Research on Parallel Forecasting Model of Short-Term Power Load Big Data

Xin-jia Li[1(✉)], Hong Sun[1], Cheng-liang Wang[1], Si-yu Tao[2], and Tao Lei[3]

[1] Jiangsu Fangtian Power Technology Co., Ltd., Nanjing 210096, China
rending0620@163.com
[2] Southeast University, Nanjing 210000, China
[3] South China Normal University, Guangzhou 510006, China

Abstract. The parallel prediction model of big data with traditional power load has a low prediction accuracy in different working conditions, so the parallel prediction model of big data for short-term power load is designed. The short-term power load forecasting theory is analyzed, and the short-term power load data are classified to select the short-term power load forecasting theory. The Map/Reduce framework is built on the basis of the theory, and the prediction process is designed through the Map/Reduce framework. The short-term power load data of the subnet and the big data of the short term power load are predicted respectively, and the construction of the parallel prediction model of the short-term power load big data is realized. The experimental results show that the proposed big data parallel prediction model is better than the traditional model, and can be switched under different working conditions, and the deviation between the forecasting curve and the actual load is small, the average deviation is 1.7, and the overall prediction effect is good.

Keywords: Short-Term load forecasting · Big data · Electrical load · Prediction algorithm

1 Introduction

With the rapid economic development, the power industry is facing increasing challenges. Among them, power system load forecasting is of great significance to the entire power industry. Power system load forecasting has become a hot topic in scientific research at present. It is to accurately grasp the future load trend, accurately grasp the overall layout of the power system, and maintain a safe and stable operating environment [1]. The tremendous changes in human life and the environment have made the impact of power loads increasingly complex. The existence of prediction error directly increases the additional cost of power system operation, which is not conducive to economic improvement. This paper proposes a parallel prediction model for short-term power load data. To study and analyze the power load forecasting, we propose a short-term power load forecasting scheme based on the Map/Reduce model theory, to establish a sub-network and global load forecasting model, and the parallel

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 126–136, 2019.
https://doi.org/10.1007/978-3-030-19086-6_14

prediction model of the design is analyzed and tested. The test results show that the parallel prediction model proposed in this paper has a very high effectiveness.

2 Short-Term Power Load Forecasting Theory Analysis

2.1 Short-Term Power Load Big Data Classification Analysis

The power load forecasting is based on relevant historical data to restore load characteristics, and it estimates the time load in the future. The study of power load forecasting is of great significance and is not limited to estimating the load trend, the key point is that power load forecasting can be used as a basis for important work such as power grid planning and power dispatch, which effectively improving resource utilization and improving grid environment [2]. Before load forecasting, we must analyze the load characteristics, the load forecasting classification and the prediction steps, then make full preparations for further load forecasting.

The load forecasting content is complex and different divisions are obtained according to different criteria. According to the classification of electricity, it can be divided into agricultural power load, industrial power load and residential electricity load. According to the length of time, the power load forecast can be divided into long-term, medium-term, short-term and ultra-short-term power load forecasting [3]. Among the short-term power load forecasting, the forecasting object is the load at each moment of the day, the characteristic is that it has strong periodicity and is influenced by weather factors. Therefore, the forecasting model must consider the influence of weather factors, This prediction is usually used to assist in the determination of fuel supply for power generation, and to make an advance estimate of the operating power plant to ensure that the unit maintenance plan is properly scheduled. The main forecasting methods include periodic time series forecasting, neural network forecasting, and related predictions that take into account weather factors. Law et al. [4].

2.2 Selected Short-Term Power Load Big Data Prediction Theory

Usually, The main content of load forecasting research is to restore the load characteristic curve based on historical load. However, complex power system load characteristics are also affected by external factors. In the prediction and analysis of power load, it should be comprehensive and comprehensive. Under the joint action of internal and external factors, load forecasting has the characteristics of inaccuracy, condition, time, and method diversity [5].

Electric power load forecasting is of great significance to industry efficiency and even to social and economic development. The research process of load forecasting must be comprehensive and specific, and its implementation must be organized and orderly, and unnecessary error interference must be reduced. The specific forecasting process is mainly divided into the following steps: determining the target to collect data, sorting out the data, data preprocessing, establishing a prediction model, load forecasting, and result analysis [6]. The accurate implementation of each step of these steps will play a crucial role in ultimately obtaining effective prediction results, and each procedure must be strictly monitored. The specific process is shown in Fig. 1.

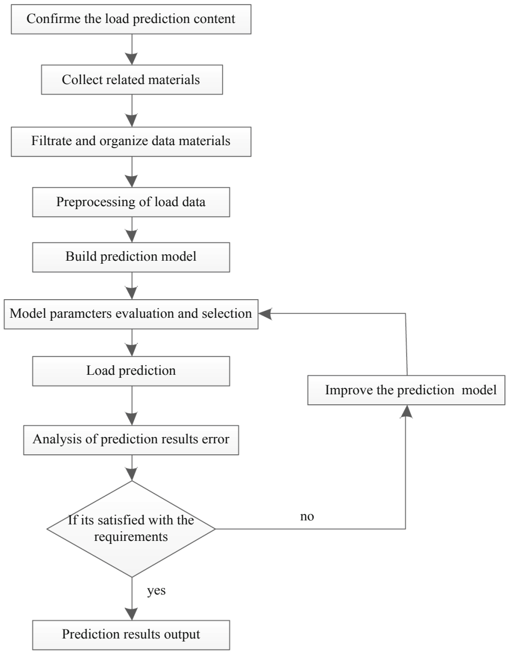

Fig. 1. Load forecasting flowchart

3 Model Design of Short-Term Power Load Big Data Parallel Prediction

3.1 Building a Map/Reduce Framework

Map/Reduce, first proposed by Goggle, is one of the most widely used frameworks for distributed processing of massive data (usually larger than 1TB) in recent years [7]. Map/Reduce is derived from the two core operations Map and Reduce of distributed processing. Map is responsible for distributing large tasks in parallel and distributed to multiple machines. Each machine only calculates the data stored in local memory. Reduce synthesizes the results of multiple machines in Map to get the final result. The entire running process takes the form of key-value pairs as input and output. The Map/Reduce workflow is shown in Fig. 2.

Data Locality (DL) is a feature of Map/Reduce. Map is responsible for breaking data, and Reduce is responsible for data aggregation [8]. The Reduce function is a summary of the results for intermediate operations. Although it does not have good parallel performance of the Map function, its calculation methods are mostly simple, so it is suitable for large-scale parallel operation. This simplified parallel computing programming model can shield the differences of the underlying physical structure, and only provide the available interfaces to the upper user. It stores data on compute nodes as much as possible, which can avoid the centralized upload of large amounts of raw data.

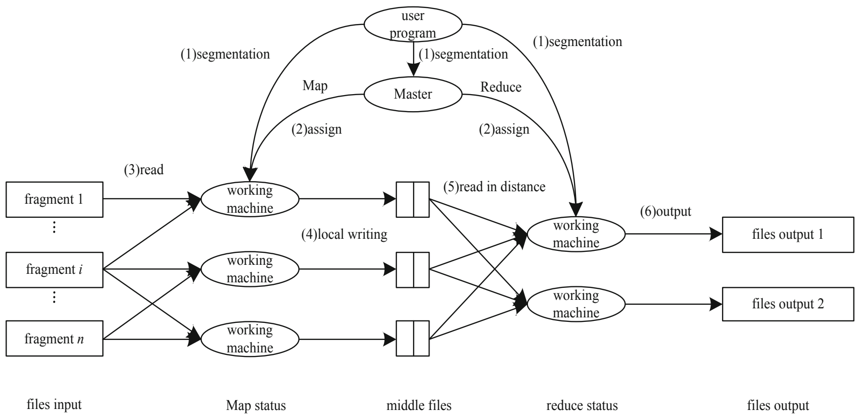

Fig. 2. The Map/Reduce workflow

Therefore, applying the Map/Reduce processing framework to big data processing in a smart grid not only reduces a large amount of communication overhead, but also can make full use of an idle device with a computing capability that is widely distributed in a power grid to exert a scale calculation capability.

3.2 Prediction Flow Design Based on Map/Reduce Framework

The grid load has statistically significant characteristics of periodicity and similarity. Based on the above-mentioned Map/Reduce functional program design to deal with the "divide and conquer" idea of big data, this paper proposes a short-term power load parallel prediction method based on big data, as shown in Fig. 3.

MapReduce is a programming model that can be used for data processing. It provides a universal, reliable and fault-tolerant distributed computing framework. MapReduce has some limitations on how to implement applications.

These restrictions are as follows:

All calculations are decomposed into map or reduce tasks

Tasks are defined mainly based on input data and output data.

Tasks depend on their input data and do not need to communicate with other tasks.

MapReduce uses the map and reduce functions to implement the application and execute these restrictions. These numbers are grouped into jobs and run as a whole: run mapper first and then reducer. Run as many tasks as possible. Because parallel tasks do not run on each other, they can be run in any order as long as the map tasks run before the reduce tasks.

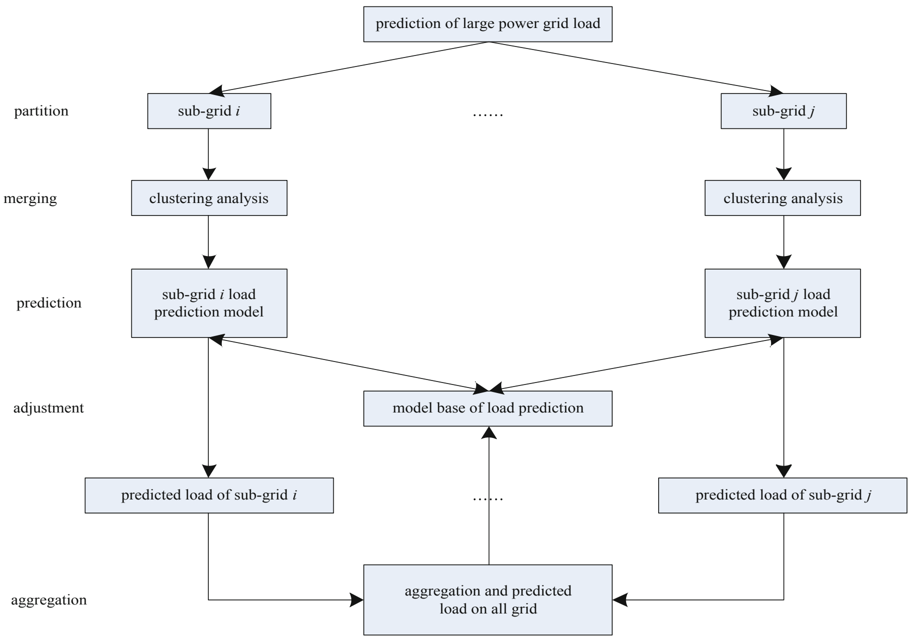

Fig. 3. The parallel prediction model of short-term power load big data

3.3 Realization of Short-Term Power Load Big Data Forecast

Based on the addition of meteorological monitoring devices at each substation of 110 kV and above, real-time meteorological data (temperature, humidity, rainfall, wind speed, etc.) and weather forecast data (predicted temperature, predicted humidity, predicted rainfall, forecasted wind speed) of each sub-network are collected. Data are quantified and interpolated, together with the load data to form the maximum data set available [9]. Based on the above studies, sub-networks and global load forecasting models were established.

3.3.1 Short-Term Power Load Forecast of the Entire Network

Without considering loss, the global load should be equal to the sum of all subnet loads, but in practice the global load is higher than the sum of all subnet loads. The degree of increase in each moment is different, and the relative stability is determined by factors such as the power consumption of the plant, network loss, and load. Set W_t as the proportional coefficient of the load of the global grid at the time of t and the sum of the n subnet loads at the corresponding time. k_t is the degree to which the global load is higher than the sum of the loads of the various subnets, and the fluctuation range is small. The relationship between the two is:

$$w_t = 1 + k_t \tag{1}$$

The load at each moment of the global power grid is the sum of all sub-network loads at the corresponding time multiplied by the proportional coefficient at the corresponding time. Set L_t to the global grid load at time t, where n is the number of subnets and $l_t(i)$ is the load of subnet i at time t. The formula is as follows:

$$L_t = w_t \times \sum_{i=1}^{n} l_t(i) \tag{2}$$

Through the above two formulas, the global grid load value at a certain moment can be obtained.

3.3.2 Subnet Short-Term Power Load Big Data Forecast

Numerous studies have shown that meteorological factors are the dominant factors affecting short-term load; and among many meteorological factors, temperature has the most significant and regular effects on the power load in each region. In addition, due to the interaction between the urban heat island effect, temperature and humidity effects, and cumulative effects, the power load characteristics in summer become more complicated, especially for a large power grid that covers a large geographic area.

It sets a certain threshold, selects the influencing factor vector with higher similarity and the load at the corresponding time as the input of the sub-network prediction model, and performs the prediction one by one. According to the different load characteristics of each sub-network and the different influencing factors, the appropriate load forecasting model is selected to perform point-by-point load forecasting for each sub-network. The three-layer BP neural network load forecasting model has a better nonlinear system approximation capability in dealing with the stationary random time series of complex variables, and can easily account for factors such as temperature, rainfall, and relative humidity that have important influence on the power load. The role of [10]. Therefore, this paper selects three layers of BP neural network as the subnet load forecasting model in the load forecasting model library.

The BP network adopts a three-layer structure. The first layer is the input layer: the input variables are the variables related to the load to be predicted, including the daily type, THI, rainfall, and the historical load of the corresponding time; The second layer is the hidden layer. According to experience, the number of neurons in the hidden layer is 7–16, and the output of the hidden layer adopts the logsig function. The third layer is the output layer. It sets the load forecast value L at a certain time of the day to be predicted, adopts a linear transformation function, sets wi to be the connection weight of the hidden layer and the output layer neuron, and yi is the output of the hidden layer neuron. n is the number of hidden layer neurons. As shown in the equation:

$$L = \sum_{i=1}^{n} w_i y_i \tag{3}$$

The adjustment of BP network connection weights and thresholds is the main component that affects the performance of BP network. In this paper, the "negative gradient descent" theory is adopted to adjust the connection weights and connection thresholds of input layer and hidden layer, hidden layer and output layer. The initial values of the connection weights and thresholds of the BP network are given randomly ($-1 < w < 1$), E is the sum of the error sum of the actual value and the network output; X is the learning rate; $w_{ij}(t)$ is the input layer. The connection weights of i-th neurons and j-th neurons of the hidden layer are calculated as follows:

$$w_{ij}(t) = -X\frac{E}{w_{ij}} + L \tag{4}$$

In the same way, we can calculate the connection weight between the jth neuron in the hidden layer and the kth neuron in the output layer, i.e. $w_{jk}(t)$ is calculated as:

$$w_{jk}(t) = -X\frac{E}{w_{jk}} + L \tag{5}$$

Y is the connection value between neurons. The meaning of the subscript is the same as the weight. Adjust the above two formulas to get:

$$Y_{ij}(t) = -X\frac{E}{Y_{ij}} + L \tag{6}$$

$$Y_{jk}(t) = -X\frac{E}{Y_{jk}} + L \tag{7}$$

Through the above several formulas, the load forecasting value at a certain moment can be obtained, and the connection weights and connection values between neurons can be obtained according to the difference between this value and a specific learning rate. According to the connection weight obtained, the system performance of the model can be directly judged. The three-layer BP neural network load forecasting model is shown in Fig. 4.

As a main model of artificial neural network, BP neural network is widely applied. It is a multi-layer forward network with one-way propagation. The input signal passes through the hidden layers from the input node to the output node. The output of each layer node only affects the input of the next layer node. It propagates back the error between the actual output and the expected output of the neural network. The whole training process is the process of the global error of the network tending to the minimum.

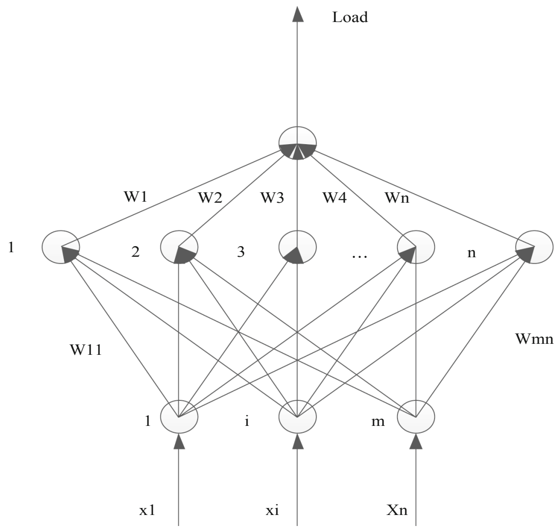

Fig. 4. Three-layer BP neural network load forecasting model

4 Test Analysis

In order to verify the validity and accuracy of the load forecasting method mentioned above, this paper uses the actual load data and meteorological data from a province in southern China as a basis for modeling and forecasting. A parallel short-term power load forecasting model based on big data was constructed for weekdays and body-building days, and compared with the load forecasting results of a centralized ARIMA model and other forecasting methods, and the forecasting effects and advantages of each method were analyzed. The study area is a province in southern China that covers a relatively large geographic area, it is with typical of a subtropical monsoon climate, where has high temperatures and rain in summer. July has the highest monthly average temperature in the year. During this period, due to the influence of high temperature, the accuracy of summer load forecasting is basically the lowest in the annual load forecasting work. Based on the 96-point real load monitoring data, real-time meteorological data and user data were collected and maintained throughout the province in summer, this paper applies the proposed method to carry out the short-term power load forecasting and error analysis in this region.

4.1 Selected Test Test Data

This article's simulation environment: HP Z220SFF sequence F4F06PA model workstation, Intel Xeon E3 quad-core processor, 3.2 GHz CPU frequency. In this paper, the area to be forecasted is divided into 9 sub-networks, and each sub-network is simulated by examples. Each subnet load data is shown in Table 1.

Table 1. Sub-network load conditions

Subnet number	Average load (MW)	Maximum load (MW)	Minimum load (MW)
1	3682.801	4747.841	2405.154
2	674.606	1051.854	157.493
3	876.213	1201.495	457.543
4	655.213	1132.079	286.105
5	677.669	941.172	322.466
6	3800.419	4803.953	2462.932
7	995.822	1279.813	544.847
8	2297.629	3100.285	1309.890
9	1362.372	1879.321	748.011

4.2 Prediction Test Results Analysis

For the daily load data, the load curves on the weekdays and holidays are significantly different. Compared with the weekdays, the load during the holidays was significantly reduced, and the shape of the load curve was also different from that of the normal day. In this section, the 96-point load of weekdays and body-wake days is selected to perform rolling forecasting and compared with the traditional load forecasting method to verify the validity and stability of the proposed load top-measurement method. The weekday forecast curve is shown in Fig. 5.

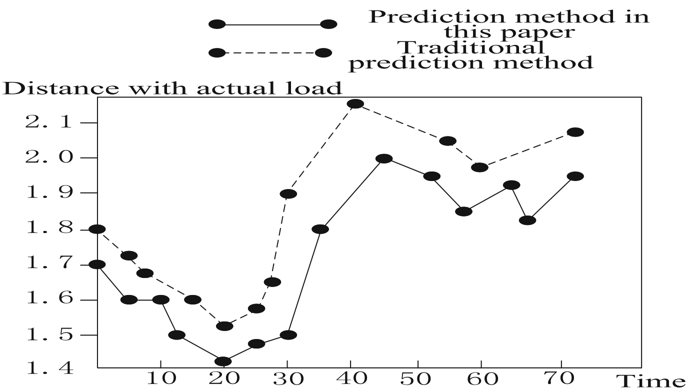

Fig. 5. The weekday forecast curve

As can be seen from the figure, for the weekday, the parallel forecasting model of short-term power load big data proposed in this paper has a better prediction result, and the deviation between the forecasting curve and the actual load is small, the average deviation is 1.7, and the overall prediction effect is good. However, the load forecasting value made by using other traditional forecasting methods has a large deviation from the actual value, especially in the load peaks and valleys, the load forecasting effect is not good, the average deviation is 1.9, which is 12% higher than this model.

The holiday forecast curve is shown in Fig. 6.

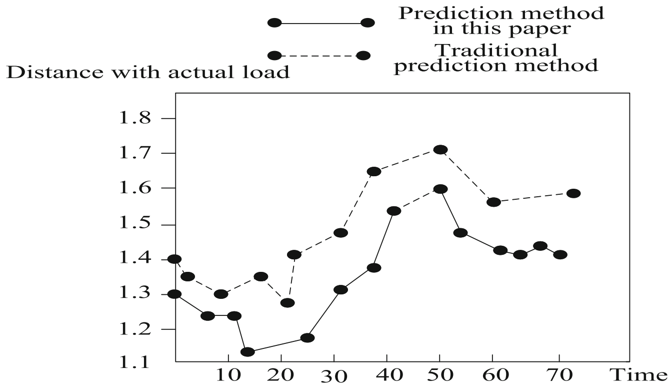

Fig. 6. Holiday forecast curve

As can be seen from the figure, the parallel forecasting model for the short-term power load big data proposed in this paper is also ideal. The deviation between the forecasting curve and the actual load is small, with an average deviation of 1.35, which has a good overall forecasting effect; However, the load forecasting value made by using other traditional forecasting methods has a large deviation from the actual value, especially in the load peak and valley periods, the load forecasting effect is not good, and the average deviation is 1.5, which is 11% higher compared to this model.

5 Conclusion

This paper proposes a parallel prediction model for short-term power load big data, analyzes the types and characteristics of different types of load forecasting, and proposes short-term power load forecasting steps. According to the forecasting steps and the Map/Reduce model theory, a short-term power load parallel prediction scheme is established to divide the sub-networks, and a short-term parallel load forecasting model based on sub-network and the whole network is established to achieve the research of this paper. The test data shows that the method designed in this paper has extremely high effectiveness. It is hoped that this paper can provide a theoretical basis for the study of short-term power load big data parallel prediction model.

Facing the large power load data in smart grid, in order to fully mine and analyze local and global information in the area to be predicted and avoid data anomalies and network congestion caused by uploading a large number of data sets, it is necessary to study and develop more effective short-term power load according to the characteristics of wide-area distribution of power grid sensing system. Big data prediction method.

References

1. Weekend, J.M.: Short-term load forecasting method combined with multi-algorithm multi-model and online second learning. Comput. Appl. **37**(11), 3317–3322 (2017)
2. Xiong, J.H., Niu, X., Zhang, C.G., et al.: Short-term load forecasting based on wavelet transform with Drosophila optimized support vector machines. Power Syst. Prot. Control **45**(13), 71–77 (2017)
3. Liu, D.D., Zhu, J.M., Huang, T.T.: Short-term power load forecasting based on time series and grey model. J. Qiqihar Univ. (Natural Science) **33**(3), 7–12 (2017)
4. Li, L., Yang, S.F., Qiu, J.P., et al.: Simulation research on power system short term load forecasting method. Comput. Simul. **34**(1), 104–108 (2017)
5. Wang, W.G., Dou, Z.H., Shen, J., et al.: Improved short-term load forecasting based on backstepping theory for fuzzy mean function. Hydroelectric Energy Sci. **14**(12), 208–211 (2017)
6. Zhi, L., Guoqiang, S., Zhinong, W., et al.: Short-term load forecasting based on variable selection and Gaussian process regression. Electr. Power Const. **38**(2), 122–128 (2017)
7. Song, R.J., Yu, T., Chen, Y.H., et al.: Similarity duplicate records detection algorithm for big data based on MapReduce model. J. Shanghai Jiaotong Univ. **52**(2), 214–221 (2018)
8. Xiao, W., Hu, J., Zhou, X.F.: Research review of parallel association rules mining algorithm based on MapReduce computing model. Res. Comput. Appl. **38**(1), 132–139 (2018)
9. Lei, J.S., Hao, X., Zhu, G.K.: Short-term power load forecasting based on "layered-pooling" model. Electric Power Constr. **38**(1), 68–75 (2017)
10. Su, X., Liu, T.Q., Cao, H.Q., et al.: Short-term load forecasting based on multi-distributed distributed BP neural network based on Hadoop architecture. Proc. CSEE **37**(17), 4966–4973 (2017)