INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 19, Issue: 5, Month: October, Year: 2024
Article Number: 6631, https://doi.org/10.15837/ijccc.2024.5.6631

CCC Publications

Ultra-short-term Load Forecasting Based on XGBoost-BiGRU

Shuyi Chen, Guo Li, Kaixuan Chang, Xiang Hu, Peiqi Li, Yujue Wang, and Yantao Zhang

Shuyi Chen
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
csy_0730@163.com

Guo Li*
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
*Corresponding author: liguo_sg@126.com

Kaixuan Chang
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
853532107@qq.com

Xiang Hu
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
robin_rhu@126.com

Peiqi Li
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
peiqidev@163.com

Yujue Wang
State Grid Commercial Big Data Co., Ltd.
42 Donggexinli, Yongwai, Dongcheng District, Beijing, China
358639599@qq.com

Abstract

High-precision load forecasting serves as the foundation for power grid scheduling planning and safe economic operation. In scenarios where only historical power load data is available without other external information, fully exploiting meaningful features from the temporal load sequence is crucial for improving the accuracy of load forecasting. Therefore, an ultra-short-term load forecasting method that combines eXtreme gradient boosting (XGBoost) and bidirectional gated recurrent unit (BiGRU) is proposed in this paper. Considering various factors that affect loads, a candidate feature set is established, which includes temporal information and historical loads. XGBoost is used to select the features that contribute significantly to load forecasting, forming an optimal feature set. These optimal features are then used as inputs to the BiGRU, and the bayesian optimization algorithm is applied to optimize the network hyperparameters. Then the load forecasting model for the next 15 minutes based on BiGRU is generated by training iteratively. The proposed XGBoost-BiGRU method is validated on real load data from a province in China. Experimental results demonstrate that the method can effectively avoid the impact of redundant features, improving both prediction accuracy and efficiency. The research has significant importance for guiding real-time supply-demand balance calculations and scheduling in power grids.

Keywords: load forecasting, eXtreme gradient boosting, bidirectional gated recurrent unit, feature selection.

1 Introduction

With the increasing proportion of new energy integration and the improvement of end-user electrification levels, the traditional power system is transforming towards a new one. The significant challenges are posed to grid operation because of the uncertainty of new energy sources and the diversification of energy consumption patterns [1]. In the new power system, exploring the potential of various resources to achieve efficient coordination among generation, grid, load, and storage is crucial for addressing the supply-demand balance issues. As an important technology on the demand side, load forecasting is the basis for the coordinated scheduling of generation, grid, load, and storage. The load forecasting with high-precision helps assess the supply-demand balance accurately. Then, corresponding scheduling strategies for various resources are formulated during supply-demand warning periods. Therefore, high-precision load forecasting helps ensure the safe and stable operation of the power system [2].

Load forecasting is the process for predicting future power consumption reasonably by exploring the inherent relationship between load and various influencing factors [3]. It can be divided into ultra-short-term, short-term, medium-term, and long-term load forecasting based on the time span. Ultra-short-term load forecasting refers to predicting loads within the next 5 minutes to 1 hour. The effectiveness of real-time grid dispatching strategies are determined by the accuracy of ultra-short-term load forecasting directly [4]. Ultra-short-term load forecasting has always been a hot research topic for experts and scholars worldwide. Various forecasting methods have been proposed, and are mainly divided into two categories: statistical methods and artificial intelligence-based methods [4, 5]. Common statistical methods include time series analysis [6] and regression analysis [7]. These methods have simple principles and fast computation speeds, but have limitations in analyzing and processing nonlinear data.

With the continuous development of artificial intelligence, machine learning algorithms suitable for handling nonlinear problems have been applied to execute load forecasting. Machine learning algorithms based on decision trees, support vector machines, etc., have achieved competitive performance in load forecasting. The eXtreme gradient boosting (XGBoost) algorithm has been used in peak load forecasting of distribution network lines in [8]. And the algorithm has also been used in short-term load forecasting with fewer feature dimensions in [9]. It performs good in prediction accuracy and computation speed. Recently, deep learning develops rapidly providing new solutions for solving load forecasting problems. Algorithms such as convolutional neural networks (CNN), long short-term memory networks (LSTM), and their variants have been widely applied to load forecasting tasks. These algorithms succeed in nonlinear feature extraction, and thus improve prediction accuracy effectively [10]. The method based on adaptive spatial-temporal graph convolutional network for short-term load prediction of multiple public charging stations is proposed in [11]. The influence of data with different time scales on prediction results is considered and LSTM is used for medium and long-term load forecasting in [12]. Load and temperature from previous 24 hours are used to establish the load forecasting model based on clustering and LSTM. Additionally, the load prediction model based on CNN- bidirectional LSTM (BiLSTM) performs well and utilizes feature data sufficiently.

Meanwhile, experts and scholars have focused on the impact of features on the accuracy of load forecasting, and researched on feature selection and mining. The ultra-short-term load forecasting method based on the double-layer XGBoost algorithm is proposed in [3]. The first layer of the method selects features and the second layer execute load forecasting. The performance of various prediction models combined with feature selection is explored. It ultimately proved that XGBoost has superior feature selection efficiency, and LSTM outperforms in prediction.

The energy consumption behavior of users is linked to economic, social, and environmental factors intricately. The diversified and flexible energy consumption patterns in new power systems have increased the difficulty of load forecasting [5, 12]. In the absence of external information other than historical load data, extracting the inherent features of time-series load data is crucial for improving prediction accuracy. Therefore, we propose an ultra-short-term load forecasting method based on the combination of XGBoost and bidirectional gated recurrent unit (BiGRU) in this paper. The method utilizes the meaningful time information and historical load embedded in time-series load data to predict power load for the next 15 minutes. The proposed method avoids redundant feature impacts, ensures load forecasting accuracy, and improves prediction efficiency. It also guides real-time grid supply-demand balance calculation and dispatch. The innovations of this paper are as follows:

- Based on the factors influencing power load, an ultra-short-term load forecasting feature set containing time information and historical load is constructed with only historical load data available.
- A feature selection method based on XGBoost is proposed. Features contributing significantly to load forecasting are selected from the candidate feature set to construct an optimal feature set according to the importance.
- A load forecasting method based on BiGRU is proposed. The Bayesian optimization (BO) algorithm is employed to optimize network hyperparameters with the optimal features as inputs.

The paper is organized as follows: Section 2 introduces the proposed ultra-short-term load forecasting method based on XGBoost-BiGRU, followed by the experimental setup clarified in Section 3; in Section 4, the experimental results are demonstrated with detailed discussion; the conclusion are presented in Section 5.

2 Load Forecasting Based on XGBoost-BiGRU

2.1 The Framework of the Proposed Method

The temporal characteristics of load need to be explored to improve the accuracy of prediction when only historical load data are available. The time information and historical load that contribute significantly to prediction are selected to construct the optimal feature set. It can avoid the interference of redundant features, and improve the precision and efficiency of the prediction model (Figure 1). The historical load data is firstly preprocessed by detecting and removing outliers, filling in missing values, and normalizing. Then XGBoost is used to rank the importance of time and historical load features. The features with high importance are selected to construct the optimal feature set. Additionally, the optimal features are used as input to train the load forecasting model based on BiGRU iteratively, and the hyperparameters are optimized by BO. The trained neural network model is finally used to predict power load and then the results are obtained.

2.2 Data Preprocessing

The collected historical load often suffers from data missing and errors, impacting the precision of prediction models directly. Therefore, it is essential to preprocess the raw load data by cleaning and handling outliers and missing values. To ensure the effectiveness of model training, the load data input into the neural network are normalized within the [0,1] range.

(1) Processing for abnormal data: For load data that significantly deviate from the normal range, the boxplot method is used to detect outliers. The first quartile of the load samples is Q_1, the third quartile is Q_3, and the interquartile range Q_R are defined as follows:

$$Q_R = Q_3 - Q_1,$$

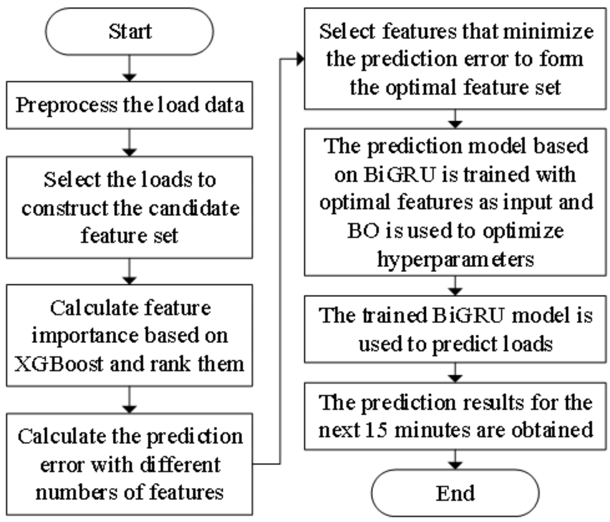


Figure 1: Flow chart of the proposed XGBoost-BiGRU method

Data outside the range of $[Q_1 - 1.5Q_R, Q_3 + 1.5Q_R]$ are identified as outliers and removed.

(2) Processing for missing data: For load data with single-point missing, based on the continuity of power load, the average of the data before and after the missing point is used to fill in the gap. For continuous missing load data, regression analysis is employed to fit the missing data.

(3) Data normalization: The load data input into BiGRU are normalized as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is raw load data.x_{max} and x_{min} refer to the maximum and minimum value of raw load data, respectively. x' is the normalized load data.

2.3 Feature Selection Based on XGBoost

The XGBoost algorithm is an ensemble learning algorithm composed of multiple classification and regression trees as base learners. The key to XGBoost is adding new base learners iteratively to correct the prediction errors from the previous iteration, gradually approaching the ground truth. XGBoost improves upon the principles of gradient boosting decision trees by performing a second-order Taylor expansion on the loss function and introducing regularization terms, effectively avoiding overfitting of the model.

In this paper, XGBoost is used to measure the contribution of features to load forecasting. By calculating the average gain of each feature during splitting, the importance of features is ranked and the optimal features are selected. For the task of ultra-short-term load forecasting with only historical load, the sampling interval is set to 15 minutes, and 96 points are collected per day. The power load value after 15 minutes is predicted based on historical load data. Initially, a feature set including two categories, temporal information and historical load, is established. The feature set are outlined (see Table 1).

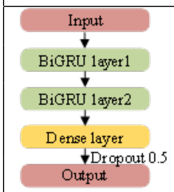
The temporal information within the feature set comprises the year, month, day of the year, week of the year, season of the year, sampling point of the day for the load, and whether the load is on a weekday. Historical load data are selected as candidate features, specifically including loads within two hours before the target time and loads at the same time within the previous seven days.

XGBoost is employed to calculate the importance of each feature, and the features are ranked in descending order of their importance. And then the corresponding error in load prediction is calculated with the number of features increasing gradually, starting from the most important one. The optimal feature set is determined by selecting the features that correspond to the minimum prediction error.

Table 1: Feature Set

Feature number	Feature name
F1	Year
F2	Month
F3	Day
F4	Week
F5	Season
F6	Sampling point
F7	Weekday
F8	Load 15 minutes ago
F9	Load 30 minutes ago
F10	Load 45 minutes ago
F11	Load 60 minutes ago
F12	Load 75 minutes ago
F13	Load 90 minutes ago
F14	Load 105 minutes ago
F15	Load 120 minutes ago
F16	Load at the same time 1 day ago
F17	Load at the same time 2 days ago
F18	Load at the same time 3 days ago
F19	Load at the same time 4 days ago
F20	Load at the same time 5 days ago
F21	Load at the same time 6 days ago
F22	Load at the same time 7 days ago

Table 2: Network Specifications for Bigru

Architectures	Network specifications		
	Optimal feature set		
	Size: 64	Merge: concat	Activation: ReLU
	Size: 64	Merge: concat	Activation: ReLU
	Units: 32	Activation: ReLU	
	Units: 1	Activation: Linear	

2.4 Load Forecasting Based on BiGRU

As a variant of LSTM, GRU similarly addresses vanishing and exploding gradients in long-term memory and backpropagation. GRU has a simpler structure while performs comparable to LSTM. By merging the input and forget gates into an update gate, the original three gates in LSTM are replaced with only a reset gate and an update gate. Therefore, GRU is able to retain memory of long-term sequences with fewer parameters and faster convergence. It results in higher computational efficiency and lower memory requirements for load forecasting tasks. BiGRU consists of a forward GRU and a backward GRU, specifically, it is trained in both forward and backward directions to capture temporal information from past and future of the input.

The optimal features selected by XGBoost are used as the input feature vectors for BiGRU, and the network outputs the target load values. Additionally, BO is employed to tune the hyperparameters of the network, iteratively finding the optimal combination of model parameters. Finally, the feature vectors of the target load are input to the trained model, and the network outputs the predicted load results. The architecture of BiGRU is demonstrated (see Table 2).

3 Experimental Setup

To validate the performance of the proposed method in this paper, BiGRU without feature selection, the double-layer XGBoost method for load forecasting proposed in [3], and the XGBoost-LSTM method proposed are selected for benchmarking. In the prediction stage, the input features of XGBoost-BiGRU, double-layer XGBoost, and XGBoost-LSTM are the optimal features selected by the method in this paper.

3.1 Experimental Data

To validate the performance of the proposed method, the load data of the entire society from a certain province in China for 2021 and 2022 are selected as experimental data. The load data samples at 15 minutes, resulting in 96 sampling points per day and a total of 70,080 data points. In the experiments, the training set is 80% of the whole data, while the 20% left is used as testing data. Specifically, data from January 1st, 2021 to August 9th, 2022 are used as the training data and the testing data are from August 10th, 2022, to December 31st, 2022. The training set is used to train the load forecasting model, while the testing set is to validate the performance of the model.

3.2 Experimental Platform

The experiments in this paper are implemented on the PyCharm platform using Python 3.7 for programming implementation. The BiGRU models in the experiments are implemented based on the TensorFlow framework. Networks are trained by ADAM optimizer algorithm, with a learning rate set to 0.001.

3.3 Evaluation Metrics

To evaluate the performance of the method proposed in this paper, three metrics are selected to assess the prediction results quantitatively.

The mean absolute percentage error (MAPE) is a commonly used evaluation metric in regression problems, which measures the ratio of the absolute difference between the prediction and the ground truth to the actual value.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \overline{y}_t|}{y_t}$$

where y_t and $\overline{y_t}$ are actual load values and prediction at each time index t, respectively, and n is the total sample number.

The mean absolute error (MAE) is used to evaluate the average absolute errors between the predicted signal and the ground truth.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \overline{y}_t|$$

The root mean square error (RMSE) calculates the square root for the average of the squared differences between the prediction and the ground truth for each sample.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}|y_t - \overline{y}_t|^2}$$

All the aforementioned metrics can measure the error in load forecasting, which indicates the lower the better.

4 Experimental Result

In this section, the performance of the proposed XGBoost-BiGRU is demonstrated through experimental results, benchmarked with BiGRU without feature selection, the double-layer XGBoost method proposed in [3], and the XGBoost-LSTM method proposed. Four methods are employed to execute ultra-short-term load forecasting on data from August 10th, 2022 to December 31st, 2022. The future power load for the next 15 minutes is predicted based on historical data. Initially, feature selection and analysis of XGBoost are presented. Subsequently, the prediction results of four methods are evaluated in three metrics. Finally, a comparison of the predicted curves of four methods and the ground truth is shown.

4.1 Feature Selection and Analysis

As described in Section 2.3, the candidate feature set comprises 22 features, including temporal information and historical load data. XGBoost is utilized to calculate the importance of each feature in load forecasting. The importance values and ranking of these 22 features are demonstrated (Figure 2).

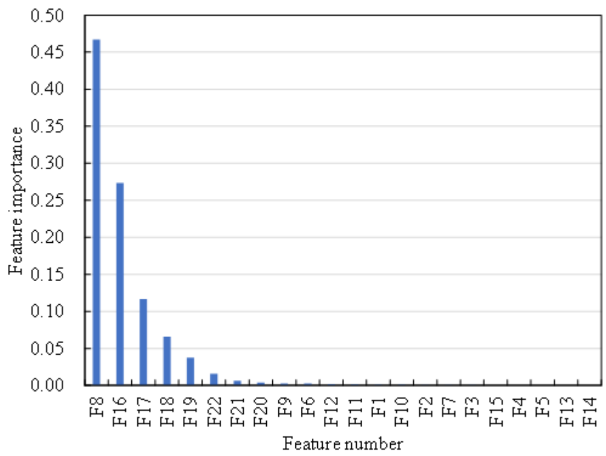


Figure 2: Ranking of feature importance

It can be observed that too few features are insufficient to comprehensively map the relationship between historical and predicted load, resulting in lower prediction accuracy. On the other hand, redundant features also have a negative impact on the accuracy of load forecasting. Among the candidate features, the importance of F8 is 46.70%, which is the highest one. It indicates that the correlation between the power load at the target time and the previous moment is strong. The second most important feature is the load at the same time one day before the target time, with an importance of 27.34%. Closely following are the loads at the same time 2 to 4 days before the target time, each with an importance exceeding 3%. It suggests that power consumption of users is closely related to their usage habits, and the consumption patterns tend to be similar at the same time each day.

The load for a continuous seven-day period has a regular pattern in daily power consumption (Figure 3). The load curves of different days have similar trends with same time for peaks and valleys. Therefore, the load values at the same time on previous days contribute significantly to the prediction results. However, it can be seen that seasonal factors, as well as loads 90 minutes and 105 minutes ago, have no importance in the prediction. The season affects the difference between daily peak and valley loads and their occurring time, but does not contribute to predicting the load values at future moments (Figure 4). As for loads 90 minutes and 105 minutes ago, they are too far from the target time, resulting in little correlation.

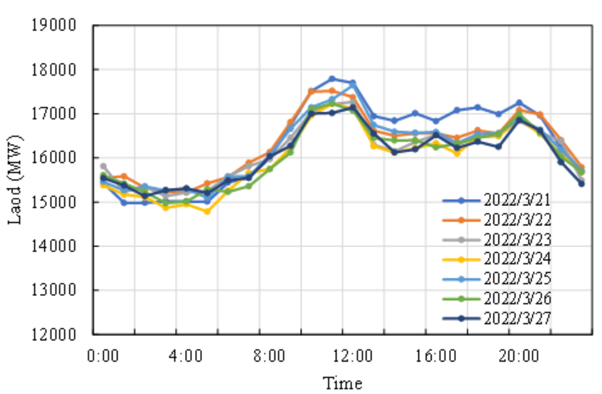


Figure 3: Daily load characteristic curves

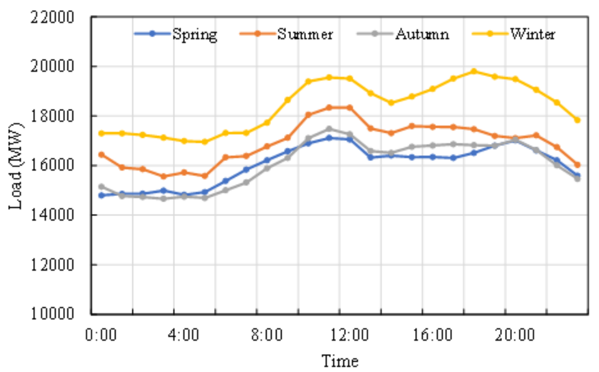


Figure 4: Quarterly load characteristic curves

Table 3: Load Forecasting Results in Evaluation Metrics

Methods	MAPE	MAE(MW)	RMSE(MW)
BiGRU	2.90%	87.20	125.36
XGBoost-BiGRU	1.09%	42.38	77.91
double-layer XGBoost	3.53%	100.68	151.24
BiGRU	2.11%	77.10	112.51

To determine the number of features in optimal feature set, the candidate features are sorted according to their importance. Starting from the most important one, additional features are gradually added, and the RMSE is calculated for different numbers of features (Figure 5).

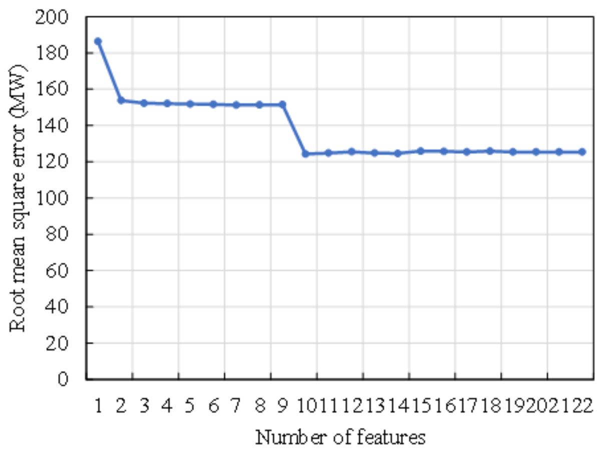


Figure 5: Number of features and corresponding RMSE for prediction

From the above analysis, it can be observed that the RMSE is minimized when the number of features in the optimal feature set is 10. Therefore, it is finally determined that the optimal feature set includes the top 10 most important features, namely load 15 minutes ago, load at the same time 1 day ago, load at the same time 2 days ago, load at the same time 3 days ago, load at the same time 4 days ago, load at the same time 7 days ago, load at the same time 6 days ago, load at the same time 5 days ago, load 30 minutes ago, and sampling point of the day.

4.2 Load Forecasting Performance Metric Comparison

The load prediction results of the proposed method and the benchmarks are evaluated by three metrics (see Table 3).

We can observe that the proposed XGBoost-BiGRU method outperforms other methods in the three evaluation metrics. The results indicate that using all 22 features as inputs to the BiGRU network leads to performance degradation in prediction. Redundant features have a negative impact on the accuracy of load prediction models and increase both model training time and computational complexity. As deep learning algorithms, LSTM and BiGRU perform better in nonlinear fitting compared to XGBoost and is more suitable for handling long-term sequence-related tasks. Therefore, the prediction performance of XGBoost-LSTM and the proposed method is superior to that of the double-layer XGBoost method. Compared to XGBoost-LSTM, BiGRU is used to map the relationship between input features and output loads and BO is used to optimize hyperparameters in our method. The proposed method fully learns knowledges from both past and future to train the model and finds the optimal combination of model parameters. Therefore, the proposed method achieves better load prediction performance and also reduces the computational complexity of the model.

For further visual demonstration, the load prediction results of all methods for November 1, 2023 are selected as examples (Figure 6), compared with the ground truth.

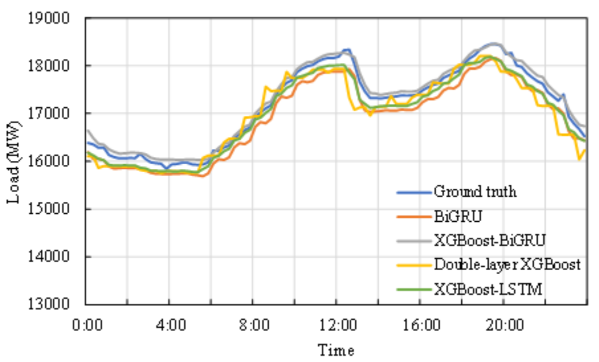


Figure 6: Examples of load forecasting results and ground truth

It can be seen that the prediction curves of both XGBoost-BiGRU and XGBoost-LSTM are close to the ground truth, while results of our method are even closer to the actual values. The trend of the prediction curve of BiGRU without feature selection is similar to the ground truth, but the single-point error is relatively large. The double-layer XGBoost method predicts sharp peaks at points of load value changes, resulting in larger prediction errors.

5 Conclusion

To improve the accuracy of load forecasting with only historical load data is available, we propose an ultra-short-term load forecasting method based on XGBoost-BiGRU. By proposing an XGBoost-based feature selection method, the meaningful information contained in historical load data are fully exploited. And an optimal feature set that includes temporal information and historical loads is established. Using the optimal features as network inputs, a load forecasting method based on BiGRU is proposed. Additionally, BO is employed to optimize hyperparameters of the model for load prediction for the next 15 minutes. Real load data are used to validate the effectiveness of the proposed method. The experimental results demonstrate that the method can effectively capture meaningful features for load forecasting and avoid the impact of redundant features. BiGRU learns knowledges from historical and future directions, resulting in improvement in prediction efficiency and accuracy. Therefore, our method has significance for guiding real-time supply-demand balance calculations and dispatch in power grids.

Funding

This work is supported in part by the Research on Key Technologies for Supply-Demand Balance Perception and Optimization Decision Support Based on Source-Grid-Load-Storage Data Coordination (527404239004). The contracting party (Party A) of this project is State Grid Xiongan Financial Technology Group Co., Ltd.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Wang, J. (2022). Application and Prospect of Source-grid-load-storage Coordination Enabled by Artificial Intelligence, *Proceedings of the CSEE*, 42(21), 7667–7682, 2022.

[2] Tang, X. (2022). Short-term Power Load Forecasting Based on Extreme Gradient Boosting and Temporal Convolutional Network, *High Voltage Engineering*, 48(8), 3059–3067, 2022.

[3] Sun, C. (2021). Ultra-short-term Power Load Forecasting Based on Two-layer XGBoost Algorithm Considering the Influence of Multiple Features, *High Voltage Engineering*, 47(8), 2885–2898, 2021.

[4] Zhou, S. (2023). Ultra-short-term Load Forecasting Based on Temporal Convolutional Network Considering Temporal Feature Extraction and Dual Attention Fusion, *Automation of Electric Power Systems*, 47(18), 193–205, 2023.

[5] Kong, X. (2023). Review on Electricity Consumption Characteristic Modeling and Load Forecasting for Diverse Users in New Power System, *Automation of Electric Power Systems*, 47(13), 2–17, 2023.

[6] Paparoditis, E.; Sapatinas, T. (2013). Short-Term Load Forecasting: The Similar Shape Functional Time-Series Predictor, *IEEE Transactions on Power Systems*, 28(4), 3818–3825, 2013.

[7] Amral, N.; Ozveren, C.S.; King, D. (2007). Short term load forecasting using Multiple Linear Regression, In *42nd International Universities Power Engineering Conference*, Brighton, UK, 1192–1198, 2007.

[8] Jiang, J. (2021). Peak load forecasting method of distribution network lines based on XGBoost, *Power System Protection and Control*, 49(16), 119-127, 2021.

[9] Chen, J.;Yang, J.; Lou, Z. (2019). A new short-term load forecasting model based on XGBoost algorithm, *Electrical Measurement and Instrumentation*, 56(21), 23–29, 2019.

[10] Jiao, R.; Chu, J.; Li, J.; Zhang, W. Personalized Federated Multi-region Load Forecasting Method Based on Data Decomposition, *Proceedings of the CSEE*, in press.

[11] Huang, N. Short-term Spatial-temporal Forecasting of Electric Vehicle Charging Load With Differentiated Spatial-temporal Coupling Correlation of Multiple Public Charging Stations, *Proceedings of the CSEE*, in press.

[12] Luo, S. (2020). Medium and Long-term Load Forecasting Method Considering Multi-time Scale Data, *Proceedings of the CSEE*, 40(S1), 11–19, 2020.

Member since 2012
JM08090

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Chen, S.; Li, G.; Chang, K.; Hu, X.; Li, P.; Wang, Y.; Zhang, Y. (2024). Ultra-short-term Load Forecasting Based on XGBoost-BiGRU, *International Journal of Computers Communications & Control*, 19(5), 6631, 2024.
https://doi.org/10.15837/ijccc.2024.5.6631