Improving the Bi-LSTM Model with XGBoost and Attention Mechanism: A Combined Approach for Short-Term Power Load Prediction

Yeming Dai[1]*, Qiong Zhou[1], Mingming Leng[2], Xinyu Yang[1], Yanxin Wang[1]

1. School of Business, Qingdao University, Qingdao 200071, China

2. Faculty of Business, Lingnan University, Hong Kong.

Yeming Dai, (Email: yemingdai@163.com)

Qiong Zhou, (Email: zq17806262874@163.com)

Mingming Leng, (Email: mmleng@ln.edu.hk)

Xinyu Yang, (Email: 865050851@qq.com)

Yanxin Wang, (Email: 457825811@qq.com)

Abstract

Short term power load forecasting plays an important role in the management and development of power systems with a focus on the reduction in power wastes and economic losses. In this paper, we construct a novel, short-term power load forecasting method by improving the bidirectional long short-term memory (Bi-LSTM) model with Extreme Gradient Boosting (XGBoost) and Attention mechanism. Our model differs from existing methods in the following three aspects. First, we use the weighted grey relational projection algorithm to distinguish the holidays and non-holidays in the data preprocessing. Secondly, we add the Attention mechanism to the Bi-LSTM model to improve the validity and accuracy of prediction. Thirdly, XGBoost is a newly-developed, well-performing prediction model, which is used together with the Attention mechanism to optimize the Bi-LSTM model. Therefore, we develop a novel, combined power load prediction model "Attention-Bi-LSTM + XGBoost" with the weight determination theory-error reciprocal method. We evaluate the developed prediction method using the dataset in Singapore's and Norway's power markets, and find that our prediction method outperforms the LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, and XGBoost models in effectiveness, accuracy and practicability.

Key words: Power load forecasting; Attention mechanism; Bidirectional long-short term memory network; Extreme gradient boosting; Weighted grey relational projection algorithm.

1. Introduction

With the continuous growth of electricity demand, traditional power grid faces challenges in centralized

distribution, manual monitoring and recovery, and two-way communication (Sobhani et al., 2020; Rosato et al., 2021). Smart grid acts as an effective solution to the above issues, since it is conducive to monitoring power production, transmission, and consumption as well as balancing the relationship (Jiang et al., 2020). However, the power load fluctuates greatly due to the influence of uncertain factors such as climate, economy, and environment (Nystrup et al., 2021). It is thereby difficult to estimate the future trend of electricity demand. Moreover, overestimation or underestimation of power load leads various challenges to power grid strategic decision-makings. Therefore, accurate and precise power load prediction is of great importance to ensuring safe operations of power systems as well as balancing the supply and demand of power load.

There are a number of power load prediction methods in existing literatures, which can be classified as four categories: (i) Classical prediction methods, (ii) Modern prediction methods, (iii) Hybrid prediction methods, and (iv) Combined prediction methods. Among them, classical prediction methods include time series analysis (Qiu et al., 2017), regression analysis (Wu et al., 2020), and other statistical methods, which all perform well in dealing with simple linear problems by using time series methods to estimate the future power. However, these methods are challenged in dealing with nonlinear problems. In order to better predict the nonlinear problems, the nonlinear mapping-based prediction technologies appear. The input data is embedded into high-dimensional space, which can transforms the nonlinear problems into linear problems. Classic modern prediction methods mainly include fuzzy logic, gray system (Zhao and Guo, 2016), artificial neural network (ANN) (Azadeh et al., 2008), support vector regression (Wang et al., 2021), etc. It's worth noting that those new prediction methods have inherent limitations such as complex calculation (Yang et al., 2019; Guo et al., 2018; Wang et al., 2021), poor generalization ability (Liang et al., 2019), and over fitting (Wang et al., 2017; Brégère et al., 2022; Ribeiro et al., 2019), which all challenge power load predictions.

To overcome the weaknesses of prediction methods above, the hybrid prediction models have been developed by involving the optimization algorithms, which include, e.g., the modified fire-fly optimization (mFFO) algorithm (Hafeez et al., 2021), particle swarm optimization (PSO) (Jnr et al., 2021), and Bayesian optimization (BOA) (Polamuri et al., 2021). These algorithms can help significantly improve prediction performance. For example, Wang et al. (2021) used the hybrid support vector regression (HSVR) method to predict the medium- and long-term loads, and applied the hierarchical method based on nested strategy and state transition algorithm (STA) to optimize the parameters of prediction models. Barman et al. (2020) optimized the parameters of support vector machine (SVM) using the grey wolf optimizer (GWO) and predicted the electricity demand that is significantly affected by social factors such as culture or religious rituals. The approach has a higher prediction accuracy compared to other methods. Moreover, in general, hybrid prediction methods consist of data preprocessing and forecasting parts. Preprocessing data through different technologies helps eliminate outliers, correct data errors, and improve data quality. Relevant technologies include (i) data decomposition technologies such as ensemble empirical mode decomposition (EEMD) (Wu et al., 2019; Qiu et al., 2017) and variational modal decomposition (VMD) (Jiang et al., 2020), and (2) feature selection technologies such as modified mutual information (MMI) (Hafeez et al., 2020), random forest (RF) (Lahouar and Slama, 2015), and weighted grey relational projection algorithm (WGRP) (Dai and Zhao, 2020). In conclusion, the above methods need to process the input data and optimize the parameters of prediction models, which can significantly reduce prediction error and improve prediction accuracy. It follows that the hybrid prediction methods perform well in predictions.

To further improve and optimize the prediction model and overcome the inherent defects of various single models in classical, modern, and hybrid prediction methods, a number of combined prediction methods were proposed by combining two or more different prediction models with a specific weighting method. Bates and Granger (1969) put forward the idea of combined forecasting for the first time. They proposed a seminal combined prediction model, which combines two independent airline datasets with a weighting system. The results show that the combined prediction set can produce a lower error than the original prediction. Nie et al. (2012), Deng et al. (2020), and Chu et al. (2021) also proved that the performance of combined prediction models is better than that of single models. Chen et al. (2020) used the LSTM and XGBoost models to predict the power load, respectively; then, they assign weights to the two models according to the error reciprocal method. For a better weighting method, the error should be reduced because a smaller error implies a higher prediction accuracy. Zhuang et al. (2021) set an initial weight of model combination to search for the best weight combined with the MAPE-RW algorithm, and then constructed the CNN-LSTM-XGBoost combined prediction model, which significantly reduced the error index compared with the single prediction models. Nie et al. (2020) used the multi-objective grey wolf algorithm (MOGWO) to determine the weights to the radial basis function network (RBF), generalized regression neural network (GRNN), and extreme learning machine (ELM). They established a combined prediction model based on the swarm intelligence optimization, which can effectively reduce adverse effects of the weak adaptability of single models and better grasp the characteristics of power load, thus significantly improving the prediction accuracy and adaptability.

We can learn from the above literatures review that the recent improvement of prediction models with the combination methods mentioned above not only needs more than one single prediction model but also combines a variety of different algorithms to calculate the weights for each model. Such combined prediction methods have improved the prediction performance (Zheng et al., 2020). However, the above combined prediction models did not emphasize the importance of data preprocessing, but instead considered a few determinant factors and mostly combined existing mature prediction models without any significant prediction improvement. In this context, we use the WGRP algorithm to preprocess the data and eliminate the impact of holidays. As for the prediction process, the Bi-LSTM model had been widely viewed as one with an excellent forecasting effect because it can fully consider the hidden information and obtain better prediction results. However, the Attention mechanism has the advantages of large-scale parallel processing, distributed information storage, and acceptable self-organization and self-learning ability. We then first add the Attention mechanism to the Bi-LSTM model (Yu et al., 2020), which can eliminate the unreasonable impact, emphasize the impact of key input data, and make the results more comprehensive. The model is called “Attention-Bi-LSTM model” (Zheng and Chen 2021). To avoid the defects of a single prediction model, we combine the XGBoost (Chen et al., 2015) model with the Attention-Bi-LSTM model. The use of the XGBoost model is attributed to the fact that the XGBoost not only controls the model complexity by adding regular terms to prevent over-fitting and improves the generalization ability of the model, but also carries out the leaf splitting optimization calculation only by relying on the input data value without selecting the specific form of loss function (Trizoglou et al., 2021). Therefore, we use the “Attention-Bi-LSTM+XGBoost” combination model for power load prediction. In order to verify the effectiveness of developed prediction method, we analyze two power market cases in Singapore and Norway.

The novelty and major technical contributions are as follows:

(1) We develop a novel Attention-Bi-LSTM + XGBoost combined prediction model.

(2) We use the weighted grey relational projection algorithm to distinguish the holiday and non-holiday data.

(3) We consider the Attention mechanism in the Bi-LSTM model to improve the prediction accuracy.

(4) We use the XGBoost model to further improve the performance of Bi-LSTM model in a combined manner.

(5) We use two power market data and six power load prediction models to verify the effectiveness and reliability of our developed method in this paper.

The organization of this paper is as follows: The basic methods and algorithms used in this paper are introduced in Section 2. Sections 3 introduces the weight method of combining the models. Sections 4 presents the specific steps of our proposed prediction method. We consider two practical case to verify the prediction accuracy and stability of our method in Sections 5. This paper ends with concluding remarks and possible future directions in Section 6.

2 Methodologies

2.1 Weighted grey relational projection algorithm

The WGRP algorithm (Dai and Zhao, 2020) is a method for measuring the degree of similarity or difference between the development trends of various factors, i.e., "grey relational degree." This method is not limited by the sample size. For the data with small sample size and discreteness, it can avoid the one-way deviation caused by comparing the index values of single factors of each scheme, and also can comprehensively analyze the relationship between the indexes while the size of the module and the cosine of the included angle are combined. The proximity between each decision scheme and the ideal scheme is fully and accurately reflected. It is also applicable to the regular sample size with small amount of calculation. Therefore, this paper uses the WGRP algorithm to sort the factors that affect power load, and assigns weights according to the importance of these factors. Thus, the prediction results are general. The details regarding the calculation steps are as follows.

Firstly, select the data of the preceding n_1 samples and the data of the samples to be predicted, calculate the relationship coefficient between them, and construct the following grey relationship matrix.

$$\boldsymbol{A}=\begin{bmatrix} A_{01}\cdots A_{0m_1} \\ \vdots \quad \ddots \quad \vdots \\ A_{n_1 1}\cdots A_{n_1 m_1} \end{bmatrix} \tag{1}$$

where $A_{n_1 m_1}$ represents grey correlation coefficient whose m_1-th factor in the n_1-th sample.

Then, the weight of each influencing factor is calculated by entropy weight method, and the weighted grey relation matrix is obtained by weighting the grey relation matrix, as shown below:

$$\boldsymbol{A'}=\boldsymbol{A}\boldsymbol{\gamma}^{\boldsymbol{T}}=\begin{bmatrix} \gamma_1\cdots\gamma_{m_1} \\ \vdots \ddots \vdots \\ \gamma_1 A_{n_1 1}\cdots\gamma_{m_1} A_{n_1 m_1} \end{bmatrix} \tag{2}$$

where, $\boldsymbol{\gamma}$ represents the weight of each influencing factor, the first row in the matrix is expressed as the row vector of the sample to be predicted of $\boldsymbol{A_0'}$, the row vector of other historical samples is expressed as $\boldsymbol{A_i'}$, and the included

angle between each A_0' and A_i' is the gray projection angle of the sample, expressed as θ_i, and calculate the $\cos\theta_i$.

$$\cos\theta_i = \frac{\sum_{j=1}^{m_1} \gamma_j A_{ij} \gamma_j}{\sqrt{\sum_{j=1}^{m_1} \left(\gamma_j A_{ij}\right)^2} \sqrt{\sum_{j=1}^{m_1} \gamma_j^2}} \tag{3}$$

Thus, the weighted grey correlation projection value B_i is

$$B_i = \frac{\sum_{j=1}^{m_1} \gamma_j A_{ij} \gamma_j}{\sqrt{\sum_{j=1}^{m_1} \gamma_j^2}} \tag{4}$$

Finally, the obtained projection values are sorted from large to small, and the samples with large projection values are selected as similar samples for replacement.

2.2 Bi-LSTM forecasting model

In the gradient algorithm of recurrent neural network, when the time steps are too small or too large, the gradient of recurrent neural network is easy to explode and disappear. Therefore, in order to solve this problem, LSTM uses gating mechanism to control information, and introduces input gate, forgetting gate and output gate to remove some contents that are not important to the current situation (Zheng et al., 2020), thereby prolonging the storage time of information and save some older information. The input of LSTM gate is the hidden state between the current time step input $\boldsymbol{X_t}$ and the previous time step H_{t-1}. The output is calculated by the full connection layer.

$$\text{Input gate:}\quad I_t = \sigma\left(\boldsymbol{X_t W_{xi}} + H_{t-1}\boldsymbol{W_{hi}} + b_i\right) \tag{5}$$

$$\text{Forgetting gate:}\quad F_T = \sigma\left(\boldsymbol{X_t W_{xf}} + H_{t-1}\boldsymbol{W_{hf}} + b_f\right) \tag{6}$$

$$\text{Output gate:}\quad O_t = \sigma\left(\boldsymbol{X_t W_{xo}} + H_{t-1}\boldsymbol{W_{ho}} + b_o\right) \tag{7}$$

where h is the number of hidden units, $\boldsymbol{X_t}$ is the small batch input of a given time step t, H_{t-1} is the hidden state of the previous time step, σ is sigmoid function, $\boldsymbol{W_{xi}}$ is the weight matrix of the input gate, b_i is the offset term of the input gate, $\boldsymbol{W_{xf}}$ is the weight matrix of the forgetting gate, b_f is the offset term of the forgetting gate, $\boldsymbol{W_{xo}}$ is the weight matrix of the output gate, and b_o is the offset term of the output gate. Short term memory needs to calculate the candidate memory cells $\widetilde{C}_t$, using the tanh function with the value range in [-1,1] as the activation function:

$$\widetilde{C}_t = \tanh\left(\boldsymbol{X_t W_{xc}} + H_{t-1}\boldsymbol{W_{hc}} + b_c\right) \tag{8}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \widetilde{C}_t \tag{9}$$

$$H_t = o_t \odot \tanh\left(C_t\right) \tag{10}$$

Equations (8) and (9) calculate the cell state C_t at the current time, $\boldsymbol{W_{xc}}$ and $\boldsymbol{W_{hc}}$ are weight matrices. It

is generated are weight matrices by multiplying the last cell state $C_{t\text{-}1}$ by the forgetting gate F_T with element ⊙, multiplying the current input cell state $\widetilde{C}_t$ by the input gate I_t with element ⊙, and adding the two results. Furthermore, the current memory $\widetilde{C}_t$ and long-term memory $C_{t\text{-}1}$ of LSTM are combined to form a new unit state C_t. Due to the control of the forgetting gate, older information can be saved, and through the control of the input gate, some irrelevant content can be discarded. Finally, the LSTM output is obtained by formula (10).

Different from the LSTM, the Bi-LSTM (Bi-directional long short-term memory) method is composed of forward LSTM and backward LSTM. When extracting data features, we take into account the overall information hidden in the data, and extract features from both forward and reverse angles (Zhuang et al., 2021). Then, the results of two-way extraction are combined in a specific way and summarized from two dimensions, which can eliminate the impact of the order of input data in a single LSTM on the final result to a certain extent and make the results more comprehensive.

2.3 Attention mechanism

The core idea of Attention mechanism is to simulate attention ability of people. For the information to be processed, people usually focus on a few key points instead of evenly distributing their attention to all information. Therefore, the introduction of Attention mechanism into the prediction model can assign different weights to the data, eliminate the unreasonable impact of input data on output data, and improve the impact of key input data.The model structure of attention is shown in Figure 1. For the specific calculation steps, see, for example, Zheng and Chen (2020).

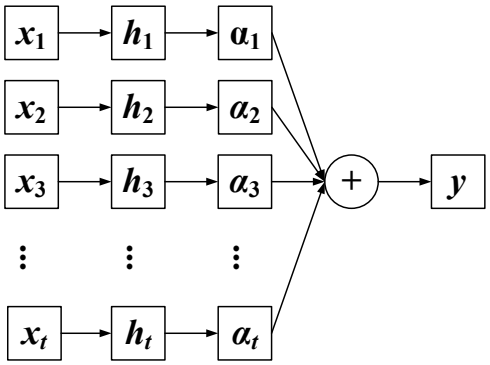

Fig.1 Structure of attention mechanism.

2.4 XGBoost power load forecasting model

Extreme gradient boosting is essentially a gradient boosting decision tree, which can improve the speed and efficiency of prediction. It is an optimization of the boosting algorithm that builds a decision tree by continuously adding trees and continuously splitting features (Chu et al., 2020). When we add a tree, a new function $f(x)$ is learned to fit the residual predicted last time. When the training is completed and k trees are obtained, each tree falls to a corresponding leaf node, and each leaf node corresponds to a score. It is only necessary to add up the corresponding scores of each tree to get the predicted value of the sample. The XGBoost model is as follows:

$$\hat{y}_i = \sum_{j=1}^{n} w_j x_{ij} \tag{11}$$

where $\hat{y}_i$ is the predicted value, n is the number of trees, w_j is the weight, and x_{ij} is sample data.

In each iteration, a tree is added on the basis of the existing tree to fit the residual between the predicted results of the previous tree and the real value. The iterative process is as follows:

$$\begin{aligned}
&\hat{y}_i^{(0)} = 0 \\
&\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\
&\hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = y_i^{(1)} + f_2(x_i) \\
&\cdots \\
&\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)
\end{aligned} \tag{12}$$

where $\hat{y}_i^{(t)}$ is the model after training t round; $\hat{y}_i^{(t-1)}$ is the reserved function added in the previous round; and $f_t(x_i)$ is the newly added function. The objective function of XGBoost is as follows:

$$\begin{aligned}
Obj^{(t)} &= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \\
&= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)
\end{aligned} \tag{13}$$

$$\Omega f(t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{14}$$

The ultimate goal is to find f_t that minimizes the objective function, the $\sum_{i=1}^{t} \Omega(f_i)$ in formula (13) is a regular term in the objective function, which determines the complexity of the tree. Moreover, a smaller value results in a lower complexity and the stronger generalization ability. In formula (14), T is the number of leaf nodes, ω is the score of leaf node, γ is used to control the number of leaf nodes, and λ ensures that the score of leaf nodes is not too large.

In order to find a f_t to minimize the objective function, Taylor's second-order expansion is carried out at $f_t = 0$, and the objective function obtained is approximately as follows:

$$\tau^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{15}$$

Where $g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right)$ is the first derivative and $h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right)$ is the second derivative.

Since the prediction score of the first $t-1$ trees and the residual error of y will not affect the optimization of the objective function, it is directly removed and the objective function is further simplified as:

$$\tilde{\tau}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{16}$$

Formula (16) superimposes the loss function values of each sample and reorganizes all samples with the same leaf node. The process is as follows:

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

$$= \sum_{i=1}^{n}\left[g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2 \right] + \gamma T + \lambda \frac{1}{2} \sum_{j=1}^{T} \omega_j^2 \tag{17}$$

$$= \sum_{j=1}^{T}\left[\left(\sum\nolimits_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum\nolimits_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T$$

For the rewritten univariate quadratic function about the leaf node fraction ω, the optimal ω_j^* and objective function can be obtained as follows:

$$\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{18}$$

$$Obj = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{19}$$

In order to facilitate calculation and meet the requirements of data input, the data shall be normalized in advance. The power load data is normalized according to the following formula. The data is limited to the range of [0,1].

$$x^n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{20}$$

where x and x^n are the power load data value before and after normalization, respectively. $x_{\min}$ and $x_{\max}$ are the minimum and maximum of the power load data value before normalization, respectively.

3 Attention-Bi-LSTM + XGBoost power load combined prediction model

3.1 Weighting method

The research result of Chen et al. (2020) shows that the error reciprocal method is not only easy to operate, but also can significantly optimize the prediction performance of the model. Therefore, in this paper, the reciprocal error method is used to assign weights to the model. The prediction model with a smaller error in this combined model is given a larger weight. Hence, the overall error of the combined prediction model can be reduced significantly. To confirm the weight coefficient, the formula of error reciprocal method is as follows:

$$f_t = \omega_1 f_{1t} + \omega_2 f_{2t}, t = 1,2,\cdots,n \tag{21}$$

$$\omega_1 = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_2} \tag{22}$$

$$\omega_2 = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2} \tag{23}$$

where ω_1 and ω_2 mean the weight value of Attention-Bi-LSTM and XGBoost respectively; f_{1t} and f_{2t} mean the predicted value obtained by Attention-Bi-LSTM and XGBoost. The weight value is obtained from formulas (22) and (23), where ε_1 and ε_2 are the error values of the prediction models Attention-Bi-LSTM and XGBoost respectively.

3.2 Attention-Bi-LSTM + XGBoost combined prediction model

Different from the existing prediction models, the Attention-Bi-LSTM can not only fully consider the overall information hidden in the input data from two dimensions to obtain more comprehensive results, but also emphasize

the impact of key input data. Thus, the use of Attention-Bi-LSTM model can improve the prediction accuracy of results. Moreover, XGBoost, as a newly proposed prediction model with a low complexity, can prevent over fitting and has an excellent prediction performance. We then use Attention-Bi-LSTM and XGBoost methods to forecast the power load, and obtain the corresponding errors. The weights for the above two models are calculated by using the error reciprocal method according to the error results, which gives a greater weight to the model with a smaller error, so as to maximize the advantages of the model and reduce the error as much as possible. Finally, we combine the different prediction results of above two models by using the error reciprocal method, which can overcome various inherent defects of a single prediction model. The framework is shown in Figure 2.

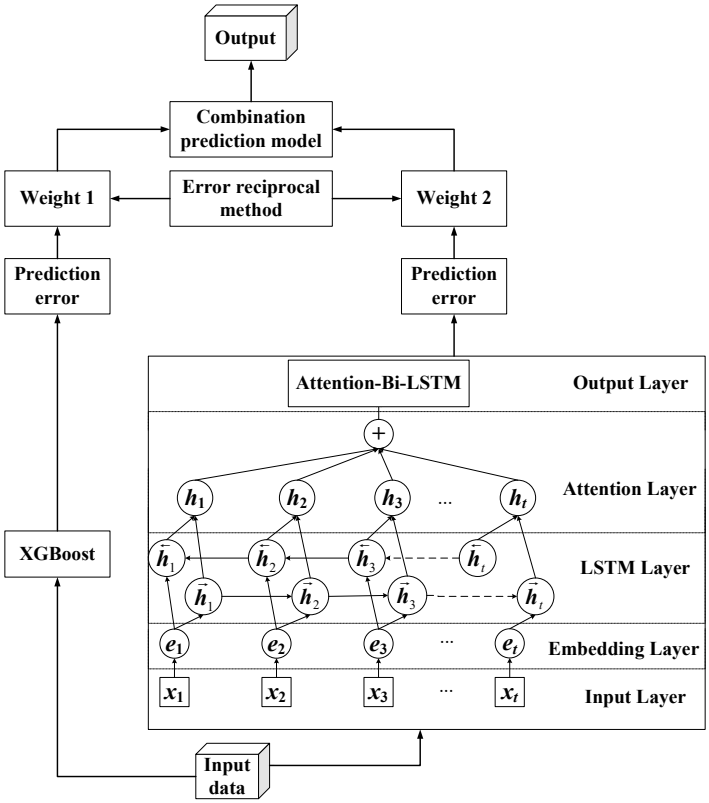

Fig.2 The framework of Attention-Bi-LSTM + XGBoost combined prediction model.

4 Prediction method

Based on our newly proposed combination forecasting model, the power load prediction process in our method is shown in Figure 3. The prediction steps of the full text are mainly divided into four stages:

Stage 1: Data preprocessing. First, we select several influencing features with the greatest correlation, such as time, day type, holiday type, real-time price. Then, we use the WGRP algorithm to process the data of holidays to distinguish holiday and non-holiday data, making the data more general. Finally, we normalize the data.

Stage 2: Prediction using single models. The Attention mechanism is used to optimize the LSTM model, so that the influence of unreasonable factors can be eliminated and then the influence of key input data can be emphasized to make the results more comprehensive. Single Attention-Bi-LSTM and XGBoost model are used to predict the same dataset and prepare for the combination of the two models according to the prediction results in Stage 3.

Stage 3: Weight the models. After the power load data are forecasted with Attention-Bi-LSTM and XGBoost methods, we use the error reciprocal method to obtain the weights according to the error predicted, which means

that single models are weighted. Then, the Attention-Bi-LSTM + XGBoost combined prediction model forms.

Stage 4: Evaluation of prediction results. By comparing the prediction errors for six benchmark models with two power markets, we show whether this method can improve the accuracy of power load forecasting or not.

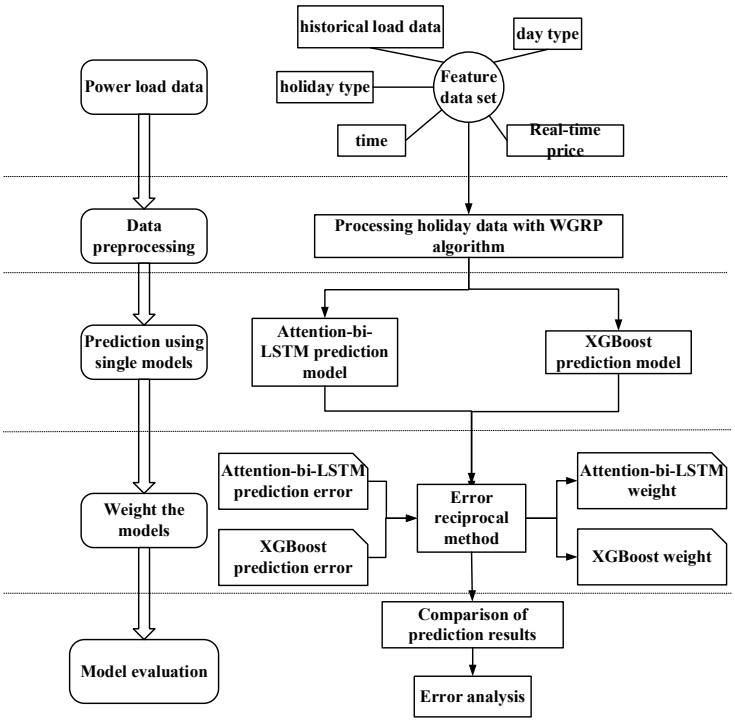

Fig.3 Power load prediction process in our novel method.

5 Evaluation and analysis

In this section, we evaluate and discuss the performance of developed prediction method based on two cases. Simulation is carried out in Python to verify the effectiveness of the proposed framework. Since LSTM, Bi-LSTM, Attention-Bi-LSTM, Attention-LSTM, Attention-RNN and XGBoost has a high coherence with the model in this paper, we select them as the benchmark models and compare them with our proposed method.

5.1 Datasets and experimental environment

After screening the data of several national power markets, we find that the data of Singapore and Norway power markets embrace all the influencing factors we need. The relevant data is complete and comprehensive consideration is more suitable for the model proposed in this paper. Therefore, we use the data to evaluate the proposed method and consider the factors such as day type, time, holiday type, real-time price, etc. These data include Singapore's data from January 1, 2019 to December 31, 2020 and Norway's data from January 1, 2020 to December 31, 2021. The sampling period of historical power load data is 1 hour; and, 80% of the prepared data is used for training and 20% for testing. Finally, we compare the results with the real data and investigate the errors.

The hardware platform of this experiment is equipped with Intel i5-1035G1 processor, with 8GB memory, 477GB solid state disk capacity and MX230 CPU graphics card. The method proposed in this paper is implemented based on Python language. The Attention-Bi-LSTM model uses keras deep learning framework and XGBoost uses py-xgboost framework.

5.2 Evaluation criteria

We consider the prediction accuracy as our objective to test the proposed combined model's efficiency. To

evaluate this objective, we use two standard statistical indicators, select mean absolute percent error (MAPE) as the main evaluation index of each prediction model, and choose mean absolute error (MAE) as the auxiliary evaluation index. These statistical indicators are used to test the accuracy of the proposed framework. The calculation formulas of MAPE and MAE are as follows:

$$X_{\mathrm{MAPE}} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \hat{y}_t\right|}{y_t} \tag{24}$$

$$X_{\mathrm{MAE}} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \tag{25}$$

$$X_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2} \tag{26}$$

where n means the quantity of power load data, y_t means the real power load data, and $\hat{y}_t$ means the predicted power load data.

5.3 Influence of data preprocessing on prediction results

To verify the significance of the WGRP algorithm, we extract the short holiday data from May 1, 2020 to May 5, 2020 in the dataset of Singapore power market from January 1, 2019 to December 31, 2020. Then, comparing the prediction accuracy of the WGRP algorithm before and after data preprocessing, we specify the details of the four experiments below.

(1) Attention-Bi-LSTM without WGRP algorithm (Model 1). According to the local conditions of Singapore, Model 1 determines the characteristics of historical load series, hour, day type, and holiday type, and also uses Attention-Bi-LSTM model to predict power load.

(2) Attention-Bi-LSTM processed by WGRP algorithm (Model 2). For holidays, Model 2 selects the historical data with a high similarity to this holiday with the WGRP algorithm, to make the overall historical data general.

(3) XGBoost without WGRP algorithm (Model 3).

(4) XGBoost processed by WGRP algorithm (Model 4)

We use Models 3 and 4 to investigate the impact of WGRP algorithm on the prediction accuracy of XGBoost model.

The characteristics in the above experiments are artificially selected, and there are certain subjective factors. Based on the model proposed in this paper, the objective feature selection method, WGRP algorithm, is used to select the historical load series of holidays to be predicted. Furthermore, we compare the results and analyze the errors of the above four models, the comparison results are shown in Figures 4 and 5; and, the MAPE and MAE values of each model are shown in Tables 1 and 2.

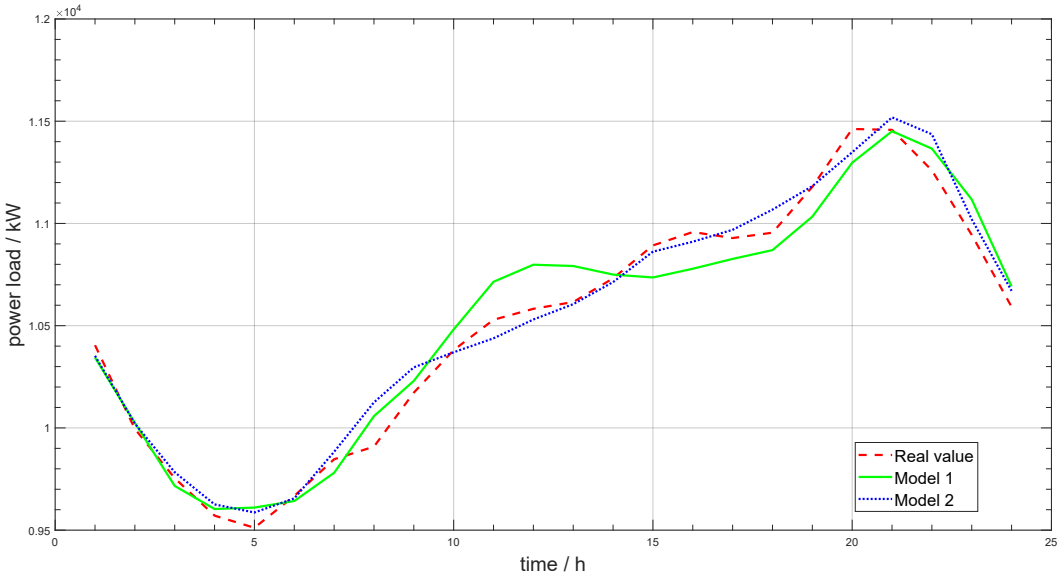

Fig.4 Attention-Bi-LSTM and Attention-Bi-LSTM of WGRP algorithm prediction results with real values

In Figure 4, Attention-Bi-LSTM (Model 1) and Attention-Bi-LSTM processed by WGRP algorithm (Model 2) are compared. We can find that Model 2 is closer to the real value than Model 1. Table 1 indicates that the MAPE, MAE, and RMSE values of the Attention-Bi-LSTM processed by the WGRP algorithm (Model 2) are 0.639, 67.147, and 96.232, respectively, which means that Model 2 has a higher prediction accuracy than Model 1. Therefore, for Attention-Bi-LSTM, after selecting historical holiday data of May 1, 2020, we can improve the accuracy of the model by using the WGRP algorithm, which shows the effectiveness of the WGRP algorithm.

Table 1 Error analysis of Attention-Bi-LSTM and Attention-Bi-LSTM of WGRP algorithm

Model	X_{MAPE} / %	X_{MAE} / kW	X_{RMSE} / kW
Model 1	0.971	103.063	119.726
Model 2	**0.639**	**67.147**	**96.232**

In Figure 5, XGBoost (Model 3) and XGBoost processed by WGRP algorithm (Model 4) are compared, and the line's trend of the XGBoost processed by WGRP algorithm (Model 4) roughly coincides with the real value. We also learn from Table 2 that the MAPE, MAE, and RMSE values of XGBoost processed by WGRP algorithm (Model 4) are 0.409, 42.976, and 43.817, respectively, which means that Model 4 has a higher prediction accuracy than Model 3. Thus, for XGBoost, the WGRP algorithm can help improve the prediction accuracy of the model.

When the WGRP algorithm is not considered, the predicted results are basically consistent with the change trend of the actual value; but, the differences of them are large. After the use of the WGRP algorithm, the gap between the prediction results and real value can be reduced. Consequently, to decrease the prediction error, the holiday data in the historical data is first processed by the WGRP algorithm. Then, in order to verify the advantages of the algorithm in improving prediction accuracy and reducing error, we use the same model to predict the processed and unprocessed data. We find that the prediction results of data processing are closer to the true value, and the accuracy of prediction can be further improved. Therefore, we use the WGRP algorithm to process the holiday data, which makes the data more general and improves the prediction accuracy.

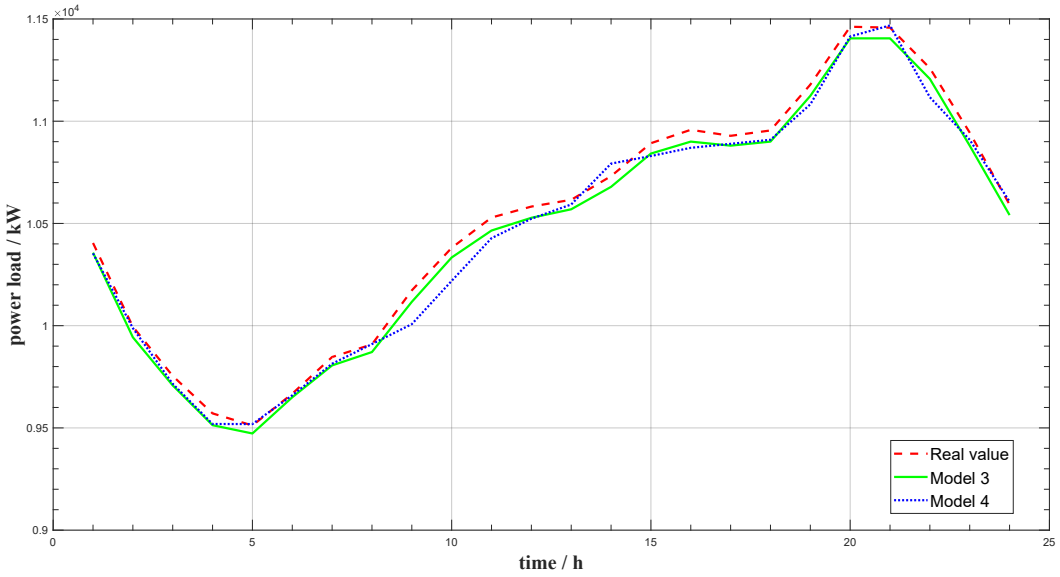

Fig.5 XGBoost and XGBoost of WGRP algorithm prediction results with real values

Table 2 Error analysis of XGBoost and XGBoost of WGRP algorithm

Model	X_{MAPE} / %	X_{MAE} / kW	X_{RMSE} / kW
Model 3	0.475	50.053	50.932
Model 4	**0.409**	**42.976**	**43.817**

5.4 Analysis of prediction results

5.4.1 Experiment I: Singapore electricity market

Taking Singapore power load data with hourly resolution as an example, we do this experiment to verify the applicability of our proposed power load forecasting approach. First, we use the Attention-Bi-LSTM and XGBoost models to predict the same set of data, and obtain the prediction errors of the two models. Secondly, according to the errors, we assign weights to the above two prediction models and combine them with the error reciprocal method. The model with a smaller error is given a higher weight. Thirdly, to verify the effectiveness of the combination model proposed in this paper, we use the LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost and "Attention-Bi-LSTM + XGBoost" combined forecasting model to predict the data, and compare the seven prediction results to show that our proposed model is most effective. From our results we observe the following issues.

(1) Trend comparison between prediction results and real data.

a) Figure 6 shows the comparison between the real values and the prediction results of LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost models, and "Attention-Bi-LSTM+XGBoost" combined prediction model. We learn that the prediction accuracy of "Attention-Bi-LSTM+XGBoost" combined prediction model is the highest, and the Bi-LSTM model has the lowest prediction accuracy.

b) The predicted value curve of the "Attention-Bi-LSTM + XGBoost" combined prediction model is the closest to the real value curve. That is, the fitting effect of our proposed model is the best, and the change trend is roughly the same.

c) Overall, the "Attention-Bi-LSTM + XGBoost" combined prediction model shows the performance of the optimal and behaves is better than others in prediction accuracy and sensitivity to proportionality changes.

The locally-enlarged drawing of the results of 24 hour on January 7, 2020 in Singapore is shown in Figure 7. We note from Figure 7 that the prediction accuracy of our proposed method is the highest according to both the overall trend diagram and the locally enlarged diagram.

(2) Error comparison and analysis

The MAPE, MAE, and RMSE values of the above six models are shown in Table 3. By comparing the error values in Table 3, we can draw the following conclusions.

a) Since LSTM is an improvement of recurrent neural network (RNN), the error of Attention-LSTM is less than that of Attention-RNN, which shows the necessity of selecting the LSTM model in this paper.

b) According to the comparison of LSTM, Attention-LSTM, and Attention-RNN models, the comprehensive comparison of the three prediction methods indicates the significant effect of attention mechanism on prediction accuracy.

c) Among all the accuracy test standards for benchmark models, the error value of XGBoost is the smallest, which means that XGBoost has an extremely excellent prediction performance. We use XGBoost to optimize the Attention-Bi-LSTM model, which can significantly improve the accuracy of the Attention-Bi-LSTM model.

d) Table 3 exposes that the values of MAPE and MAE of the "Attention-Bi-LSTM + XGBoost" combined prediction model 0.445 and 49.546, respectively, which are the smallest from a holistic perspective. Although the RMSE value of XGBoost is the smallest 62.585, the value of the proposed combined prediction model is only slightly higher than the minimum value, which is obviously acceptable. Therefore, the test results show that the combined prediction method of the two models can reduce the prediction error as a whole, thus being better than the single prediction model and having the highest prediction accuracy.

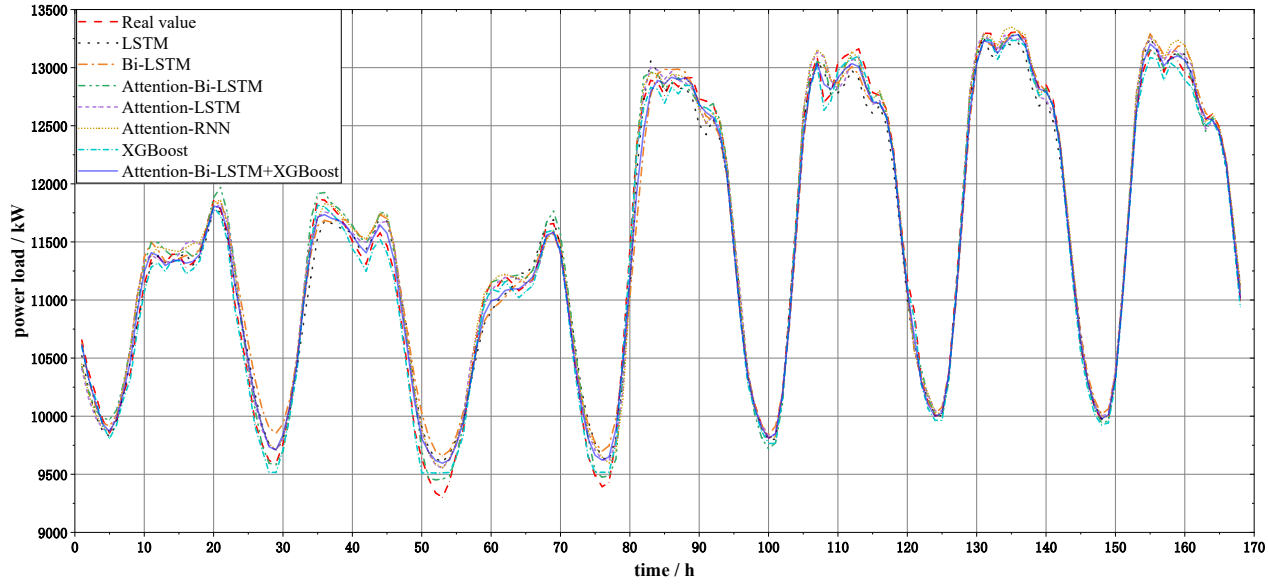

Fig.6 Comparison between LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost and Attention-Bi-LSTM+XGBoost prediction results with real values of Singapore power market

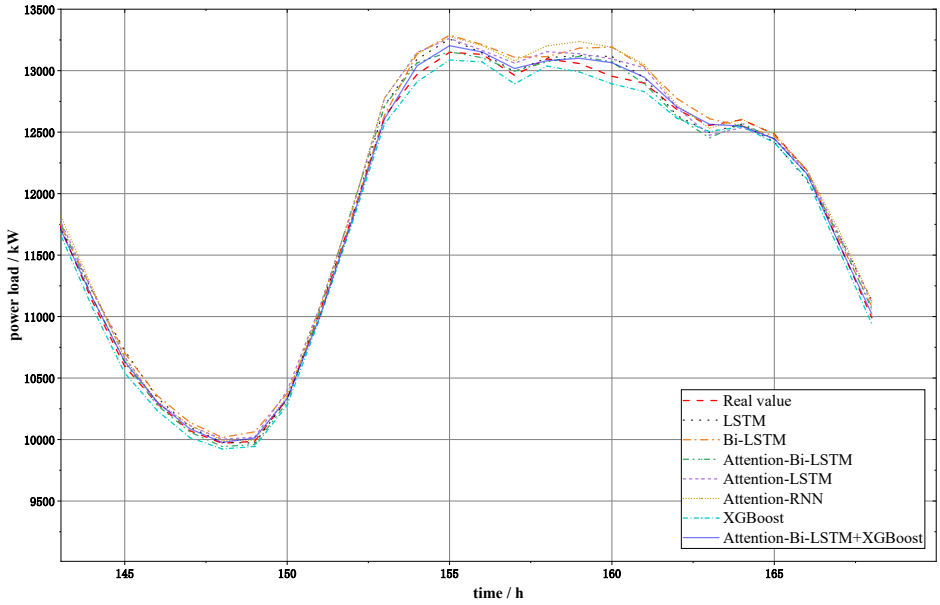

Fig.7 The local enlarged drawing of Singapore's results on January 7, 2020 (24 hours)

Table 3 Error analysis between LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost and Attention-Bi-LSTM+XGBoost Singapore power market

Model	X_{MAPE} / %	X_{MAE} / kW	X_{RMSE} / kW
LSTM	1.011	113.480	147.083
Bi-LSTM	1.155	127.812	168.841
Attention-RNN	0.886	98.16	130.704
Attention-LSTM	0.877	97.691	125.574
Attention-Bi-LSTM	0.711	79.418	105.023
XGBoost	0.5256	59.463	**62.585**
Attention-Bi-LSTM+XGBoost combined model	**0.445**	**49.546**	63.405

5.4.2 Experiment II: Norway electricity market

The hourly load data of Norway is used as another test data in this experiment. The purpose is to further verify and evaluate the effectiveness of the proposed method by using a different dataset to compare our results with those from other prediction approaches. Figures 8 and 9 as well as Tables 4 describe the graphics and error results of experiment II of the proposed combined prediction model compared to the existing prediction models. The results reveal the insights below.

(1) Trend comparison between prediction results and real data

Through the comparison of the seven prediction models in Figures 8 and 9, we find that compared to the benchmark models, the line of the combined prediction model proposed in this paper is the closest to the trend and value of the real value. This shows that the idea of optimizing the model with the combined method is feasible. Particularly, we learn from Figure 9 that there is a significantly large deviation between the real data and the predicted values of LSTM and other models. That is, from the perspective of graphic trend, the proposed model has a higher fitting degree.

(2) Error comparison and analysis.

Table 4 lists the numerical results of three accuracy tests, which show that in the accuracy test, the method proposed in this paper achieves the minimum results of MAPE, MAE, and RMSE (i.e., 0.682, 96.278, and 125.343, respectively). According to the evaluation, we conclude that the proposed model is more accurate than those benchmark frameworks. In addition, we further ensure the effectiveness of the attention mechanism and combination optimization model that were proved in Experiment I.

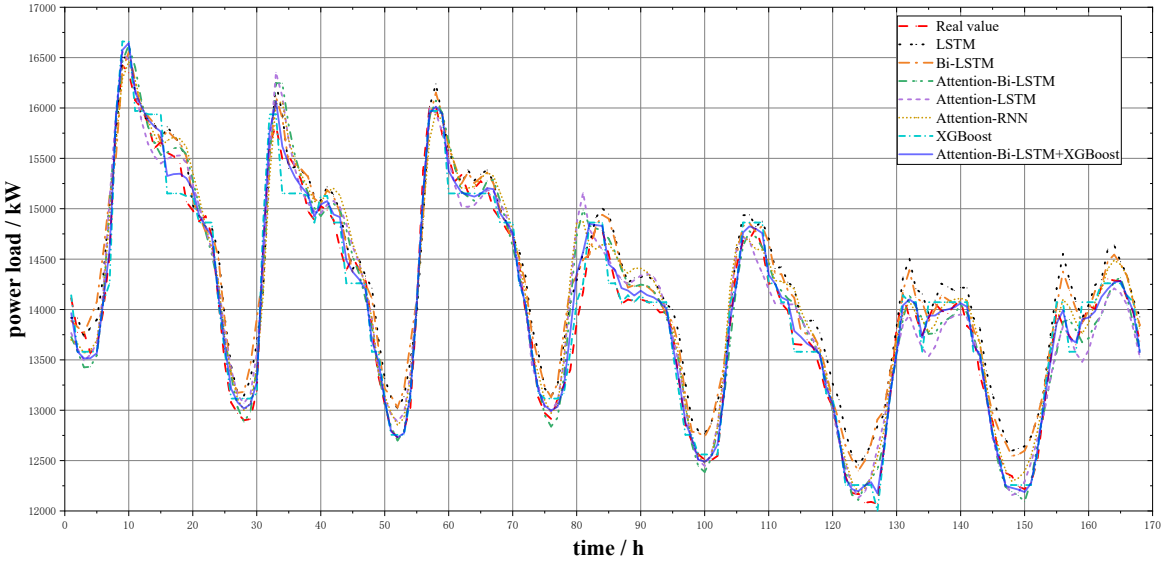

Fig.8 Comparison between LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost and Attention-Bi-LSTM+XGBoost prediction results with real values of Norway power market

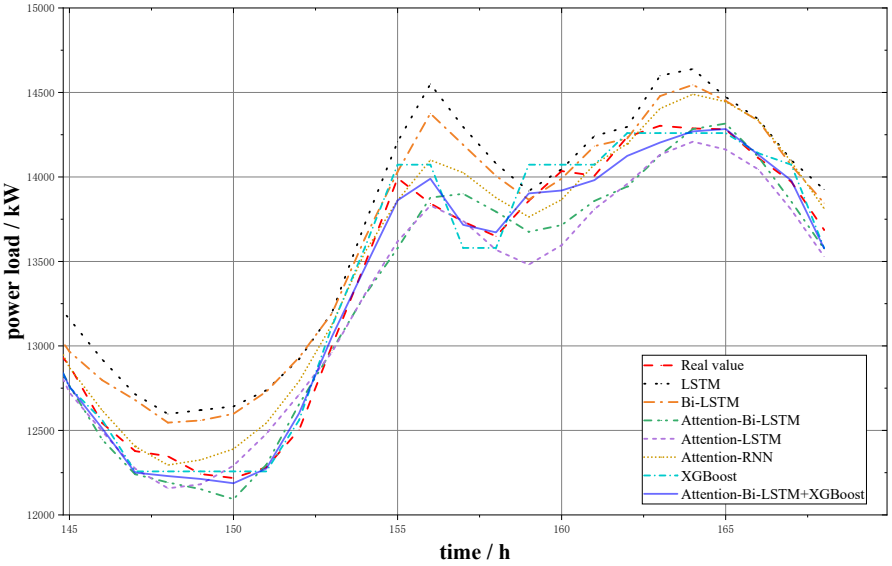

Fig.9 The local enlarged drawing of Norway's results on May 16, 2021 (24 hours)

Table 4 Error analysis between LSTM, Bi-LSTM, Attention-RNN, Attention-LSTM, Attention-Bi-LSTM, XGBoost and Attention-Bi-LSTM+XGBoost of Norway power market

Model	X_{MAPE} / %	X_{MAE} / kW	X_{RMSE} / kW
LSTM	1.886	257.791	301.141
Bi-LSTM	1.648	225.233	279.765
Attention-RNN	1.230	182.122	238.673
Attention-LSTM	1.229	171.672	233.651
Attention-Bi-LSTM	1.091	153.595	210.720
XGBoost	0.814	115.892	148.007
Attention-Bi-LSTM+XGBoost combined model	**0.682**	**96.278**	**125.343**

6 Conclusions

Power load forecasting plays an important role in balancing energy distribution, economy, and the safe and reliable operation of power systems. An accurate load forecasting can reduce the cost and risk of power operations, and can improve the environmental and economic benefits of power grids. Thus, in this paper, with an aim to enhance the accuracy and stability of power load predictions, we propose a novel hybrid Attention-Bi-LSTM + XGBoost power load combined forecasting method based on WGRP algorithm. On the phase of data preprocessing, the historical load series of holidays are selected by WGRP algorithm, and better prediction results are obtained. In addition, for the accuracy comparison between our combined forecasting model and the benchmark models such as LSTM, Bi-LSTM, Attention-Bi-LSTM, Attention-LSTM, Attention-RNN and XGBoost, we perform two case studies using the datasets of Singapore and Norway power markets. We can draw the following conclusions.

(1) Using the WGRP algorithm to preprocess holiday data can effectively improve the prediction accuracy of the model.

(2) Attention mechanism allows the Bi-LSTM model to emphasize the influence of important factors, which can eliminate redundancy and improve prediction performance.

(3) Adding regular terms to XGBoost can effectively prevent over fitting and reduce calculation, so as to greatly improve the efficiency of the algorithm. Therefore, using XGBoost model for optimization can greatly reduce the error of the model.

(4) Compared with the prediction results of all benchmark models, the prediction result of the “Attention-Bi-LSTM + XGBoost” combined model has the lowest errors and is the closest to the actual value than those of the single models, and the trend of the proposed model is roughly the same as the real values.

In conclusion, the combined forecasting method proposed in this paper is more effective than the single classical and modern prediction methods, hybrid prediction methods, and other existing combined prediction methods. Our model can reduce the error and obtain a higher power load prediction accuracy to reduce the unnecessary waste in power markets and improve the reliability and safety of power system operations.

In future, we may further improve our model. For example, we may consider whether we can find a suitable method to further optimize the parameters of our model and develop a better weight assignment algorithm that can further improve the prediction accuracy or not. We expect that, in the future, we can solve the above shortcomings, and also strive to apply the method to more fields.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 72171126), Ministry of Education Project of Humanities and Social Science (No.20YJA630009), and Social Science Planning Project of Shandong Province (No.20CSDJ15). This work is also financially supported by the Faculty Research Grant (FRG) of Lingnan University (No.DB21B1).

Reference

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2008). A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. *Energy policy*, *36*(7), 2637-2644.

Barman, M., & Choudhury, N. B. D. (2020). A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. *Sustainable Cities and Society*, *61*, 102311.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451-468.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

Chen, Z., Liu, J., Li, C., Ji, X., Li, D., Huang, Y., & Di, F. (2020). Ultra short-term power load forecasting based on combined LSTM-XGBoost model. *Power System Technology*, *44*(2), 614-620.

Chu, Y., Xu, P., Li, M., Chen, Z., Chen, Z., Chen, Y., & Li, W. (2020). Short-term metropolitan-scale electric load forecasting based on load decomposition and ensemble algorithms. *Energy and Buildings*, *225*, 110343.

Dai, Y., & Zhao, P. (2020). A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Applied Energy*, *279*, 115332.

Deng, D., Li, J., Zhang, Z., Teng, Y., & Huang, Q. (2020). Short-term electric load forecasting based on EEMD-GRU-MLR. *Power System Technology*, *44*(2), 593-602.

Guo, Z., Zhou, K., Zhang, X., & Yang, S. (2018). A deep learning model for short-term power load and probability density forecasting. *Energy*, *160*, 1186-1200.

Hafeez, G., Alimgeer, K. S., & Khan, I. (2020). Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Applied Energy*, *269*, 114915.

Hafeez, G., Khan, I., Jan, S., Shah, I. A., Khan, F. A., & Derhab, A. (2021). A novel hybrid load forecasting framework with intelligent feature engineering and optimization algorithm in smart grid. *Applied Energy*, *299*, 117178.

Https: //www.emcsg.com/MarketData/PriceInformation.

Jiang, P., Li, R., Liu, N., & Gao, Y. (2020). A novel composite electricity demand forecasting framework by data processing and optimized support vector machine. *Applied Energy*, *260*, 114243.

Jnr, E. O. N., Ziggah, Y. Y., & Relvas, S. (2021). Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting. *Sustainable Cities and Society*, *66*, 102679.

Jnr, E. O. N., Ziggah, Y. Y., & Relvas, S. (2021). Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting. *Sustainable Cities and Society*, *66*, 102679.

Ju, L., Huang, L., Lin, H., Li, H., & Tan, Z. (2021). An interactive dispatching strategy for micro energy grids considering multi-energy flexible conversion based on the three-level optimization perspective. *Sustainable Cities and Society*, *64*, 102504.

Lahouar, A., & Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, *103*, 1040-1051.

Liang, Y., Niu, D., & Hong, W. C. (2019). Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy*, *166*, 653-663.

Nie, H., Liu, G., Liu, X., & Wang, Y. (2012). Hybrid of ARIMA and SVMs for short-term load forecasting. *Energy Procedia*, *16*, 1455-1460.

Nie, Y., Jiang, P., & Zhang, H. (2020). A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting. *Applied Soft Computing*, *97*, 106809.

Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2021). Multi-Model Generative Adversarial Network Hybrid Prediction Algorithm (MMGAN-HPA) for stock market prices prediction. *Journal of King Saud University-Computer and Information Sciences*.

Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. (2017). Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, *54*, 246-255.

Qiu, X., Zhang, L., Suganthan, P. N., & Amaratunga, G. A. (2017). Oblique random forest ensemble via least square estimation for time series forecasting. *Information Sciences*, *420*, 249-262.

Ribeiro, G. T., Mariani, V. C., & dos Santos Coelho, L. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Engineering Applications of Artificial Intelligence*, *82*, 272-281.

Rosato, A., Panella, M., Andreotti, A., Mohammed, O. A., & Araneo, R. (2021). Two-stage dynamic management in energy communities using a decision system based on elastic net regularization. *Applied Energy*, *291*, 116852.

Trizoglou, P., Liu, X., & Lin, Z. (2021). Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines. *Renewable Energy*, *179*, 945-962.

Wang, K., Xu, C., Zhang, Y., Guo, S., & Zomaya, A. Y. (2017). Robust big data analytics for electricity price forecasting in the smart grid. *IEEE Transactions on Big Data*, *5*(1), 34-45.

Wang, Y., Wang, L., Yang, F., Di, W., & Chang, Q. (2021). Advantages of direct input-to-output connections in

neural networks: The Elman network for stock index forecasting. *Information Sciences*, *547*, 1066-1079.

Wang, Z., Zhou, X., Tian, J., & Huang, T. (2021). Hierarchical parameter optimization based support vector regression for power load forecasting. *Sustainable Cities and Society*, *71*, 102937.

Wu, F., Cattani, C., Song, W., & Zio, E. (2020). Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting. *Alexandria Engineering Journal*, *59*(5), 3111-3118.

Wu, Z., Zhao, X., Ma, Y., & Zhao, X. (2019). A hybrid model based on modified multi-objective cuckoo search algorithm for short-term load forecasting. *Applied Energy*, *237*, 896-909.

Yang, Y., Che, J., Deng, C., & Li, L. (2019). Sequential grid approach based support vector regression for short-term electric load forecasting. *Applied Energy*, *238*, 1010-1021.

Yu, X. M., Feng, W. Z., Wang, H., Chu, Q., & Chen, Q. (2020). An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese Q&A system. *Soft Computing*, *24*(8), 5831-5845.

Zhang, J., Tan, Z., & Wei, Y. (2020). An adaptive hybrid model for short term electricity price forecasting. *Applied Energy*, *258*, 114087.

Zhao, H., & Guo, S. (2016). An optimized grey model for annual power load forecasting. *Energy*, *107*, 272-286.

Zheng, R.C., Gu, J., & Jin, Z. J. (2020). Research on short-term load forecasting variable selection based on fusion of data driven method and forecast error driven method. *Proceedings of the CSEE*, *40*(2), 487-499.

Zheng, X., & Chen, W. (2021). An attention-based bi-LSTM method for visual object classification via EEG. *Biomedical Signal Processing and Control*, *63*, 102174.

Zhuang J., Yang G., & Zheng H. (2021). CNN-LSTM-XGBoost short-term power load forecasting method based on multi model fusion. *Electric Power*, 54(5), 46-55.