Short-term load forecasting model based on multi-model integration

Hongwei Wang, Yuansheng Huang, Mengshu Shi *and Shijian Liu
North China Electric Power University, Beijing, China

*Corresponding author e-mail: shimengshu1227@foxmail.com

Abstract. Artificial intelligence and machine learning methods have gradually matured and have been widely used in short-term power load forecasting. In order to make better use of the advantages of different artificial intelligence prediction models and traditional prediction models and improve prediction accuracy, this paper proposes a short-term load prediction model based on multi-model stacking. Different from the combined prediction method, the model first uses three machine learning models, support vector machine (SVM), back propagation neural network (BPNN), and extreme learning machine (ELM) as the base learners, and uses different training data sets. Training is performed on the model, and then the prediction results of the three basic learners are used as the input of Gaussian Process Regression (GPR), and multiple models are integrated to obtain the final prediction result. In order to verify the effectiveness of the Stacking prediction model, this paper applies the short-term load data of the PJM market to this model. Compared with the three base learners, the prediction results show that the model can make full use of the advantages of different prediction models and effectively reduce Forecasting errors have practical significance for solving short-term load forecasting problems.

1. Introduction
Power load forecasting is the basis for the rational planning and dispatching of power systems. Accurate short-term load forecasting can not only reduce the occurrence of mismatches between planned and actual power generation, but also further improve the reliability of the power system. Because the short-term load is greatly affected by holidays, weather and other factors, it has the characteristics of uncertainty and randomness [1]. Therefore, power systems need more effective and accurate short-term load forecasting models to ensure the safe and reliable operation of power [2-4].

The current short-term load forecasting models can be roughly divided into two categories: traditional forecasting models and artificial intelligence forecasting models. Traditional prediction models include time series method [5], regression analysis method [6] and so on. Although the theoretical system of this type of prediction model is relatively mature, it is difficult to effectively represent the non-linear relationship between the data in the process of processing a large amount of load data [7]. Artificial intelligence technology has developed rapidly in recent years, providing a new solution to the forecasting problem, and is widely used in new energy output forecasting and load forecasting [8-9]. Artificial intelligence-based prediction models rely on the training of a large amount of load data and the powerful computing power of modern computers. Compared to traditional
algorithms, machine learning can discover non-linear mapping between traditional prediction models and data that are difficult to find through human experience. Relationship to find the best prediction. The most classic artificial intelligence prediction model is the Back Propagation Neural Network (BPNN). The characteristic of BPNN is to use the error back-propagation algorithm in the training process to continuously adjust the weights and thresholds of the neural network to approximate the optimal solution. Therefore, BPNN provides an effective forecasting tool for load forecasting problems [10-11]. Extreme learning machine (ELM) is a special model of neural network. Its main feature is that the weights and thresholds of the neural network can be given randomly or artificially, and no adjustment is needed. The learning process only needs to calculate the output weight [12-13]. Compared with other methods, ELM has the advantage of high learning efficiency. Support vector machine (SVM) is an artificial intelligence prediction model based on statistical theory and the principle of structural risk minimization [14-15]. SVM uses kernel functions to map the training data from the original space to a higher-dimensional space, so that the SVM can better represent the relationship between variables in the high-dimensional space. Gaussian Process Regression (GPR), as a traditional prediction method, is characterized by using the Gaussian process to perform a priori, and further regression analysis of the data. In complex nonlinear problems, GPR has good adaptability and generalization [16-17].

However, the above literature only uses a single artificial intelligence forecasting model for load forecasting. In the actual load forecasting work, the forecasting model needs to forecast the load data of different seasons, and a single model may cause different data sets due to randomness there are large differences in prediction performance. Therefore, the literature [18-19] tried to give weight to the prediction results of multiple prediction models and obtain the final weighted result by combining prediction methods, thereby improving the prediction accuracy. However, the methods of combined prediction mostly use the method of calculating the mean to calculate the prediction mean of multiple algorithm models or different parameter models of the same type of algorithm. It is difficult to comprehensively utilize the higher differences between different prediction algorithms, and the algorithms cannot achieve complementary advantages. In addition, this combination method lacks sufficient theoretical support and the principle is relatively simple. It is difficult to make full use of the advantages of different prediction models to obtain more satisfactory results.

To overcome the limitations of the above method, this paper proposes a short-term load forecasting model based on multi-model stacking. Stacking integration method is a kind of ensemble learning technology, but different from bagging and boosting, which are two homogeneous ensemble learning technologies, Stacking algorithm can integrate different prediction models to improve the effect of machine learning [20-21]. Compared with a single model, the Stacking algorithm can provide better prediction results [22]. At the same time, the Stacking algorithm is different from the combined prediction model that simply adds weights to the prediction results. It relies on the differences of different prediction models to ensure the diversity of the base learners, and integrates different bases in the best way through the secondary learners. The prediction result of the learner [22]. This paper uses three single artificial intelligence prediction models, namely BPNN, SVM, and ELM as the base learners. In the framework of the integrated model, this paper considers the data observation space of multiple models and uses the secondary learner GPR to establish a multi-model fusion Load forecasting model, and used in the short-term load data of winter and summer in the PJM power market. The results of load forecasting show that the model can take advantage of different forecasting models to obtain more accurate forecasting results, and it has a good application effect in load forecasting problems in different seasons.
2. Forecasting theory and method

2.1. Single prediction model

2.1.1. Support vector regression. Support vector regression (SVR) is a small sample statistical learning theory that uses the principle of minimizing structural risk. The principle is as follows: Let \(\{(x_i, y_i), (x_2, y_2), \ldots, (x_n, y_n)\} \in (x \times y), x_i \in x \subset \mathbb{R}^n, y_i \in y \subset \mathbb{R}\) be the input vector and \(y_i \in y \subset \mathbb{R}\) be the output.

When the sample sets satisfy a linear relationship, the problem of support vector machines comes down to the optimization problem:

\[
\min_{\omega, b} \frac{1}{2} ||\omega||^2 + C \sum_{i=1}^{l} (\zeta_i^+ + \zeta_i^-)
\]

\[
s.t. (\omega \phi(x_i) + b - y_i) \leq \zeta_i^- + \epsilon, i = 1, 2, \ldots, l
\]

\[
y_i - (\omega \phi(x_i) + b) \leq \zeta_i^+, i = 1, 2, \ldots, l
\]

\[
\zeta_i^+, \zeta_i^- \geq 0, i = 1, 2, \ldots, l
\]

When the data set does not meet the linear relationship, the idea of the support vector machine is to map the original data set through non-linear mapping to perform linear regression in the high-dimensional feature space. The inner product operation on a high-dimensional feature space is a kernel function:

\[
(\phi(x_i), \phi(x_j)) = \phi^2(x_i, x_j).
\]

In this case, you only need to perform kernel function operations on the variables in the original low-dimensional space. At this time, the constraint expression is:

\[
\min_{\omega, b} \frac{1}{2} \sum_{i=1}^{l} (a_i^+ - a_i^-) (\phi(x_i), \phi(x_j)) + C \sum_{i=1}^{l} (a_i^+ + a_i^-) - \sum_{i=1}^{l} y_i (a_i^+ - a_i^-)
\]

\[
s.t. \sum_{i=1}^{l} (a_i^+ - a_i^-) = 0
\]

\[
0 \leq a_i^+ \leq C, 0 \leq a_i^- \leq C
\]

Get the optimal solution \(\hat{a} = (a_1^+, a_1^-, \ldots, a_l^+, a_l^-)^T\).

Calculation \(\hat{b} = y_j - \sum_{i=1}^{l} (a_i^+ - a_i^-) (\phi(x_i), \phi(x_j)) + \epsilon\)

The regression decision function is:

\[
f(x) = \sum_{i=1}^{l} (a_i^+ - a_i^-) (\phi(x_i), \phi(x_j)) + \hat{b}
\]

2.1.2. BP neural network. BPNN is different from other neural networks, and it is a training model of error "inverse push". This method propagates the signal in the forward direction and propagates the error in the backward direction. During forward propagation, the signal propagates backward from the input layer through the hidden layer to the output layer; during backward propagation, the error in some form follows the direction of the decreasing error, and the BP nerve is corrected forward from the output layer through the hidden layer. The connection weight of the network. With continuous learning and training, the parameters are continuously optimized, and the network output error is continuously reduced. In essence, the BP algorithm takes the squared error of the network as the objective function and uses the gradient descent method to calculate the minimum value of the objective function.

2.1.3. Extreme Learning Machine. Extreme Learning Machine (ELM) is a hidden layer feedforward neural network. This algorithm is a simple and efficient unsupervised learning algorithm [3].

The mathematical expression of ELM is:

\[
\sum_{i=1}^{l} \omega^i f(W_i u_i + b_i) = t_i, k = 1, 2, 3 \ldots N
\]
Where: \( W \) is the input weight; \( \omega \) is the output weight; \( f \) is the excitation function; \( h \) is the hidden layer bias value; \( N \) is the total number of samples; \( u_k \) is the input vector; \( l \) is the number of hidden layer nodes; \( t_o \) is the output vector.

| Algorithm | Advantages | Disadvantages |
|-----------|------------|---------------|
| SVR       | 1. No need for network structure selection | 1. Sensitive to missing values |
|           | 2. Solved the problems of over- and under-learning and easy to fall into local minima. | 2. The choice of kernel function is high |
|           |                             | 3. Large memory consumption, difficult to explain |
|           |                             | Slow learning |
| BP        | 1. has non-linear mapping capability | 2. Easy to fall into the local minimization problem |
|           | 2. Has good self-learning ability and adaptability | |
| ELM       | 1. has better training efficiency | 1. Low prediction accuracy |
|           | 2. Strong generalization ability | 2. Stability is not high |

Table 1. Advantages and disadvantages of a single prediction algorithm

For the selection of multiple models, this paper chooses the model with better prediction performance as the base learner. This is because the base model with strong learning ability contributes to the overall prediction effect of the model. Support vector regression (SVR) has unique advantages for solving small sample, non-linear and high-dimensional regression problems. BP neural network has a relatively mature theoretical basis, and can implement any complex non-linear mapping, which is suitable for solving problems with complex internal mechanisms. In addition, the BP neural network has a high degree of self-learning and adaptive capabilities, as well as generalization capabilities. An Extreme Learning Machine (ELM) randomly generates the connection weights between the input layer and the hidden layer, and the thresholds of the hidden layer neurons, and does not need to be adjusted during the training process, only the number of hidden layer neurons It does not require long-term training through large-scale historical data, but only needs to optimize the output weights to minimize the error between the output result and the actual value to ensure the accuracy of the regression prediction. Compared with the traditional BP neural network algorithm, ELM method has fast learning speed and good generalization performance. In addition, support vector regression, BP neural network, and extreme learning machine are widely used methods in power load forecasting. Based on the advantages and applicability of the three models, this paper chooses the three models as the basic learner of the integrated algorithm.

2.2. Load Forecasting Model Based on Multi-model Fusion Stacking Integrated Learning

The ensemble method is one of the main research directions of machine learning. Unlike the traditional traditional method of constructing a single learner through the training set, ensemble learning attempts to construct a set of classifiers and combine the output of different individual classifiers in a certain way. Due to the complexity of data distribution and the learning preferences of traditional learning algorithms, the adaptation range of each traditional learning algorithm is relatively limited, and the generalization performance is relatively low. The ensemble method improves the accuracy and generalization ability by retaining the advantages of different learners while retaining the differences. In addition, the Stacking integration algorithm is constructed using a cross-validation method and has strong robustness.

The idea of ensemble learning is to integrate and combine multiple single learners together, make full use of the advantages of different single learners, and make them jointly complete the prediction task, so as to obtain better prediction results and other performance than a single prediction model. Stacking ensemble learning algorithm, as one of the ensemble learning methods, is characterized by the use of cross-validation during the training phase, which generates training data for the next layer of meta-learners, thereby reducing the risk of overfitting.
The basic idea of the Stacking integration method is to train the first layer classifier with the initial data set, and generate a new data set to train the second layer classifier. The output of the first layer classifier is the input of the second layer classifier, the original features are still used as labels for the new dataset. The Stacking algorithm steps are shown in Figure 1.

The training method of Stacking is to fully use the training results of the layer 1 algorithm in the induction process of the layer 2 algorithm. The layer 2 algorithm can find and correct the prediction errors in the layer 1 learning algorithm. Based on this principle, the model is improved. Precision.

The ensemble learning method combines the advantages of the three base learners selected in this paper, and at the same time minimizes the impact of algorithm disadvantages on the prediction results. In other words, the method of ensemble learning uses the advantages of support vector machines for small samples, non-linearity and high dimensions, and also integrates the adaptive and generalization capabilities of BP neural networks, and integrates the fast learning speed of extreme learning machines and Generalization. Through the ensemble learning method, the stability of the prediction model is improved, the prediction effect of the prediction model is further enhanced, and the accuracy of the prediction result is improved.

In addition to the above-mentioned base learner, this paper uses Gaussian Process Regression (GPR) as a secondary learner for the integrated model, and builds a new prediction model by fusing the three models mentioned above. A Gaussian process is a set of arbitrary random variables with a joint Gaussian distribution, whose properties are determined by the average function and the covariance function. Where and represent the parameters of the mean and covariance functions. GPR can choose different covariance functions. This article uses the Square Index (SE) covariance function:

$$
k(x, x_j; \theta) = \sigma^2_w \exp \left( -\sum_{i=1}^{d} \left( \frac{x_i - x_{i,j}}{2\theta^2} \right)^2 \right)$$  \hspace{1cm} (5)
\[ \theta = \{ \sigma_{\omega}, l_1, \cdots, l_r \} \] contains all parameters and is a measure of the total variance of the implicit function. This is a proportional parameter that controls the degree of reduction associated with an increase in the input dimension.

The stacking integrated learning principle of the multi-model fusion established in this paper is shown in Figure 2. This article selects the SVR, BPNN, and ELM described above as the base learners of the Stacking integrated learning algorithm, and trains on the respective divided data sets to obtain preliminary prediction results. The prediction results of the three prediction models are used as the input of the meta-learner GPR, and Gaussian process regression training is performed to obtain the final prediction result.

2.3. Error index

This article uses two indicators as the evaluation criteria for the assessment and verification results, namely the average relative error index (mean absolute percentage error, MAPE) and the average absolute error index (mean absolute error, MAE). The principle is as follows:

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x(i) - y(i)}{x(i)} \right| \times 100\%
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |x(i) - y(i)|
\]

Where \( x(i) \) and \( y(i) \) are the actual and predicted values at \( i \) time, respectively, and \( n \) is the number of samples.

\textbf{Figure 2.} Multi-model fusion of stacking integrated learning
3. Example analysis

3.1. Data sources
The experimental data in this article is derived from the PJM power market. The content of this study is based on PJM's summer hourly load data from June 22 to September 14 as training data and the 24-hour load data on September 15 as the training data. Test data for verification. Considering the seasonal factors, this paper also selects PJM's winter hourly load data from December 22 to March 14 as training data, and uses the 24-hour load data on March 15 as test verification.

For the input of the prediction model, the load history information often has a greater impact on the prediction target, and the historical load information features that are closer to the prediction target are more important. At the same time, temperature information and holidays have a greater impact on the power load. Therefore, the historical load historical information, temperature and holiday information are selected as the input of the prediction model.

3.2. Parameter selection
During the training of the data set, Stacking integrated learning uses a cross-validation method to select the optimal parameter set of each model. In this paper, the kernel function of support vector regression (SVR) is set as the radial basis function. The initial penalty coefficient and kernel function coefficient are set to 1 according to the original toolkit program. The combination of K-cross validation and grid search is used. The method optimizes the parameters. Finally, in this paper, the optimized penalty coefficient is set to 0.364, and the kernel function coefficient is set to 1. For the setting of BP neural network parameters, this paper adopts a "broad strategy", the core of which is to simplify and monitor, and gradually increase the number of layers and the number of neurons in the network. After multiple parameter adjustments, the number of hidden layers in the neural network is set to 2, and the number of neurons in each hidden layer is 10 and 5, respectively. Since the Extreme Learning Machine (ELM) has no hyperparameters, the weights and thresholds of the neural network take a randomly given form.

3.3. Comparative analysis of experimental results
In order to avoid the influence of seasonal factors, the load of PJM summer and winter in different months were selected for training and prediction analysis. Figures 3 and 4 show the comparison between the prediction results of summer and winter load data and the real values using the prediction results of a single prediction model, namely BPNN, SVR, and ELM.

![Figure 3. Test data fitting results in summer](image1)

![Figure 4. Test data fitting results in winter](image2)

According to the fitting results of Fig. 3 and Fig. 4, it can be found that the single method has different prediction effects on load data in different seasons. Support vector regression has the largest overall error, and the prediction of volatility points in the prediction result is relatively slow, only achieving consistency in the trend. In addition, by comparing the prediction results of summer and winter test data,
it can be seen that the prediction error of support vector regression gradually increases with time. Although the overall fitting effect of the BP neural network is better, its prediction results are not stable, and large errors still occur at some prediction points. The overall fitting effect of the extreme learning machine is average, and the stability of the prediction result is poor.

Affected by factors such as seasons, the power load shows a certain degree of fluctuation, and the algorithm's adaptability requirements are high. In the actual load forecasting work, the adaptability of a single intelligent algorithm is poor. As shown in the figure above, due to the randomness and other factors, the single intelligent algorithm has a large difference in prediction performance on different data sets.

In order to solve the problem of single model prediction, many scholars have proposed a combination prediction method. Therefore, this paper also introduces the combination prediction algorithm for further comparative analysis. This paper adopts the method of linear combination forecasting for power load forecasting. Among them, the traditional linear combination prediction method uses the prediction error square sum method. The summer load forecast error is shown in Table 2, and the winter load forecast result error is shown in Table 3.

### Table 2. Summer load forecasting error

| Predictive model    | MAPE | MAE  |
|---------------------|------|------|
| SVR                 | 1.80%| 67.69|
| BPNN                | 1.52%| 57.28|
| ELM                 | 1.58%| 57.89|
| Combination forecast| 1.48%| 55.78|
| **Stacking**        | **1.33%**| **50.29**|

### Table 3. Winter load forecasting error

| Predictive model    | MAPE | MAE  |
|---------------------|------|------|
| SVR                 | 2.31%| 89.21|
| BPNN                | 1.70%| 65.69|
| ELM                 | 1.79%| 68.04|
| Combination forecast| 1.76%| 66.48|
| **Stacking**        | **1.57%**| **60.66**|

In order to take advantage of different prediction models, multi-model fusion is achieved. Based on the prediction results of different single artificial intelligence models, this paper uses the stacking integrated learning model to fit and train the summer and winter load data, and compares the results with three single prediction models and traditional linear combination prediction methods. Compare the errors.

By comparing the stacking integrated learning model with a single intelligent algorithm and a traditional linear combination prediction method, and comparing the results of MAPE and MAE of different season data, it can be clearly seen that the integrated algorithm has a more obvious advantage of prediction accuracy. In terms of prediction error, the prediction result of a single intelligent algorithm is not stable, the prediction error of support vector regression is the largest, and the prediction result of BP neural network performs well. Due to the different effects of a single intelligent algorithm, it is difficult to choose a suitable model for different data. The traditional linear combination prediction method can improve the prediction performance to some extent, but the stability is poor. Although the combined forecasting effect is better than the single model in summer load forecasting, its forecasting effect is not superior to the single model in winter load forecasting. This is because the combination forecast requires artificially set weights and has uncertainty.

The stacking integrated learning model has stronger self-learning ability, which shows stronger accuracy and applicability, and can get the best accuracy in different data sets.
4. Conclusion
This paper proposes a Stacking integrated learning model that integrates three machine learning models: SVR, BPNN, and ELM. This model makes full use of the prediction advantages of different machine learning models, and solves the problem of prediction uncertainty of a single machine learning model. This paper further validates the model and applies it to short-term power load forecasting. By forecasting the load data of the PJM market, the experimental results show that compared with a single model, the integrated model reduces the average relative error in summer load prediction by 0.19% -0.47%, and the average absolute error decreases by 6.99-17.4; in winter In load forecasting, the average relative error decreased by 0.13% -0.74%, and the average absolute error decreased by 5.03-28.55. Compared with the combined forecasting model, the relative error of the integrated model in summer load forecasting was reduced by 0.15% and the absolute error was reduced by 5.49; in the winter load forecasting, the relative error was reduced by 0.19%, and the absolute error was reduced by 5.82. Numerical examples show that the integrated model has obtained good prediction results, proving the practicability and rationality of the model. The next research in this paper will realize the integration of more machine learning models to further improve the accuracy of short-term load prediction.

Acknowledgments
This work was financially supported by Hebei Science and Technology Plan Project: Research on Energy Economic Development Strategy (No. 12457206D-22).

References
[1] HUANG Nantian, QI Bin, LIU Zuoming, et al. Probabilistic Short-term Load Forecasting Using Gaussian Process Regression with Area Grey Incidence Decision Making [J]. Automation of Electric Power Systems, 2018, 42(23): 64-71.
[2] KANG Chongqing, XIA Qing, LIU Mei. Power System Load Forecasting [M]. China Electric Power Press, 2007.
[3] LI Peng, HE Shuai, HAN Pengfei, ZHENG Miaomiao, HUANG Min, SUN Jian. Short-Term Load Forecasting of Smart Grid Based on Long-Short-Term Memory Recurrent Neural Networks in Condition of Real-Time Electricity Price [J]. Power System Technology, 2018, 42(12): 4045-4052.
[4] López M, Sans C, Valero S, et al. Classification of Special Days in Short-Term Load Forecasting: The Spanish Case Study[J]. Energies, 2019, 12(7): 1253.
[5] LU Jixiang, ZHANG Qipei, YANG Zhihong, TU Mengfu, LU Jinjun, PENG Hui. Short-term Load Forecasting Method Based on CNN-LSTM Hybrid Neural Network Model[J]. Automation of Electric Power Systems, 2019, 43(08): 131-137.
[6] SHI Wenqing, WU Kaiyu, WANG Dongxu, WANG Di. Eclectic Power System Short-Term Load Forecasting Model Based on Time Series Analysis and Kalman Filter Algorithm [J]. Techniques of Automation and Applications, 2018, 37(09): 9-12+23.
[7] PENG Wen, WANG Jinrui, YIN Shanqing. A short-term load forecasting model based on Attention-LSTM in electricity market [J]. Power System Technology, 2019, 43(05): 1745-1751.
[8] TANG Qingfeng, LIU Nian, ZHANG Jianhua, YU Zhuangzhuang, ZHANG Qingxin, LEI Jinyong. A Short-Term Load Forecasting Method for Micro-Grid Based on EMD-KELM-EKF and Parameter Optimization [J]. Power System Technology, 2014, 38(10): 2691-2699.
[9] CHEN Jie, SHEN Yanxia, LU Xin, JI Zhicheng. An Intelligent Multi-Objective Optimized Method for Wind Power Prediction Intervals [J]. Power System Technology, 2016, 40(08): 2281-2287.
[10] SU Xueneng, LIU Tianqi, CAO Hongqian, JIAO Huiming, YU Yaguang, HE Chuan, SHEN Ji. A Multiple Distributed BP Neural Networks Approach for Short-term Load Forecasting Based on Hadoop Framework [J]. Proceedings of the CSEE, 2017, 37(17): 4966-4973+5216.
[11] Yang Z, Wang J. A hybrid forecasting approach applied in wind speed forecasting based on a
A data processing strategy and an optimized artificial intelligence algorithm [J]. Energy, 2018, 160: 87-100.

[12] Wu J, Cui Z, Chen Y, et al. A new hybrid model to predict the electrical load in five states of Australia[J]. Energy, 2019, 166: 598-609.

[13] Lu Fangcheng, LIU Yi, QI Yanxun, YAN Yuehao, ZHANG Jiantao, XIE Qing. Short-term Load Forecasting Based on Optimized Learning Machine Using Improved Genetic Algorithm. Journal of North China Electric Power University, 2018,45(06): 1-7.

[14] Li Y, Che J, Yang Y. Subsampled support vector regression ensemble for short term electric load forecasting [J]. Energy, 2018, 164: 160-170.

[15] HE Yaoyao, LIU Rui, HAN Aoyang. Short-Term Power Load Probability Density Forecasting Method Based on Real Time Price and Support Vector Quantile Regression [J]. Proceedings of the CSEE, 2017, 37(03): 768-776.

[16] PENG Hongqiao, GU Jie, HU Yu, et al. Forecasting Model for Saturated Load Based on Chaotic Particle Swarm Optimization-Gaussian Process Regression [J]. Automation of Electric Power Systems, 2017, 41(21): 25-32

[17] SUN Bin, YAO Haitao, LIU Ting. Short-term Wind Speed Forecasting Based on Gaussian Process Regression Model [J], Proceedings of the CSEE, 2012, 32(29): 104-109.

[18] Huang Y, Liu S, Yang L. Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM [J]. Sustainability, 2018, 10(10): 3693.

[19] Su Limin, Song Yanhong, He Huishuang. Variable Weight Combination Forecasting Method Considering Weight Uncertainty [J/OL]. Statistics and Decision, 2019(11):60-63 [2019-05-29].

[20] Lu Jixiang, Zhang Qipei, Yang Zhihong, et al. Short-term load prediction method based on CNN-LSTM hybrid neural network model [J]. Automation of Electric Power Systems, 2019, 43 (8): 131-137 [19] Divina F, Gilson A, Goméz-Vela F, et al. Stacking ensemble learning for short-term electricity consumption forecasting[J]. Energies, 2018, 11(4): 949.

[21] XU Yanlu, ZHANG Jiansen, JI Xing, WANG Binbin, DENG Zhuofu. Research on Short-Term Load Forecasting Method Based on Multi-Model [J]. Control Engineering of China, 2019, 26(04): 619-624.

[22] Shi Jiaqi, Zhang Jianhua. Load forecasting method based on multi-model fusion Stacking integrated learning method [J]. Chinese Journal of Electrical Engineering, 2019, 39 (14): 4032-4042.

[23] Feng C, Cui M, Hodge B M, et al. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting[J]. Applied Energy, 2017, 190:1245-1257.J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2000) 51-59.