 0.806        | 0.840            | 96.07%   |
| 3                  | 0.899        | 0.875            | 97.28%   | 9                  | 0.873        | 0.901            | 96.94%   |
| 4                  | 0.933        | 0.947            | 98.53%   | 10                 | 0.891        | 0.905            | 98.54%   |
| 5                  | 0.872        | 0.889            | 98.02%   | 11                 | 0.842        | 0.855            | 98.54%   |
| 6                  | 0.834        | 0.848            | 98.39%   | 12                 | 0.935        | 0.938            | 99.79%   |
5. Conclusion
Based on the background that the current gradual normalization of the electricity marketing market has led to the fierce competition among power generation companies and the difficulty of coal storage caused by the rapid rise of coal prices, this paper proposes to use LSTM time series neural network to predict the power generation load of power generation companies. Subsequently, this article introduces LSTM, explains the applicability of LSTM for power generation load forecasting in principle, and uses the 3σ principle to detect and correct the input data of the neural network. Finally, the actual power generation load data of the generator set is used for simulation analysis, and the data is preprocessed and input into the model for training prediction. The result proves the superiority of LSTM neural network for power generation load prediction.

References
[1] Yan K., Yuming L., Liang G. (2021) Ideas on Construction of Electricity Retail Market Under Spot Market Mode. J. Power System Technology, 45(9):3417-3423.
[2] Ciwei G., Haiming L., Lulu L., and et al. (2016) The Energy-Saving Assessment and Optimization of Energy Storage and Electricity Utilization Technology Based on the Energy Efficiency Evaluation of Power System. J. Transactions of China Electrotechnical Society, 31(11): 140-148.
[3] Xingping Z., Yuqin N., Xu Z.. (2008) Cointegration Analysis of Electricity Consumption in China. J. Proceedings of the CSEE, 28(13): 114-119.
[4] Jing X., Yijiong G., Zhong W., and et al. (2017) Research on Indexes of Energy Efficiency and Its Reference-value for Coal-fired Power Units Based on Data-mining. J. Proceedings of the CSEE, 37(7): 2009-2015.
[5] Yufan Z., Qian A., Lin L., and et al. (2019) A Very Short-term Load Forecasting Method Based on Deep LSTM RNN at Zone Level. J. Power System Technology, 43(6): 1884-1891.
[6] Haifeng J., Xinyin X., Yaowu W.. (2002) A Short-term Load Forecasting Method Based on Cascade Neural Network. J. Power System Technology, 26(3): 49-51.
[7] Hao P., Jinfeng G., Yaoheng D.. (2020) Short-term Load Forecasting Method Based on Fusion of Multiple Neural Networks. J. Electric Power Automation Equipment, 40(6): 37-42.
[8] Wu X., Zhu X., Wu G.., et al. (2013) Data mining with big data. J. IEEE transactions on knowledge and data engineering, 26(1): 97-107.
[9] Dewen W., Zhiwei S.. (2015) Big data analysis and parallel load forecasting of electric power user side. J. Proceedings of the CSEE, 35(3): 527-537.
[10] Chen K., Huo Q.. (2016)Training deep bidirectional LSTM acoustic model for LVCSR by a Context-Sensitive-Chunk BPTT approach. J. IEEE/ACM Transactions on Audio Speech & Language Processing, 24(7): 1185-1193.