The Power of Short-term Load Algorithm Based on LSTM

Zhankui Wang1, Junying Wu1,*, Rui Xin1, Tao Bai2, Jianbin Zhao2, Minglei Wei2, Jingquan Li2 and Lei Zhuang1
1Information and Communications Branch, Hebei Electric Power Company, State Grid, China
2Hebei Electric Power Company, State Grid, China

*Corresponding author

Abstract. The power data is affected by many factors and has long time series, which fully meets the conditions of long-term and short-term memory neural network. This paper analyzes and models various influencing factors such as time, vacation and meteorology, and uses adaptive moment estimation optimization algorithm. The traditional optimization algorithm is improved, the generalization of the model is improved, and the short-term load forecasting of power can be stably and efficiently operated to ensure the reliability and safety of national electricity. In this paper, the short-term load forecasting experiment is carried out in a certain area of Hebei Province. The experimental results show that the proposed algorithm outperforms the existing similar model and has higher prediction accuracy.

Keywords: LSTM; influencing factors; adam; generalization; short-term load.

1. Introduction
The power system is an indispensable part of national construction and people's life. It is more and more important to realize efficient use of electric energy[1]. On the one hand, the prediction of power load can conveniently compare the load on the system, make the electrical equipment be used reasonably, and ensure the economic, safe and reliable operation of the power system. On the other hand, all power generation and power companies have higher and higher requirements for load forecasting accuracy[2]. Accurate load forecasting is conducive to each power generation enterprise to arrange scheduling plans reasonably to ensure that high-quality power is provided to the public at all times. Thorough maintenance and inspection can also help power companies take proactive measures to improve the quality and economy of electricity consumption[3].

With the improvement of the smart grid and the improvement of hardware equipment, it provides a large amount of high-quality data and powerful computing power for the load budget, providing a strong foundation for the application of deep learning and machine learning in electric load. The neural network load forecasting model considering temperature and other related factors reduces the adverse effects of uncertain factors such as weather to a certain extent. This also proves from the side, starting from the research method, paying attention to the influence of time factors on the model distribution, and establishing a time series model for research, is the tool that is more efficient for power engineering to solve problems in forecasting work. LSTM is suitable for forecasting of power.

2. Literature Review of LSTM

2.1. Long Short-term Memory
Long short-term memory(LSTM)[4] is born to solve this problem and has good time-sequence memory.
It has excellent processing ability for time series data and is an ideal model for power short-term load forecasting. Moreover, LSTM has greatly improved the problem of RNN\cite{5} gradient explosion and gradient dispersion\cite{6}. The cell structure of the LSTM is shown in the figure:

![LSTM Diagram](image_url)

**Figure 1. LSTM.**

The LSTM adds three types of gates\cite{6} to each neuron based on the RNN. The first is the input gate after $X_t$, which is generally completed by the sigmoid function. The main function is to filter the input values and then arrange the useful information and transmit it to the network. After the $S_{t-1}$, the forgetting gate is set, and the program is mostly completed by the tanh function, and the function is to further filter the reserved portion in the memory state at the previous moment. The last one is the corresponding output gate. Like the input gate, its program is also completed by the sigmoid function. It is set after the end of the network output process, and the main function is to filter and extract the important parts of the network output information. Prepare for the next step.

The calculation formula is:

\begin{align}
    f_t &= \sigma(w_f \cdot [H_{t-1}, X_t] + b_f) \\
    i_t &= \sigma(w_i \cdot [H_{t-1}, X_t] + b_i) \\
    g_t &= \tanh(w_g \cdot [H_{t-1}, X_t] + b_g) \\
    C_t &= f_t C_{t-1} + i_t g_t \\
    o_t &= \sigma(w_o \cdot [H_{t-1}, X_t] + b_o) \\
    H_t &= o_t \tanh(C_t)
\end{align}

3. Load Prediction Model Based on LSTM+Adam

3.1. Analysis of Influencing Factors

3.1.1. Temperature effect. Among the determinants of power demand, the temperature value outside the working environment is also one of them\cite{7}. In particular, the time variation span is not long. For example, under the short-term scale standard, the results of power load forecasting have obvious interference. Several aspects: On the one hand, temperature will change the conversion problem of power load consumption and demand mode; secondly, considering the demand characteristics of power work for temperature elasticity, the situation of elastic temperature change will affect the demand of electricity\cite{8}.

3.1.2. Day and night effects. Usually, we define the effect of electricity on daytime as the daylight effect. Under normal circumstances, the faster the lighting work consumes power, the greater the power consumption, which means the shorter the daylight hours. China's vast land area, resulting in different time periods, different regions have different lengths, so the chalk effect will also be applied to the experimental considerations. Specifically, it is based on sunrise and sunset to determine the impact of system power consumption in the adjacent period\cite{9}. Then, according to the sunrise time interval from
5 am to 7 am in Beijing and the interval between sunsets from 17 to 19 in the evening. Because the experiment uses the electricity consumption data of the factory, the situation reflected is that there is more electricity during the day and less electricity at night, which is opposite to the electricity consumption in the residential area.

3.2. Adam
Traditionally, in the iterative optimization of hyperparameters, most models use SGD, and its algorithm is unstable and the error is very large, resulting in inaccurate calculation results. In the process of technology update, the network neural weights are further updated, which leads to the prototype of the Adam optimization algorithm, which is the most important reason why the Adam optimization algorithm can achieve better results. In 2015, two scientists presented the first paper on the Adam algorithm at the Duolun University Institute, which attracted the attention of a large number of people engaged in related work in the world. The Adam optimization algorithm was born, and replaced the stochastic gradient descent model. The most popular optimization algorithm. Compared with the basic stochastic gradient descent algorithm, it has the advantage of fast speed; it can be used for non-stationary objective function/data, that is, the mean value of the gradient and the covariance change; it can be used for noisy and/or sparse gradients; The update speed is also faster in the same type of model, and it is not easy to fall into the local best.

Adam formula:

\[ t = t + 1 \]  
\[ g_t = \nabla_{\theta_t} f_t(\theta_{t-1}) \]  
\[ m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \]  
\[ v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \]  
\[ \theta_t = \theta_{t-1} - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon} \]

3.3. LSTM+Adam
The experiment uses the LSTM power load prediction algorithm, and the entire network model is shown in Figure 3. First, we need to pre-process and screen data such as temperature, day and night, and electricity consumption[10]. On this basis, we will establish a dimension of n (number of neurons per single layer)*f (included in each power data set) The network input layer of the total number of feature quantities). After the data set is forward-propagated, the data is integrated into the fully connected layer through the hidden layer calculation, and the final predicted value is obtained through the Re-LU linear rectification function mapping. This process is expressed as:

Where \( x_i \) is the i-th power data and the value of its influencing factor, \( y_i \) is the corresponding i-th predicted output value, \( f \) is the i-th neural network in the model, and \( \pi \) is the fully-connected layer.

In order to realize the automatic update of the network parameters, the experiment uses the back propagation method to deal with the error, and thus realizes the gradient-based derivation of the parameter
as a whole, and uses Adam optimization to finally output the predicted value.

3.4. Model evaluation index
In evaluating the regression prediction model, this experiment is described by RMSE, and the calculation formula is as follows:

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{x_i - \hat{x}_i}{x_i} \right)^2} \times 100\% \]  
(12)

Where \( x_i \) is the actual value of the load at time \( i \), and \( \hat{x}_i \) is the corresponding predicted value.

4. Experiment

4.1. Data Processing
In the process of analysis and inquiry, the author randomly designs and integrates the collected data sets according to a reasonable ratio of 3:1:1 to form the training set part, the verification set and the test. The characteristics of the data directly affect the performance of the analytical research model, and it is essential to determine whether it can reach the upper limit. Therefore, it is necessary to perform early detection and preprocessing of the data.

Since day and night is a Boolean variable, it is processed by the classification of 0 in the day and 1 in the evening. For factors such as weather that are not numerical attributes, such factor values are visually displayed by means of one hot thermal coding program design. Finally, by means of accelerated convergence, all the feature quantities are uniformly integrated and analyzed and preprocessed, so as to achieve the purpose of improving the prediction accuracy of the model.

5. Experimental Results and Analysis
In this experiment, the control variable method is used to determine the feature number first, and then the number of layers of LSTM is changed to test the influence of the model depth on the prediction result. The experimental data is shown in Table 1. Appropriate increase of the number of layers is conducive to the accuracy of prediction. However, after adding to the fourth layer, the accuracy is reduced and the model is over-learned, so the 3 layers are the optimal depth. Under the optimal depth conditions, the performance of the improved Adam was tested by changing the gradient descent algorithm. As shown in Table 2, it is obvious that the error rate of the LSTM neural network using Adam is significantly better.

![Figure 3](image.png)

**Table 1. Number of plies.**

| Layer | Time | Eigenvalues | \( y_{\text{rmse}} \) |
|-------|------|-------------|---------------------|
| 1     | 50   | 4           | 0.0012366506392     |
| 2     | 50   | 4           | 0.00122859173442    |
| 3     | 50   | 4           | 0.00105792123897    |
| 4     | 50   | 4           | 0.00110366677911    |

**Table 2. SGD and Adam.**

| Algorithm | Layer | Time | Eigenvalues | \( y_{\text{rmse}} \) |
|-----------|------|------|-------------|---------------------|
| Adam      | 3    | 50   | 4           | 0.001057921         |
| SGD       | 3    | 50   | 4           | 0.039886423         |
For the electric load data with time series characteristics and multiple influencing factors and different weights of different influencing factors, this paper proposes an improved LSTM model based on Adam algorithm and verified it through practice. The main findings of this paper are as follows:

For a variety of factors that affect the electrical load, multiple elements should be considered for analysis, where historical data of electrical load is the most important factor.

Based on the superiority of Adam algorithm, a short-term prediction model of electric load based on Adam-LSTM is established and tested separately to prove the key of model depth and gradient descent algorithm.

The method of this paper fully exploits the regularity information between power load data and temperature, day and night, etc., with good robustness and high precision.

References

[1] X. An, L. Zhang, X. Fan, X. Su, C. Zhang, F. Li Feng, Impact of New Energy Grid Connection on Power Quality of Power System, Modern Industrial Economy and Informationization. 2018,8(16),49-50+53.

[2] Y. Li, Q. Liu, Summary of Short-term Load Forecasting Methods of Electric Power Systems, Electric Switchgear. 2017,55(05), 5-9.

[3] B. Ji, Z. Wu, Application of Grey Prediction Method in Power System Load Forecasting, Technology and Economic Guide. 2018,26(29),42

[4] X. Liang, Y. Hu, T. Huang, D. Li, W. Gao, Y. Sun, Application of LSTM and ACO in Intelligent Electric Power Dispatching, Process Automation Instrumentation. 2018, 39(05),98-102.

[5] S.A. Zagrebina, V.G. Mokhov, V.I. Tsimbol, Electrical Energy Consumption Prediction is based on the Recurrent Neural Network, Procedia Computer Science. 2019,150.

[6] J. Liu, B. He, Y. Hu, Time Series Prediction System Based on TensorFlow Deep Learning, Modern Computer. 2019, (14), 26-29+33.

[7] S. Lu, H. Gao, J. Li, H. Zhang, Y. Hao, X. Zhang, Response of Citizens Power Consumption to Meteorological Factors and Its Forecast in Xi'an City, Journal of Arid Meteorology. 2017, 35(05),886-892.

[8] Y. Li, Z. Zhu, D. Kong, H. Han, Y. Zhao, EA-LSTM: Evolutionary attention-based LSTM for time series prediction, Knowledge-Based Systems. 2019-06-03.

[9] J. Lu, Q. Zhang, Z. Yang, M. Tu, J. Lu, H. Peng, Short-term Load Forecasting Method Based on CNN-LSTM Hybrid Neural Network Model, Automation of Electric Power Systems. 2019, 43(08), 131-137.

[10] R. Rajeev, J. Abdul Samath, N.K. Karthikeyan, An Intelligent Recurrent Neural Network with Long Short-Term Memory (LSTM) BASED Batch Normalization for Medical Image Denoising, Journal of Medical Systems. 2019-06-18.