Research on Short-term Power Load Forecasting Based on Bi-GRU

Huan He¹, Haomiao Wang², Hongliang Ma³*, Xuesong Liu⁴, Yilin Jia¹ and Gangjun Gong¹

¹ State Grid Anshan Electric Power Supply Company, Anshan, 114009, China
² State Grid Liaoning Province Electric Power Company Limited, Shenyang, 110006, China
³ Beijing Engineering Research Center of Energy Electric Power Information Security, North China Electric Power University, Changping District, Beijing 102206, China
⁴ Liaoning Power Energy Development Group Co.Ltd, Shenyang,110006,China

* Corresponding author’s e-mail: mmahongliang@163.com

Abstract: The precise, accurate and efficient short-term load forecasting can guide the power supply companies to rationally arrange power dispatch plans, help improve the stability of grid operation, and significantly, improving power utilization, thereby optimizing corporate marketing strategies and increasing corporate economic returns. Short-term load forecasting methods based on deep learning have gained greater attention from academia and power companies. Among them, the load forecasting model based on recurrent neural network has gained excellent forecasting results compared with traditional machine learning models. The advantage of the cyclic neural network is that it can extract the degree of relevance of the data in the time dimension, but the unidirectional network only considers the impact of historical data on the current forecast. This paper proposes a load forecasting model based on a bidirectional gated recurrent unit, which further improves the relevance of data. However, it introduces meteorological factors and the influence of holidays to improve the accuracy of forecast results. Taking the power load data of a district of a city in China as an example, the prediction results of this model meet the actual requirements and show better prediction performance over Bi-LSTM, LSTM, GRU, and other models.

1. Introduction

With the new power generation era, power companies are gradually transforming their power production methods moving toward green energy and non-conventional methods. The installed capacity of renewable energy power generation has increased gradually with time. It is expected that China's renewable energy power generation will account for more than 30% of the total power generation by 2030 [1]. However, the consumption of renewable energy power is a major problem faced by power supply companies. With the development of society and the economy, the power grid has undergone great changes in scale and complexity. The power user side has an increasing impact on the safety and stability of the grid operation. At the same time, as more and more renewable energy sources are connected to the grid, the risk of grid operation increases. Power load forecasting research has important...
practical significance for maintaining the balance between the power supply side and the power user side and ensuring the safe and stable operation of the power grid. Load forecasting research provides scientific guidance for the power dispatch of power supply companies, and also helps to further understand the power consumption rules of power users, optimize the marketing strategies of power supply companies, and create greater economic benefits.

Power load forecasting involves research on multiple time scales, including ultra-short-term load forecasting, short-term load forecasting, mid-term load forecasting, and long-term load forecasting [2]. Super short-term load forecasting is a forecast of the hourly or every minute electric load of power users, short-term load forecasting is a forecast of the daily electric load of electric users, medium-term load forecasting is a forecast of the weekly electric load of electric users, and long-term load forecasting is forecast of the electricity load of electricity users on a monthly basis and above. Compared with the above load forecasting studies on different time scales, short-term load forecasting has more practical application value.

Artificial intelligence technology, especially deep learning, has developed rapidly since Hinton et al. made great research progress in 2006 [3]. In image processing, natural language processing, computer vision, and other fields, deep learning has revealed its excellent performance. The image recognition method based on the convolutional neural network even exceeds the accuracy of human recognition. At present, China is gradually advancing the construction of smart grids. The application of a large number of smart terminals and sensors enables the grid to have "sensing capabilities". The electricity consumption information collection system built on this basis collects and stores massive amounts of electricity data. It is the most important original data for load forecasting research. Short-term load forecasting based on deep learning is a research hotspot in academia, and there are many related research results. Tan Mong et al. proposed a hybrid ensemble learning prediction model based on long short-term memory (LSTM) network [4], which extracts deep features from multivariate data, and proposes a new loss function based on the deviation-variance compromise scheme, This function integrates the peak demand forecast error. X. Tang et al. proposed a hybrid neural network prediction model based on the deep belief network (DBN) and bidirectional recurrent neural network (Bi-RNN)[5]. Simulation experiments show that this method can effectively improve the accuracy of load forecasting. In this paper, Y. Ma et al. proposed a short-term load forecasting method based on isolation forest (iForest) and LSTM[6], which can effectively reduce the impact of data measurement errors on load forecasting results. Zhao Bing and others proposed a CNN-GRU short-term power load forecasting method based on attention mechanism [7], selecting a convolutional neural network (CNN) and gated recurrent units for feature extraction and data prediction, and introducing Attention Mechanism to retain the deep important information of the data. Through the above introduction, it can be found that Recurrent Neural Network (RNN) is widely used in short-term load forecasting research. This is because RNN considers the connection of data in the time dimension, and power load data has a close time correlation. However, the one-way RNN network only considers the impact of historical data on the current prediction results. Fatma Yaprakdal et al. proposed a deep recurrent neural network model DRNN Bi-LSTM based on bidirectional long-term short-term memory unit (Bi-LSTM)[8]. Compared with the one-way RNN, the prediction accuracy of the method has been improved to a certain extent.

To further improve the accuracy of short-term load forecasting, this paper proposes a short-term load forecasting model based on Bi-GRU and residual correction. This method uses Bi-GRU for load data forecasting, takes meteorological factors, day types, and historical power load data as input to the network to obtain power load data for the next few days. The model aims to improve the correlation of the data in the front and back directions during the prediction process and minimize the model parameters to improve the accuracy of power load prediction. Finally, the simulation experiment proves that the short-term load forecasting method proposed in this paper has superior performance compared with other methods.
2. Principle of Bi-GRU

2.1. Principle of GRU

RNN is generally used for modelling and analysis of time series data. Its characteristic usually considers the impact of historical states on the current state[9]. Compared with other deep learning algorithms, RNN shows better grasp of global data and is often used for classification and prediction of text data. In this paper, RNN refers specifically to the traditional recurrent neural network. Figure 1 demonstrates an expanded schematic diagram of RNN. RNN and artificial neural networks (ANN) have certain similarities. The difference is that RNN has a hidden state, which also makes historical input and current output having internal connections.

As the network deepens, the RNN model will have the problem of gradient disappearance or gradient explosion. The emergence of LSTM solves this problem. As shown in Figure 2, the LSTM neuron introduces the gate structure and cell state and retains the hidden state in the RNN. The three gate structures are input gate, forget gate, and output gate. The LSTM model has greatly improved the forecasting ability of time series, but the network has become very complicated, and many network parameters need to be trained.

Cho et al. merged the gating unit of LSTM and gave birth to the Gated Recurrent Unit (GRU). The parameters that need to be learned during the training of the GRU model are greatly reduced, and the speed of data processing is significantly improved. As shown in Figure 3, the hidden layer state $h$ in the GRU model is obtained by combining the cell state and the hidden layer state in the LSTM model. There
are only update gates and reset gates in the model. In Figure 3, $x_t$ is the input at time $t$, $y_t$ is the output at time $t$, $h_{t-1}$ is the hidden state at the previous time, $h_t$ is the currently updated hidden state, $\tilde{h}_t$ is the candidate hidden state, $\odot$ represents vector splicing, $\otimes$ represents vector bitwise multiplication, $z_t$ and $r_t$ represents the update gate respectively and reset gate, $\sigma$ is the activation function. The mathematical description of GRU forward propagation is:

$$z_t = \sigma(W_z[h_{t-1}, x_t])$$  \hspace{1cm} (1)

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$ \hspace{1cm} (2)

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t])$$ \hspace{1cm} (3)

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t$$ \hspace{1cm} (4)

$$y_t = \sigma(W_o \cdot h_t)$$ \hspace{1cm} (5)

In the above formula, the parameters $W_z, W_r, W_h, W_o$ are learned by the model through the training data set, and $[a, b]$ represents the splicing of two vectors, so the corresponding forms of the $W_z, W_r, W_h$ are as follows:

$$W = W_{zz} + W_{zb}$$ \hspace{1cm} (6)

$$W_r = W_{rz} + W_{rb}$$ \hspace{1cm} (7)

$$W_h = W_{hz} + W_{bh}$$ \hspace{1cm} (8)

2.2. Bi-GRU

The process of GRU processing data is carried out in the order of the timestamp of the data. The output at the current moment is not only related to the input at the current moment but also related to the historical input. Bi-GRU is a stacking model of GRU. As shown in Figure 4, the output at the current moment in the model is obtained from the hidden state of the two positions at the previous moment and the next moment. The mathematical description of the model is:

$$\tilde{h}_t = \text{GRU}(x_t, h_{t-1})$$ \hspace{1cm} (9)

$$\tilde{h}_t = \text{GRU}(x_{t-1}, h_{t-1})$$ \hspace{1cm} (10)

$$h_t = \alpha \tilde{h}_t + \beta h_{t-1} + b$$ \hspace{1cm} (11)
In the above formula, $\tilde{h}_t$ and $\tilde{h}_t$ respectively represent the state of the hidden layer in the positive and negative directions at time $t$, $\alpha$ and $\beta$ are the weights in the positive and negative directions, $b$ is the offset and $W_b$ is the weight matrix.

3. Load forecasting model based on Bi-GRU

3.1. Model building
The simulation is performed in Python environment, and the relevant deep learning algorithm model is built under the Keras framework. The number of neurons in Bi-GRU is set to 100.

3.2. Model evaluation criteria
The root mean square error (RMSE) is used as the evaluation index of model prediction performance, and the RMSE is mathematically described by equation 12.

$$E_{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}$$

(13)

Where, $n$ is the number of forecast days, $y_i$ is the actual load value on the $i$ day, and $\tilde{y}_i$ is the load forecast value on the $i$ day.

4. Simulation

4.1. Data
The load data used in this article comes from a district of a city in China, including 689 days of daily load, maximum temperature, minimum temperature, day type, the average load of the previous 7-11 days, etc. Figure 6 is the daily load curve in this area, the time range is from 2018-6-1 to 2020-4-19. From the figure, it can be seen that the electrical load changes significantly with the seasons, reflecting the difference between temperature and electrical load close relationship.

![Power load curve](image)

Figure 5. Power load curve.

4.2. Data preprocessing

4.2.1. Missing value filling
There are many processing methods for default values, and different processing methods are selected according to different actual problems. Common ones include filling the mean, mode, and median. Since
the power measurement parameters have time continuity, and there is often no short-term numerical jump at the time of non-fault, this paper will fill the default value at time $t$ with the value at time $t-1$.

### 4.2.2. Normalization processing

The normalization process is to reduce the influence of the magnitude difference of the data value itself on the simulation results. Controlling the data value within a certain interval is also beneficial to speed up the simulation operation. The mathematical description of the normalization is:

$$\tilde{x}_n^{(i)} = \frac{x_n^{(i)} - x_{\text{min}}^{(i)}}{x_{\text{max}}^{(i)} - x_{\text{min}}^{(i)}}$$ (14)

In the above formula, $x_n^{(i)}$ is the nth value of the i-th sample, $x_{\text{max}}^{(i)}$ is the maximum value in the i-th sample, $x_{\text{min}}^{(i)}$ is the minimum value in the i-th sample, and $\tilde{x}_n^{(i)}$ is the normalized data value.

### 4.3. Simulation results

Observing the result in Figure 6, as the number of iterations increases, the loss of the model decreases gradually. In the process of training the model obtains loss consistency at around 200 iterations, therefore, the number of iterations is set to 200 during the simulation.

Figure 6. Model loss.  

![Figure 6. Model loss.](image)

Figure 7 is a distribution diagram of model prediction data and real data. From the figure, It can be seen that the fitting effect between the two sets of data is better.

To compare the prediction performance of the Bi-GRU-based short-term load forecasting model proposed in this paper with other models, Bi-LSTM, GRU, and LSTM were also trained simultaneously with same data. Table 1 shows the comparison of the root mean square error between the models. Bi-GRU is used for power load forecasting to achieve the smallest error.

|                      | Bi-GRU | Bi-LSTM | GRU  | LSTM |
|----------------------|--------|---------|------|------|
| RMSE                 | 1.989  | 2.093   | 2.076| 2.241|

### 5. Conclusion

Aiming at the problems of unidirectional recurrent neural network in power load forecasting, this paper proposes a short-term power load forecasting model based on Bi-GRU and introduces the influence of
meteorological factors and daily types in the forecasting process. The simulation results proves that the prediction model proposed in this paper shows better performance with certain application value.

References

[1] Zhang T, Gu J. Markov short-term load forecasting method for high-proportion renewable energy power systems [J]. Power System Technology, 2018, 42(04): 1071-1078.

[2] Feng J. Research on short-term power load forecasting model based on deep learning [D]. University of Electronic Science and Technology of China, 2020.

[3] G.E. Hinton, S. Osindero and Y. Teh, "A Fast Learning Algorithm for Deep Belief Nets," in Neural Computation, vol. 18, no. 7, pp. 1527-1554, July 2006, doi: 10.1162/neco.2006.18.7.1527.

[4] M. Tan, S. Yuan, S. Li, Y. Su, H. Li and F. He, "Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning," in IEEE Transactions on Power Systems, vol. 35, no. 4, pp. 2937-2948, July 2020, doi: 10.1109/TPWRS.2019.2963109.

[5] X. Tang, Y. Dai, Q. Liu, X. Dang and J. Xu, "Application of Bidirectional Recurrent Neural Network Combined With Deep Belief Network in Short-Term Load Forecasting," in IEEE Access, vol. 7, pp. 160660-160670, 2019, doi: 10.1109/ACCESS.2019.2950957.

[6] Y. Ma, Q. Zhang, J. Ding, Q. Wang and J. Ma, "Short Term Load Forecasting Based on iForest-LSTM," 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 2019, pp. 2278-2282, doi: 10.1109/ICIEA.2019.8833755.

[7] B Zhao, Zengping W, Weijia J, Xin G, Xiaobing L. CNN-GRU short-term power load forecasting method based on attention mechanism[J]. Power System Technology, 2019, 43(12): 4370-4376.

[8] Fatma Yaprakdal,M. Berkay Yilmaz,Mustafa Baysal,Amjad Anvari-Moghaddam. A Deep Neural Network-Assisted Approach to Enhance Short-Term Optimal Operational Scheduling of a Microgrid[J]. Sustainability,2020,12(4).

[9] Gong, G.; An, X.; Mahato, N.K.; Sun, S.; Chen, S.; Wen, Y. Research on Short-Term Load Prediction Based on Seq2seq Model. Energies 2019, 12, 3199.

[10] R. Fu, Z. Zhang and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 2016, pp. 324-328, doi: 10.1109/YAC.2016.7804912.

[11] L. Kuan et al., "Short-term electricity load forecasting method based on multilayered self-normalizing GRU network," in 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, 2017, pp. 1-5, doi: 10.1109/EI2.2017.8245330.