Short-term load forecasting based on depth measurement information

Jian Xu¹², Gang Ma¹, Xin Shan², Yi Wang²
¹ Nanjing Normal University, Nanjing City, Jiangsu province, China
² NARI Technology Co., Ltd, Nanjing City, Jiangsu province, China
nnumg@njnu.edu.cn

Abstract. According to the predicted time span, the power load forecasting problem can be roughly divided into long-term and short-term forecasting. Among them, ultra-short-term load forecasting generally outputs load changes in the next few minutes to for on-line monitoring of power equipment, while short-term load forecasting mainly refers to pre-day load forecasting and pre-week load forecasting. It provides reference for hydropower dispatching, unit start and stop, water-fire coordination and so on, and is the basic work needed for the daily operation of power grid. In this paper, long and short memory networks network is used to predict short-term load, and depth measurement information is used as neural network input to construct mapping relationship based on load sequence timing.

1. Introduction
Power load forecasting is an important basis to ensure the stability and economic operation of power system. Load forecasting with different time span has different application purposes for power grid. Among them, accurate short-term load forecasting results can help power system staff to make reasonable production plans, maintain the balance of supply and demand and ensure the safety of power grid, while reducing the waste of resources and the cost of electricity. While the environmental factors of new energy harvesting such as photovoltaic and wind power have a great influence, which makes the source load forecasting problem more complex[1]. The traditional source load forecasting technology is not enough to perceive the "depth", and lacks the deep sensing approach to the internal characteristics of photovoltaic, wind power and load. At the same time, in the face of all kinds of massive data (such as multidimensional weather data, electricity data), traditional algorithms are difficult to construct effective mapping relationship. It is worth noting that with the continuous development of smart grid construction, load-related data in the power grid are becoming increasingly abundant and even can be said to enter the big data period. These accumulated massive data provide a big data basis for source load depth perception and accurate prediction. The interconnection between the systems makes it possible to use the data of the carriers, the dispatchers and the users. Cross-departmental, cross-domain system integration and data integration has become an important way to explore the competitiveness of power enterprises in the data age. The rapid development of the third generation artificial intelligence technology, such as deep learning, provides a new technical means for source load prediction. Advanced learning and other third generation artificial intelligence technology provide a new technical means for source load prediction[2].

At present, the computing-intensive load forecasting technology adapted to big data has gradually become the mainstream of research. Most of these methods need a lot of parallel computing and logical
judgment at run time, and occupy more CPU resources[3]. So it is called computing-intensive load forecasting technology. It mainly includes artificial neural network, support vector machine, wavelet transform algorithm and fuzzy regression model [4]. In this paper, ultrashort weather forecast will be introduced, and the prediction accuracy will be improved.

2. Low-voltage household variable topology identification

Long short-term memory is a deep learning structure with outstanding achievements in the field of natural language processing, which has been proved to be able to process time series related data well. Load data is a typical time series related data. hence, in this study, we also use LSTM algorithm to predict the load of different voltage levels.

In the traditional feedforward neural network, the input layer, the hidden layer and the output layer in the network are fully connected, but there is no connection between the nodes inside each layer. This structure makes the traditional feedforward neural network unable to deal with the problems of the correlation between the inputs. Unlike traditional feedforward neural networks, cyclic neural networks introduce directional loops. At this time, the nodes between hidden layers in the network are no longer connected but connected. The input of hidden layer includes not only the output of input layer but also the output of hidden layer at the previous time[5].Thus, the cyclic neural network is able to memorize the preceding information and apply it to the calculation of the current output. Compared with the traditional feedforward neural network, the cyclic neural network takes into account the time factor and can memorize the previously stored content.

Cyclic neural networks solve the problem of correlation before and after input information, and can connect previous information to the current task. However, in practical applications, it is found that when the time interval between the previous information and the current task increases, the cyclic neural network loses the ability to connect to such far information, hence, in order to solve this problem, long-term and short-term memory network (LSTM) was born.

For a given input time series \( x = (x_1, x_2, ..., x_T) \), a standard cyclic neural network calculates the hidden layer vector \( h = (h_1, h_2, ..., h_T) \) and output layer vector \( y = (y_1, y_2, ... , y_T) \) by iterative solution.

\[
\begin{align*}
    h_t &= S(W_{oh}x_t + W_{hh}h_{t-1} + b_h) \\
    y_t &= W_{hy}h_t + b_y
\end{align*}
\]

(1)  
(2)

The \( W_{oh}, W_{hh}, W_{hy} \) represents the weight coefficient of the network from the input layer to the hidden layer, the hidden layer inside, the hidden layer to the output layer, the \( b_h, b_y \) represents the deviation vector of the hidden layer and the output layer respectively, and the \( S(\cdot) \) use Sigmoid function as activation function in standard cyclic neural networks. LSTM, a short-term and short-term memory module is used to replace the simple hidden layer neurons in the standard circulatory neural network, so it has the ability to learn long-term information. LSTM is a special type of circulatory neural network, which mainly depends on the carefully designed "gate" structure to remove or increase the information to the state of the cell. Door is a method of making information selective, LSTM has three gates to protect and control the state of cells: input gate, output gate and forgetting gate.

Cell state neurons are equivalent to the neurons in the hidden layer of the standard circulating neural network. The excitation output \( c_t \) of the cell state neurons can be expressed as follows:

\[
c_t = f_{c_{t-1}} + i_t \tanh(W_{ox}x_t + W_{oh}h_{t-1} + b_i)
\]

(3)

\( c_{t-1} \) means the output of t-1 cell-state neurons; \( f_t \) and \( i_t \) represent the excitation output of the forget gate and the input gate respectively; \( W_{ox} \) represents the network weight coefficient from the network input layer to the current module input; \( W_{oh} \) represents the network output on t-1 module; \( b_i \) represents the current short-term and short-term record. The deviation variable corresponding to the memory module.

The input gate mainly determines which attributes and the contents of the new attributes are updated according to the last output \( h_{t-1} \) and the current input \( x_t \). Formula (4) gives the activation output \( i_t \) of the input gate.

\[
i_t = s(W_{ix}x_t + W_{ih}h_{t-1} + W_{io}c_{t-1} + b_i)
\]

(4)

\( s \) represents the Sigmoid function. Formula (4) gives the activation output \( i_t \) of the input gate.
The $S(\cdot)$ represents the logistic sigmoid function; $W_{xi}$ representation from network input layer to when the network weight coefficient of the front module input gate; $Whi$ represents the network weight coefficient output from the $t-1$ long-term and short-term memory module to the current input gate; $W_{ci}$ represents the network weight coefficient from the $t-1$ cell state neuron to the current input gate; $bi$ represents the deviation vector of the current input gate.

The output gate is mainly based on the previous output $h_{t-1}$ and the current input $x_t$ to determine what to output now. Formula (5) gives the excitation output $o_t$ of the output gate.

$$o_t = s(W_{xo}x_t + Who h_{t-1} + W_{co}c_{t-1} + b_o)$$

In the formula: $W_{xo}$ represents the network weight coefficient from the network input layer to the current module output gate; $Who$ represents the network weight coefficient output from the $t-1$ long and short term memory module to the current output gate; $W_{co}$ represents the network weight coefficient from the $t-1$ cell state neuron to the current output gate; $b_o$ represents the deviation vector of the current output gate.

The forgetting gate mainly determines whether to abandon the previous state content according to the last output $h_{t-1}$ and the current input $x_t$. Formula (6) gives the activation output $f_t$ of the forgetting gate.

$$f_t = s(W_{xf}x_t + Whf h_{t-1} + W_{cf}c_{t-1} + b_f)$$

In the formula: $W_{xf}$ represents the network weight coefficient from the network input layer to the current module forgetting gate; $Whf$ represents the network weight coefficient output from the $t-1$ long and short term memory module to the current forgotten gate; $W_{cf}$ represents the network weight coefficient from the $t-1$ cell state neuron to the current amnesia gate; $b_f$ represents the deviation vector of the current forgotten gate.

Finally, the output $ht$ of short-term and short-term memory module can be obtained from formula (7). The output of short-term memory module can be obtained.

$$h_t = o_t \tanh(c_t)$$

A variety of morphs have also been developed to enhance the information-memory capabilities of cyclic networks LSTM one of the most widely used variants. Besides, Bidirectional recurrent neural network (BRNN) is another structure with good effect.

We select 3-layer LSTM neural network and 1-layer fully connected layer network to minimize the mean variance, and use the "Adam" algorithm to perform random gradient descent to optimize parameters in-term and short-term memory networks. During training, the Epoch is 30 and the batch is 1.

3. Empirical research
This study selects the actual data of Jinnan Line of Jinshan Substation in Lianyungang 2017. This data contains multiplex table data. Table data one mainly measures urban and rural residents, table data two mainly measures light industry, table data three main measurement business, table data four main measurement public institutions, each table data and total load data are known.
3.1. Correlation analysis

Because of the actual situation, the general summary table data is known, and each sub-table data cannot be fully obtained. In order to verify whether the type of input variables of the prediction model has an impact on the accuracy of the final prediction results, the correlation analysis of the load data is carried out first. Pearson correlation coefficient is a measure of vector similarity. The output range from -1 to 1, 0 represents no correlation, negative value is negative correlation, positive value is positive correlation. The Pearson coefficient is expressed as follows:

\[ \rho_{X,Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{\sum_{i=1}^{n}(X-\mu_X)^2} \sqrt{\sum_{i=1}^{n}(Y-\mu_Y)^2}} \]  

The relevant factors involved in the calculation are shown in the following table:

| code | L1 | L2 | L3 | L4 | L5 | L6 | L7 | D |
|------|----|----|----|----|----|----|----|---|
| correlati | day load | Two day | One-week | The previous day | The previous day | The previous day | The previous day | Type of day |
| on factor | Table 1 | Table 2 | Table 3 | Table 4 | Load | Load | Load | |

Table 2. Correlation calculation results

| correlatio | L1 | L2 | L3 | L4 | L5 | L6 | L7 | D |
| factor     |    |    |    |    |    |    |    |   |
| Jinnan Line | 0.9817 | 0.9807 | 0.9783 | 0.9661 | 0.5761 | 0.9287 | 0.8353 | 0.7225 |

The correlation ranking can be concluded as follows:
(i) The selected correlation factors are all load-related and positive correlation.
(ii) The closer the historical day full-day load distance to the predicted day time, the higher the correlation. In this example, the correlation of the previous load series is higher than that of the two days and one week.
(iii) As shown in the following figure, table 1 data (urban and rural residents) accounts for the highest proportion, so it has the highest correlation with the overall load to be predicted. Table 2 data (light industry) accounts for the lowest proportion and the lowest correlation with the overall load to be predicted.
3.2. Effect of input variable selection on actual prediction

Ideally, each feeder table data is known and can be modeled and predicted separately for each category. But in practice, it is generally impossible to obtain the historical actual values of all categories. This report focuses on the influence of the number and type selection of historical actual values of various load components on the prediction results. The model uses the above LSTM neural network. The data set uses the actual data for the first quarter of 2017 of the Jinnan Line.

An mean absolute error percentage (MAPE) is used to evaluate the accuracy of the prediction. The expression is as follows:

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|x_i^\hat{} - x_i|}{x_i}
\]

(i) Influence of Number of Input Components on Prediction Accuracy

The following table shows the prediction accuracy and average accuracy of data of sub-table 1, the historical data of sub-table 1 and 2, and the prediction accuracy and average accuracy of three consecutive ten days under the historical data of input sub-table 1, 2 and 3.

| import         | Sub-table 1 | Sub-table 1, 2 | Sub-table 1, 2, 3 | NonSub-table |
|----------------|-------------|----------------|------------------|--------------|
| Mape_ave1      | 7.678414    | 7.821646       | 7.818141         | 7.91565      |
| Mape_ave2      | 8.697911    | 8.685814       | 8.690775         | 8.829966     |
| Mape_ave3      | 10.81135    | 10.653324      | 10.51323         | 11.05675     |

Take the forecast result of April 8 as an example. The more data information is input, the higher the accuracy of prediction.

Figure 2. Impact of input of different number of sub-table data on prediction results

It can be observed from figure 2 and table 3 that when the number of input sub-table data increases, the prediction accuracy increases. Compared with the data input without sub-table, adding sub-table data input can improve the accuracy.

(ii) Influence of Data Scale of Input Sub-table on Prediction Accuracy

According to the above correlation analysis, the higher the proportion of a sub-table data to the total load, the stronger the correlation with the total load to be predicted. Among the four tables, table 1 data (urban and rural residents) accounted for the highest proportion, and table 2 data (light industry) accounted for the lowest proportion. These can be added to the model separately compare the table data. Table 4 shows the forecast results in early April. It can be seen from the table that the higher the proportion of known sub-table data, the higher the accuracy of the prediction and the lower the average absolute error percentage. Compared with the input of non-score table data, even the smaller input table data can significantly improve the prediction accuracy.
Table 4. Effects of Different Types of Input Historical Load Components on Prediction Results

| import | First component | Second component | Non component |
|--------|-----------------|------------------|--------------|
| Mape ave | 7.678414 | 7.81299 | 7.915648 |

Figure 3. Effects of Different Types of Input Historical Load Components on Prediction Results

4. Conclusion

This paper mainly analyzes the correlation between the historical data of each classified load measured in depth and the total load to be predicted. By designing different prediction models and different inputs, the influence of the number of load components on the accuracy of total load forecasting and the influence of different load types on total load forecasting are verified respectively. Finally, it is concluded that increasing the historical data input of load component is helpful to improve the accuracy of load forecasting, but the more the number of types increased, the higher the accuracy of forecasting. The increase of single class load with large proportion of total load can improve the prediction accuracy. The new paradigm of load forecasting is proved, which provides a technical scheme for the prediction of voltage meter data to assist the prediction of high voltage power data, and verifies its technical feasibility.

References

[1] Zhiliang Chen. (2020) Research on Geotechnical Investigation of Deep Foundation Pit Engineering [J]. New Technologies and New Products in China, No.414(08):132-134.
[2] Jiuyang Tao, Lin Wu, Xiaofeng Hu. (2016) Analysis of Alphago Technology Principle and Prospect of Military Application of Artificial Intelligence [J]. Journal of Command and Control, 2(2), 114-120.
[3] Yuyu Chen. (2017) Research on Short-term Power Load Prediction Based on Artificial Neural Network. Zhejiang University.
[4] Baoyan Zhang. (2020) Research on the Application of Image Recognition Based on Deep Learning Model[J]. Shanxi Electronic Technology, No.213(06), 89-91+95.
[5] Jiajia Yang, Guolong Liu, Junhua Zhao. (2018) Short-term forecasting method of industrial load using long and short-term memory deep learning model [J]. power construction, 39(10):20-27.
[6] Zhitong Su, Chunlei Wang. (2020) A Quality Prediction Model Based on Istm Roasting Timing Data. Software, v.41;No.481(05), 113-115+205.