Research on the combined predicting model of short-term load in smart grid

Liping Sun\textsuperscript{1,a}, Lifang Wang\textsuperscript{1,*}, Siwei Jing\textsuperscript{1,b}, Zhou Peng\textsuperscript{1}, Fangling Yao\textsuperscript{2}

\textsuperscript{1}School of Management, Northwestern Polytechnical University, Xi’an, China
\textsuperscript{2}School of Economics, Xi’an University of Finance and Economics, Xi’an, China

*Corresponding author e-mail: lifang@nwpu.edu.cn, \textsuperscript{a}sliping@mail.nwpu.edu.cn, \textsuperscript{b}jingsiwei44@163.com

Abstract. As an important basis of power grid planning and dispatching, short-term load predicting model with high accuracy is very important to ensure the efficient and reliable operation of power grid. In this paper, the influencing factors of the short-term load of smart grid are determined by the method of grey correlation analysis. BP, RBF and Elman neural network construct the single prediction model of the short-term load of smart grid. The single prediction model is weighted by GA genetic algorithm, and the combined prediction model of the short-term load of smart grid is constructed and verified by an example. The results show that the error of the combined prediction model can be kept at about 0.4%, which has higher prediction accuracy and stability.

1. Introduction
With the construction of smart grid and the development of energy Internet strategy, smart grid demonstration areas have been successively built in major cities in China. Various demonstration projects generally achieve two-way interaction with users through the combination of various communication and control technologies to promote the development of electricity market. The continuous improvement of grid operation efficiency and security \cite{1}, and a large number of new energy equipment are put into the grid. The power dispatching and operation are also developing towards the direction of high efficiency and intelligence. The power market on the power selling side and the power generation side is gradually opening up. The power supply mode and price mechanism are determined by the independent consultation of users and power generation enterprises \cite{2}. As the power, supply and demand need to maintain the instantaneous balance, how to provide sufficient transmission capacity in the market-oriented operation of smart grid to ensure the effective trading of power market. How to prepare transmission planning, coordinate generation planning, and promote the investment and construction of power generation? How to realize the intelligent dispatching of power grid, on the basis of ensuring that the power supply and demand remain relatively stable within a certain range to meet the overall operation economy and security of the power grid? These are the important problems that the power grid companies are facing \cite{3}, and the important basis for solving these problems is the accurate short-term load predicting results \cite{4}.

At present, the short-term load predicting method commonly used by power grid companies mainly uses historical load data for analysis and calculation. There are many problems such as large calculation amount, long time, and inability to consider the impact of environmental changes on load,
which can no longer meet the higher prediction accuracy requirements of smart grid short-term load [5]. Due to its good scheduling ability, artificial neural network has been widely concerned in the short-term prediction of power system, which is one of the hot spots of current scholars' research, and one of the practical prediction methods widely recognized in the world [6]. It has strong self-adaptive and self-organizing characteristics, does not need the prior knowledge of power load data, has high fault tolerance rate for sample data, strong nonlinear mapping ability, has higher accuracy than the traditional prediction method [5], and can consider a variety of influencing factors [7].

The construction of smart grid in China started relatively late. At present, there are relatively few researches on system level short-term load predicting under this condition. The new influencing factors and load characteristics that need to be considered under the condition of smart grid are the research hotspot of scholars at home and abroad [8]. Mei Zhen believes that it is an important challenge for future research to consider the operation environment under the power market conditions in the study of smart grid load predicting; and realize the understanding of power generation and consumption demand in advance through the mathematical expression of market participants' decision-making; and realize the probabilistic load predicting more in line with the engineering demand [9]. Through clustering and genetic algorithm, Wei Qi optimizes BP neural network model to predict short-term load considering real-time electricity price, which improves the accuracy of prediction and shortens the training time [10]. Zhang Zhisheng comprehensively considered the response behavior of demand response participants and the power and environmental factors involved in demand response and expressed them digitally. On this basis, he built a short-term load-predicting model by using radial basis function neural network, which provided a theoretical basis for the load predicting method in the implementation of demand response in the power market [11].

In this paper, the gray correlation analysis method is used to extract the influencing factors of short-term load of smart grid and improve the prediction accuracy of the single short-term load prediction model of power grid. Based on the neural network method, the single short-term load prediction model of suitable power grid is constructed, and then the combined short-term load prediction model of smart grid is constructed by optimizing the weight of each model with GA genetic algorithm.

2. Extraction of factors affecting short-term load of smart grid

2.1. Analysis of factors influencing short-term load of smart grid

The determination of short-term load influencing factors of smart grid is the basic condition of short-term load predicting. Accurate extraction of load influencing factors is conducive to the realization of efficient short-term load predicting. Theoretically, the factors affecting the short-term load of power grid can be classified into three categories: time factor, weather factor and random factor [5]. The time factor mainly considers the date type factor and the time point sequence factor, among which the date type mainly refers to the working day, the weekend, the holiday and the season. Weather factors generally refer to various indexes involved in meteorological statistics, such as temperature, humidity, dew point, precipitation, wind force and visibility [12]. Random factors mainly refer to the occurrence of unforeseen conditions, such as power failure caused by sudden equipment failure.

To sum up, this paper selects 10 kinds of short-term load predicting influencing factors and codes their characteristics as the input of load predicting model, as shown in Table 1.
### Table 1 Description of influencing factors for short-term load predicting of smart grid

| No. | influence factors | description            | No. | influence factors | description      |
|-----|------------------|------------------------|-----|------------------|------------------|
| 1   | hours            | 24hr                   | 7   | rain             | 1                |
| 2   | minutes          | 60 minutes             | 7   | no rain          | 0                |
| 3   | temperature      | °C                     | 8   | workday          | 0                |
| 4   | dew point        | °C                     | 8   | day off          | 1                |
| 5   | wind power       | Km/h                   | 9   | load value at the previous day | MW         |
| 6   | visibility       | Km                     | 10  | load value at last moment | MW         |

#### 2.2. The selection of influencing factors based on grey correlation method

The short-term load of smart grid is affected by a variety of factors, the change rules of various loads are different, and the amount of data is large, so it is difficult to determine the specific influencing factors through simple index and curve analysis. In this case, the grey correlation analysis method has a high applicability. The data of each factor and load that changes with time are set as two kinds of data series. The grey correlation method is used to calculate the correlation between each influencing factor and load data, to know the influence of each factor on load quantitatively.

The process of analyzing the influencing factors of short-term load of smart grid by using grey correlation method is as follows:

1. Determination of data collection and analysis sequence.
2. The collected data were supplemented with missing values and standardized.
3. Use the formula to calculate the correlation coefficient. The details are as follows:

\[
\zeta_k(k) = \frac{\min \min |x_i(k) - x_j(k)| + \rho \cdot \max \max |x_0(k) - x_i(k)|}{|x_0(k) - x_j(k)| + \rho \cdot \max \max |x_0(k) - x_j(k)|}
\]  

\(\rho \in (0, \infty)\) is called the resolution coefficient. The smaller the value is, the greater the difference between the correlation coefficients is, and the greater the resolution is. \(\zeta(k)\) is a positive number greater than 0 but less than 1, which can reflect the degree of association between the factor sequence \(X_i\) and the reference sequence \(X_0\) on the \(k\)-th attribute. The larger the value is, the higher the similarity of variation between the factor and load is, and vice versa.

4. Calculate the average value to get the final correlation degree.
5. According to the final correlation degree calculated above, the comprehensive evaluation results are obtained. The higher the correlation degree is, the more similar the evaluation object is to the reference series, and the better the evaluation result is.

#### 3. Construction of short-term load combination predicting model for smart grid

In the neural network method, the advantages and disadvantages of different network structures and types of neural network models are different. In order to make full use of the advantages of each network structure, this paper selects the BP neural network with global characteristics and the RBF neural network with local characteristics from the feedforward neural network based on the characteristics of model algorithm. It selects the BP neural network with local characteristics from the feedback neural network there are three kinds of Elman neural networks with dynamic modeling characteristics to establish a single short-term load-predicting model.
3.1. Construction of a single prediction model based on neural network

(1) Construction of BP neural network prediction model

BP network and its changing form are the most widely used neural network at present. It is composed of input layer, middle layer and output layer, and the middle layer can be expanded to multi-layer. The greatest advantage of BP neural network lies in its highly nonlinear mapping function. Theoretically, the neural network can approach a nonlinear function with arbitrary accuracy by adjusting the number of nodes in the middle layer. It has strong learning and analysis ability for the data of the input network, and has high adaptability for the nonlinear characteristics of short-term load data. When using BP neural network for short-term load predicting, the training parameters need to be set, that is, the allowable maximum training steps of the network model, the training target error and the learning rate. The maximum number of training steps can set a threshold value for the training process of the model to avoid that the network training is always in the calculation state without satisfactory results.

(2) Construction of RBF neural network prediction model

From the theoretical point of view, RBF neural network is the best performance neural network in feed forward network. It has three-layer network structure. The input layer is responsible for the transmission of data information, and the output layer is responsible for the calculation of data information and the transmission of network structure. This kind of calculation belongs to linear calculation. The hidden layer, as the most important structure, is responsible for the realization of the intermediate transmission function the transfer function is radial basis function, and the number of neurons in the hidden layer can be set according to the attributes of the problem. Its advantages are as follows:

1) Compared with BP neural network, RBF neural network only affects the input around the input weight vector in self-learning and calculation, and its approach to the target is in the local range. The function approximation has the only best and fast characteristics. In short, the closer a test set sample is to a training set sample, that is, the smaller the Euclidean distance is, the larger the output value will be after the RBF function, avoiding the occurrence of local optimum in BP network easily;

2) Because RBF neural network can divide the whole training into parts when training the network. In load predicting, although the number of neurons is larger than BP neural network, it is better than BP neural network in terms of the speed of network training and the time and effect spent overall. It can complete the network training when there are many input sample data vectors.

When using RBF neural network for short-term load predicting, several parameters need to be determined, which are training target error setting, spread value (expansion speed of radial basis), maximum number of neurons and network parameters added each time.

(3) Construction of Elman neural network prediction model

Elman neural network is based on the structure of BP neural network, which has four layers. The input layer is responsible for the transmission function of data information. The output layer is responsible for the linear calculation of data information and the function of transmitting it out of the network structure. The hidden layer is responsible for the intermediate transmission function, and the function describing this transmission can select linear or non-linear function expression according to different problems. The receiving layer uses the memory function to input the previous output results of the hidden layer Line record, and input it into the hidden layer again for calculation verification. Its advantage is that it has the characteristics and advantages of dynamic neural network, that is, it can reflect the dynamic characteristics of using the network modeling. In addition, because it belongs to the optimization results of BP neural network, its calculation performance and prediction stability are better than BP neural network.

When using Elman neural network for short-term load predicting, firstly, determine the input sample data and output sample data of neural network, the input sample data is mainly used as neural network self-learning data, the output data is used for the detection and comparison of network training results, and the error of each training result and output data comparison enters the next calculation process; secondly, in the network In the calculation process, the errors calculated in the
previous step are fed back to the next generation to modify the network parameters. Finally, the results that meet the requirements of network training errors are generated by repeated calculation in the first step and the second step.

3.2. Construction of combined predicting model based on Neural Network

(1) Weight determination of prediction model based on GA genetic algorithm

GA genetic algorithm can simulate the evolutionary characteristics of organisms, consider the prediction effect of each time point, calculate and optimize generation by generation, find the weight solution to minimize the prediction error in the global scope, and effectively improve the prediction accuracy of the combined model. When using the "Ga" function to calculate the weight of the combined model, first set the parameters, give the upper and lower limits of each weight value and the number of variables to be calculated (in this paper, it is 3, that is, the weight of 3 single models), and use the minimum prediction error to establish the fitness function, that is, the objective function of the optimization problem, as shown in formula 3-1:

\[ F = \frac{1}{n} \sum_{i=1}^{n} |\omega_{i,t} \hat{Y}_{i,t} - L_t| \]  

(3-1)

In the formula: \( n \) is the number of prediction points; \( \hat{Y}_{i,t} \) is the prediction value of the i-th prediction model in the t-th period; \( \omega_{i,t} \) is the weight of the i-th prediction model in the t-th period; \( L_t \) is the actual value of the load in the t phase.

(2) Construction of combined predicting model based on Neural Network

In this paper, the basic principle of the establishment of the combined predicting model is as follows: let m different predicting models be used (i = 1, 2, ..., m) for period n (time, number of samples, t = 1, 2 ..., n) for data predicting, the combined predicting model composed of these m single models is:

\[ Y_t = \sum_{i=1}^{m} \omega_{i,t} \hat{Y}_{i,t} \]  

(3-2)

In the formula: \( Y_t \) is the predicted value of the t-th period of the combined predicting model; \( \hat{Y}_{i,t} \) is the predicted value of the i-th period of the t-th period of the combined predicting model; \( \omega_{i,t} \) is the weight of the i-th period of the t-th period of the combined predicting model.

All weight coefficients are satisfied,

\[ \sum_{i=1}^{m} \omega_{i,t} = 1, \quad \omega_{i,t} \geq 0 \]  

(3-3)

In order to realize the construction of combination model, first, based on the analysis of neural network method, the single prediction model is selected and constructed, and the influence factors extracted quantitatively are used as the model input data to improve the prediction accuracy of single model. On this basis, the combination prediction model is constructed by calculating the weight of each single model with GA genetic algorithm. To sum up, the prediction steps of the combined prediction model proposed in this paper are shown in Figure 1.
4. Example verification of short-term load predicting model of smart grid
Seasonal factors and special holidays have great influence on load change. In order to eliminate these influences and extract the influencing factors of conventional short-term load change, this paper selects the load data of smart grid in Binhai New Area of Tianjin in August 2016 and collects the daily date type and meteorological data as samples for short-term load prediction and analysis of smart grid. Among them, the load data collection interval is 1min, and there are 44640 load data in 31 days. MATLAB was used to carry out grey correlation analysis to quantitatively extract the influencing factors, and the analysis results are shown in Figure 2.

![Figure 1](image1)

**Figure 1** Short term load combination predicting process of smart grid

![Figure 2](image2)

**Figure 2** Correlation of short-term load influencing factors of smart grid
According to the results of grey correlation analysis, the influencing factors of short-term load of smart grid in August 2016 in this region are sorted according to the similarity rule as follows: load value at the same time on the previous day > dew point > load value at the previous time > whether there is rain > date type > temperature > wind force > hour > minute > visibility. The correlation degree of the five factors of the previous day load value, dew point, load value at the previous time, whether there is rain or not and date type is more than 0.8, which shows that dew point is the main factor affecting the load change in this month. In addition, Tianjin usually enters a rainy period in the middle and late August. Whether there is rain has a great influence on the load change. As a whole, in the ranking of correlation degree, the first seven factors have a greater impact on the short-term power load change, with values above 0.5, which can be considered as the main factors when building the load-predicting model for short-term load predicting. The correlation degree of the last three factors is less than 0.4, although it has a certain degree of influence, but the degree is relatively small. In the construction of prediction model, it can be discarded properly for the consideration of prediction time and accuracy.

In order to shorten the training time and improve the training efficiency of the model, 7 kinds of influencing factors data obtained from quantitative analysis are used as the training data of each single load-predicting model. Considering that, this paper studies the short-term load predicting in the smart grid environment, that is, daily load predicting, select 2 / 3 of the original data as the training samples and the rest as the test samples. It takes 21-day data points of the total samples as the training samples for neural network training, the rest load data randomly select 1 day for short-term load predicting, and compare the predicted value with the actual value. In addition, build the neural network. Network prediction model needs to determine the network input data and output data. In this paper, the input data is the influencing factors, and the output data is the load data.

Generally, the gap between the predicted value and the actual value is used to evaluate the prediction result of the model and the quality of the established model. At present, there are more mature error analysis theories and calculation methods. In this paper, MSE, RMSE, MAE, SMAPE and NSE are used as evaluation indexes of short-term load predicting model.

The neural network of each single prediction model is trained by setting the network parameters, and then the short-term load prediction is carried out by using the trained neural network model. Each prediction is randomly selected from the test samples for one day, and each neural network model is respectively predicted for five times to fully compare the prediction effect of different models. The weight coefficients of each combination prediction model calculated by GA genetic algorithm in 5 times of prediction are shown in Table 2, and the comparative analysis between neural network combination prediction model and single neural network model is shown in Table 3, in which cm represents combination model and ENN represents Elman neural network model.

| Prediction Time | $\omega_1$ | $\omega_2$ | $\omega_3$ |
|-----------------|-----------|-----------|-----------|
| T1              | 0.8023    | 0.0162    | 0.1813    |
| T2              | 0.2264    | 0.5892    | 0.1843    |
| T3              | 0.0054    | 0.0163    | 0.9772    |
| T4              | 0.0001    | 0.4049    | 0.5958    |
| T5              | 0.0283    | 0.3266    | 0.6459    |
| Evaluation Indicators | Comparison of Neural Network Prediction Performance | T1 | T2 | T3 | T4 | T5 |
|-----------------------|---------------------------------------------------|----|----|----|----|----|
| SSE                   | BPNN                                              | 104.6118 | 153.5544 | 654.9924 | 245.1084 | 305.7584 |
|                       | RBFNN                                             | 177.7306 | 128.185 | 177.7306 | 177.7306 | 177.7306 |
|                       | ENN                                               | 122.8782 | 224.3782 | 122.8241 | 188.8183 | 188.8183 |
|                       | CM                                                | 103.1223 | 112.7127 | 110.7520 | 124.5871 | 126.5771 |
| RMSE                  | BPNN                                              | 10.2279 | 12.3917 | 25.5928 | 15.6559 | 17.4859 |
|                       | RBFNN                                             | 13.3315 | 11.3219 | 13.3315 | 13.3315 | 13.3315 |
|                       | ENN                                               | 11.0850 | 14.9792 | 11.0826 | 13.7411 | 13.7411 |
|                       | CM                                                | 10.1549 | 10.6166 | 10.5238 | 11.1618 | 11.2506 |
| MAE                   | BPNN                                              | 7.9189 | 9.7764 | 20.8768 | 12.8662 | 14.6350 |
|                       | RBFNN                                             | 10.6036 | 8.9724 | 10.6036 | 10.6036 | 10.6036 |
|                       | ENN                                               | 8.7033 | 12.4679 | 8.7607 | 10.6781 | 10.6781 |
|                       | CM                                                | 7.8628 | 8.0838 | 8.2199 | 8.7075 | 8.7565 |
| SMAPE                 | BPNN                                              | 0.3566% | 0.4329% | 1.0019% | 0.5805% | 0.6857% |
|                       | RBFNN                                             | 0.5041% | 0.3973% | 0.5041% | 0.5041% | 0.5041% |
|                       | ENN                                               | 0.3924% | 0.5608% | 0.4013% | 0.4720% | 0.4720% |
|                       | CM                                                | 0.3543% | 0.3525% | 0.3749% | 0.3935% | 0.3945% |
| NSE                   | BPNN                                              | 0.9986 | 0.9966 | 0.9916 | 0.9968 | 0.9960 |
|                       | RBFNN                                             | 0.9977 | 0.9972 | 0.9977 | 0.9977 | 0.9977 |
|                       | ENN                                               | 0.9984 | 0.9951 | 0.9984 | 0.9975 | 0.9975 |
|                       | CM                                                | 0.9986 | 0.9975 | 0.9985 | 0.9984 | 0.9983 |

It can be seen from the table that when BP neural network is used to predict the 5-day load value respectively, the prediction effect fluctuates greatly, the prediction effect is sometimes good, and the model prediction stability is insufficient. The prediction effect by RBF and Elman neural network is relatively stable, and the 5-day prediction SMAPE value is basically stable within the range of 0.4% - 0.5%. In a word, four kinds of neural network models are used to predict the five-day load value in five times. The results show that the prediction accuracy of the combined model is higher than that of each single prediction model. When the prediction result of the single prediction model is poor and the volatility is strong, the combined model can still ensure the stability of the prediction error and effectively avoid the short-term load prediction of the single prediction model. The shortcomings of the SSE value in the evaluation index of prediction effect can always be maintained within 150, RMSE value is kept around 10, reflecting the high prediction accuracy of the model, MAE value is between 7-8; SMAPE value is stable below 0.4%, indicating that the prediction error is small. The prediction accuracy of the combined model is relatively high. NSE value is always around 0.998, indicating that the combined model has high reliability and good quality.

Figure 3 shows the comparison between the estimated value and the actual value of the combined model T1 in the above prediction. It can be seen from the figure that the actual value and the estimated value are approximately highly coincident. At this time, the average error of the model is small, 7.8628, and the symmetrical average absolute percentage error is 0.35%, indicating that the short-term load combined prediction model of smart grid established in this paper is of high accuracy.
5. Conclusion
The short-term load combined predicting model of smart grid based on neural network and GA is compared with the single Predicting model. The results show that the short-term load combined predicting model of smart grid can well meet the requirements of smart grid operation, has high accuracy of short-term load predicting, and multiple predicting results. The SMAPE value of the combined prediction model in the test is kept at about 0.4%, which indicates that the combined prediction model not only has higher prediction accuracy, but also can overcome the shortcomings of each single model, has higher stability, and the validity and reliability of the model are verified.

Acknowledgements
[1] This study was funded by State Grid Shaanxi Electric Power Company, in 2019 construction of electric energy substitution system --Household investigation and analysis of energy consumption of "coal to electricity" users (No. D5201190001).

References
[1] Zhang Dongxia, Miao Xin, Liu Liping, et al. Research on the development of smart grid big data technology [J]. Chinese journal of electrical engineering, 2015, 35(1): 2-11.
[2] National development and reform commission, national energy administration. Notice on actively promoting market-oriented electricity trading and further improving the trading mechanism. [2018-7-16].
[3] Xu Chengshun. Practical exploration on opportunities and challenges faced by the power market in the smart grid environment [J]. Finance & economics (academic edition), 2019(21): 49-50.
[4] Zhang Yongfeng, Cui Kai. Short-term load forecasting using consensus PSO and truster Tech [J]. Journal of power systems and automation, 2017, 29(7): 7-12.
[5] Zhang Lingyun, Xiao Huiren, Wu Junhao, Liao Qian, Wang Peng. Overview of power system load forecasting [J]. Power big data, 2008, 21(01): 52-56.
[6] Elattar E E, Goulermas J, Wu Q H. Electric Load Forecasting Based on Locally Weighted Support Vector Regression [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2010, 40(4): 438-447.
[7] Wu Qianhong, Han Bei, Feng Lin, Li Guojie, Jiang Xiuchen. Prediction and analysis of smart grid in the era of "artificial intelligence +" [J]. Journal of Shanghai jiaotong university, 2008, 52(10): 1206-1219+1266.
[8] Xue Yusheng, Lai Yining. Fusion of big energy thinking and big data thinking --big data and power big data [J]. Power system automation, 2016, 40(1): 1-8.
[9] Mei Zhen. Short-term load forecasting based on power big data and artificial intelligence [J]. Chinese and foreign architecture, 2018 (10): 215-216.
[10] Wei Qi, Yang Ming. Improved neural network algorithm for short-term load prediction of smart
grid [J]. Journal of Harbin university of science and technology, 2017, 22(4): 65-69.
[11] Zhang Zhisheng, Yu Daolin. Rbf-nn short-term load forecasting model considering comprehensive influencing factors of demand response [J]. 2008, 38(6): 1631-1638.
[12] Huan Jiaja, Zuo Zhengmin, Cheng Xin. Analysis on the influence of external factors on the load characteristics of Guangdong power grid [J]. Hydropower energy science, 2008, 36(05): 210-213.