

Medium and Long-term Power Load Forecasting based on the Thought of Big Data

Feng Xian Zheng

Power grid operation and training department
State Grid Sichuan Electric Power Company Skill
Training Center
ChengDu ,China
fengxianzheng@qq.com

Zhang Ting Ting

Power grid operation and training department
State Grid Sichuan Electric Power Company Skill
Training Center
ChengDu ,China
e-mail: 1062029444@qq.com

Li Hong Jun

(of Affiliation): Power grid operation and training
department
State Grid Sichuan Electric Power Company Skill
Training Center
ChengDu ,China
e-mail: 271237358@qq.com

per Zhao Bin

Power grid operation and training department
State Grid Sichuan Electric Power Company Skill
Training Center
ChengDu ,China
e-mail: 348941671@qq.com

Abstract—The traditional medium and long-term load forecasting methods are mainly carried out based upon model or algorithm, and forecasting results rely heavily on the accuracy of mathematical model, but model adaptability is very poor. Medium and long-term load forecasting lasts long and suffers from lots of uncertain influential factors in a broad spatial scope, so this paper proposes a big data technology-based medium and long-term load forecasting method. By analyzing the typical characteristics of the big data of load forecasting and the different levels of structure relations between the data, the paper sets up a big data system for load forecasting, a frame structure for load forecasting, and a big data-based medium and long-term refined load forecasting model, which falls into forecasting partition model and load forecasting model. The validity and practicability of this method is verified based on an analysis of the actual grid load in a certain region.

Keywords—load forecasting; big data; correlation; model; refine

I. INTRODUCTION

As the primary task in power grid planning and development, medium and long-term power system load forecasting has provided essential basic data for power grid design and planning, to ensure the demand for electric power in all sectors of the national economy and the people's life. But due to the large time span and wide geographical distribution of medium and long-term load forecasting, as well as the impacts from various uncertain and random factors, it is very complex to make predictions of load accurately. The development of informatization and interaction under an intelligent grid environment has helped obtain lots of relevant data, including load type, distribution and user demand for electric power, making it possible to forecast load by classification in a refined way.

Load forecasting methods mainly include three ones[1,2]: classic forecasting method, traditional forecasting method and modern forecasting method.

Electricity consumption per unit output method, elastic coefficient method and classed load forecasting method are classic forecasting methods, which rely on experience or a simple variable relation, and cannot produce a high-accuracy result; trend analysis approach, regression analysis method and time series method are traditional forecasting methods, which need a complex model and a tedious computational process, have high requirements for historical data, and can produce a relatively accurate forecasting result; artificial neural network method, grey theory forecasting method, fuzzy theory forecasting method and data mining technology are modern forecasting methods. When the methods of this kind are adopted, model structure and parameters needn't be known in advance, neither is a complex mathematical model required, so they are suitable for the forecasting of multivariable, nonlinear, time-varying and probabilistic power loads. These three kinds of load forecasting methods are based on some mathematical theory with model or algorithm at the core. When one of them is adopted, a load forecasting model needs to be built and improved according to the actual situation, and finally a forecasting result can be calculated. But the forecasting result relies heavily on mathematical model, and the model is poorly adaptable to diverse basic data types.

Big data contains huge social, economic and scientific research values, so effective organization and utilization of big data will greatly promote the development of the social economy and scientific research. As a powerful tool for medium and long-term load forecasting, big data technology provides a new idea for load forecasting. It is a method by which an adaptive model is built based on data with data at the core. There is already lots of data information in the operation process of a real power grid, such as load distribution and type data. In addition, the research and development of a friendly user interaction system can make it easier to collect more and more load information data; additionally, such relevant factor data as

climate, economy and population will also get increasingly comprehensive and precise along with the improvement of observational means and the renovation of tools[3].

Based upon this trend that data becomes more and more detailed and accurate, this paper proposes a big data-based load forecasting methodology system. Then, the paper makes comprehensive use of large-scale and multi-type historical data, and mines more related data to form a multi-level spatio-temporal data system and set up a forecasting model system compatible with big data, to forecast the loads within the forecasting region meticulously. Finally, the paper tests the rationality of the big data-based load forecasting method by forecasting the loads in a certain region.

II. LOAD FORECASTING FRAMEWORK AND DATA ANALYSIS

If big data-based load forecasting method is adopted, various influential factors of data should be taken into adequate consideration. The data required for load forecasting includes not only basic historical load data, but also economic and political data, as well as the data of smart city and intelligent electric equipment, the data of natural environment, that is, meteorological data, and the data of various natural calamities and emergencies, etc. Big data-based load forecasting method is to comprehensively analyze the effects of various factors by comprehensively using the data.

The first step of this method is to collect, collate and refresh data, gather various relevant influential factor data from the measuring equipment terminals in power grids and the information issuing platforms of the government departments at all levels, update the data at any time, and eliminate the data without reference value through data timeliness testing. The second step is to analyze and manage the data, that is, to classify the data and set up a data hierarchical structure system in accordance with the correlations between various data. The final step is to build a forecasting model in conformity with the data of multi-dimensional spatio-temporal system.

A. Load Data Partitioning and Structuring

Loads are distributed in a broad spatial area. Load data isn't a simple sum of various local data, but should be at certain structural levels. The medium and long-term load forecasting method based on the thought of big data is exactly to handle different characteristics of loads respectively and forecast them by levels according to the structural characteristics of load partition and the personality of each load. After the forecasting region is divided accurately and rationally into different levels of forecasting partitions based on the actual distribution characteristics of loads in power grids, the change of regional loads can be reflected clearly and timely.

B. Load Data Correlation

In the load data partitioning and structuring system, there may be a correlation and varying levels of cooperation between the loads distributed in different regions, so consideration must be given to this correlation in forecasting process to improve forecasting precision. There are mainly two correlations between the loads, one of which is transverse correlation, that is, the interactional

relationship between the industries at the same level distributed in different intervals of forecasting. For instance, such equipment manufacturing industries as power equipment, electronic equipment and automobile equipment have the same demand for steel materials, so all these loads will be affected by material supply. The other is longitudinal correlation, that is, the interactional relationship between the interrelated upstream and downstream industries. For instance, the growth of non-metallic ore manufacturing industry and cement industry will promote the development of construction industry, such as the construction of railway, and the growth of railway load.

C. Eigenvalues of Load Data

The data volume of load forecasting is large, but the value density is low, so there isn't much effective data that has an impact on forecasting results. Eigenvalue is a piece of typical data of load, and an analysis of big load data can help obtain a reference value for various loads and an influence coefficient for each effect or component load.

D. Timeliness of Load Data

The timeliness of data is mainly reflected in the adjustment of industrial structure, the change of policy and domestic and foreign political environment, the probabilistic natural climates, and specific events, etc. The analysis of the timeliness of data is in essence the analysis of whether the policy environment and economic environment at that time are not consistent with the current one. If there is a big difference, correction and integration should be carried out by the use of parameters, or some data should be given up selectively or all data should be directly abandoned according to the specific situation.

E. Small Probability Events in Load Forecasting

Sometimes, load forecasting will come under the influence of some small probability events, particularly in some specific regions. Small probability events have a low occurrence probability, but once one occurs, it will produce a significant impact on the corresponding locality.

III. LOAD FORECASTING MODEL

A. Partition Model and Correlation Treatment

Partitioning should be implemented by steps according to the actual distribution characteristics of load. The series of partitioning should be different in different regions with different development degrees and load densities. We can

denote each partition with I_n^m , where m denotes the series of partitioning, and n denotes the different partitions at each level. If sub-partition I_n^m has S sub-partitions, then[5-7]

$$I_n^m = \sum_{k=1}^S I_{n.k}^{m+1} \quad (1)$$

So, for a forecasting region, there is

$$I = \sum_{i=1}^n I_i^1 = \sum_{j=1}^p I_{1,j}^2 + \sum_{j=1}^q I_{2,j}^2 + \sum_{j=1}^r I_{3,j}^2 + \dots + I_k^1 + \dots = \dots \quad (2)$$

In the above equation, I_1^1 , I_2^1 , I_3^1 denote the partitions that have p, q, r sub-partitions respectively; I_k^1 refers to the partitions that have no sub-partition. This shows that there can be different levels of lower partitions in different sub-partitions, and the partitions are determined according to the practical situation in each partition until the determination of one at the highest level is finished.

$$I = \sum_{i=1}^n I_i^1 = \sum_{j=1}^p I_{1,j}^2 + \sum_{j=1}^q I_{2,j}^2 + \sum_{j=1}^r I_{3,j}^2 + \dots + I_k^1 + \dots + I_{v_1}^1 + I_{v_2}^1 + \dots \quad (3)$$

After partitioning, the loads in each partition are divided, and the loads of different influencing factors are partitioned in accordance with the different features of each partition. When partitioning goes deep into the lowest level, the loads can be divided into such specific types of loads as office building load, residential building load and

The loads in each partition that have transverse and longitudinal correlations with each other can be extracted respectively to set up independent virtual partitions. The independent virtual partitions can be adjusted dynamically in accordance with the change of actual electrical load, so that forecasting result should be consistent with the reality. After virtual partitions are set up for loads, the forecasting region should be

external facility load. When there are only a few levels of partitioning, the loads can also be divided into such types as the load of the three main industries and the load of resident power consumption. Load classification should be carried out within partitions. For partitioning, if there are types of loads[8-10], then,

$$P_{I_n^m} = \sum_{i=1}^L P_{i, I_n^m} \quad (4)$$

B. Load Forecasting Model and Parameter Adjustment

The model falls into two kinds: electric capacity forecasting model and load forecasting model, corresponding to partition classification, and the forecasting model also has a corresponding a hierarchical structure. Now the I^r -level forecasting that contains two kinds of sub-partitions is partitioned into I_k^r . That is, there are p-l+1 partitions that contain sub-partitions and l-1 partitions that do not contain sub-partitions. The detailed modeling process is shown below:

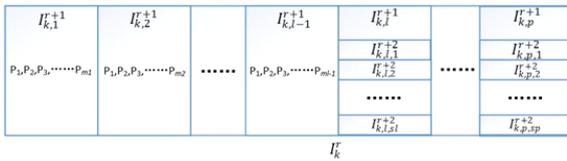


Figure 1. Model Drawing of the Forecasting Intervals at Different Levels

For a certain type of load P_i in the $l-1$ partitions that do not contain sub-partitions, suppose it is affected by n influencing factor, write it as a vector, that is, $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,n})$. About the prediction for P_i , first determine a reference value $P_{i,t,base}$ for it according to its overall variation trend, then

$$P_{i,t,base} = P_{i,0} \left(1 + k_i \cdot r_{\Delta G_i} \right)^t \quad (5)$$

$P_{i,0}$ denotes basic reference value, and the load capacity of a certain year or the mean value of many years can be chosen as a reference; k_i denotes a elasticity coefficient, $r_{\Delta G_i}$ denotes economic growth rate, and t denotes the duration relative basic reference value.

Then, correct the basic forecasting value according to the degree of each influencing factor's effect on load, then

$$P_{i,t} = P_{i,t,base} + \sum_{i=1}^n (\alpha_{i,t} \cdot \Delta X_{i,t}) \quad (6)$$

$P_{i,t}$ denotes the forecasting value of a corresponding forecasting partition t years later, $\Delta X_{i,t}$ denotes the forecasting data change vector of each influencing factor t years later, and $\alpha_{i,t}$ denotes the coefficient of the influence of a corresponding change on $P_{i,t}$.

So for forecasting partition $I_{k,1}^{r+1}$, the forecasting value $P_{I_{k,1}^{r+1},t}$ of load in t years will be

$$P_{I_{k,1}^{r+1},t} = P_{I_{k,1}^{r+1},t,base} + \sum_{j=1}^{m_1} (\alpha_{j,t} \cdot \Delta P_{i,t})$$

$$\begin{aligned}
&= P_{I_{k,l}^{r+1}.t.base} + \sum_{j=1}^{m_j} [\alpha_{j,t} \cdot (P_{i,t} - P_{i,t.base})] \\
&= P_{I_{k,l}^{r+1}.t.base} + \sum_{j=1}^{m_j} \left[\alpha_{j,t} \cdot \left(\sum_{i=1}^n (\alpha_{i,t} \cdot \Delta X_{i,t}) \right) \right] \quad (7)
\end{aligned}$$

For the $P-l+1$ partitions that contain sub-partitions, the load at the lowest level should be calculated prior to the ones at higher levels, and for forecasting partition $I_{k,l}^{r+1}$, there is

$$\begin{aligned}
P_{I_k^r.t} &= P_{I_k^r.t.base} + \sum_{j=1}^p (\alpha_{j,t} \cdot \Delta P_{I_{k,j}^{r+1}.t}) \\
&= P_{I_k^r.t.base} + \sum_{j=1}^{l-1} \left\{ \alpha_{j,t} \cdot \sum_{q=1}^{m_j} \left[\alpha_{q,t} \cdot \left(\sum_{i=1}^n (\alpha_{i,t} \cdot \Delta X_{i,t}) \right) \right] \right\} + \sum_{j=l}^p \left\{ \alpha_{j,t} \cdot \sum_{i=1}^s [\alpha_{i,t} \cdot (P_{I_{k,i}^{r+1}.t} - P_{I_{k,i}^{r+1}.t.base})] \right\} \quad (9)
\end{aligned}$$

In the above model, the degree of various data's impact on the final forecasting result is reflected in impact coefficient $\alpha_{i,t}$, and to change impact coefficient $\alpha_{i,t}$ can change the forecasting result; $\alpha_{i,t}$ is decided by the practical situation, and certain empirical values need to be used as reference. Make estimation in accordance with some historical data and limit errors in a certain scope.

An Application Analysis of Medium and Long-term Load Forecasting

An Analysis of the Load Data from a Certain Region

The loads in different regions have something to do with multiple factors, and different types of loads have different change trends, so it is a precondition to the determination of important model parameters to analyze the growth trend of different types of loads in different regions and their share in their respective region. The line graph below illustrates the change trend of the total load in a certain region.

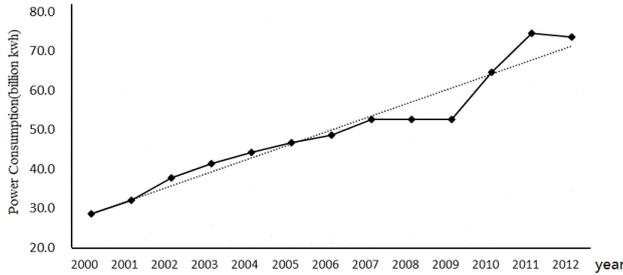


Figure 2. Change Chart of the Power Consumption in a Certain Region

$$\begin{aligned}
W_{I_j^2.t} &= W_{I_j^2.t.base} + \sum_{i=1}^5 (\alpha_{i,t} \cdot \Delta W_{i,t}) \\
j &= 1, 2, 3, 4, 5, 6, 7 \quad (11)
\end{aligned}$$

$$W_{I^1.t} = W_{I^1.t.base} + \sum_{j=1}^7 (\alpha_{j,t} \cdot \Delta W_{j,t}) = W_{I^1.t.base} + \sum_{j=1}^7 \left[\alpha_{j,t} \cdot \left(\sum_{i=1}^5 \alpha_{i,t} \cdot \Delta W_{i,t} \right) \right] \quad (12)$$

where, $\Delta W_{I_k^2.t} = W_{I_k^2.t} - \overline{W_{I_k^2.t}}$ and $W_{I_k^2.t}$ are the forecasting value of the various types of power utilization based on relevant economy and per-capita income. For $\alpha_{i,t}$ and $\alpha_{j,t}$, the ratio of the total electricity (or maximum load) in the corresponding forecasting interval to the electricity in the higher-level forecasting interval is adopted, and the forecasting period is from 2014 to 2020.

$$\begin{aligned}
P_{I_{k,l}^{r+1}.t} &= P_{I_{k,l}^{r+1}.t.base} + \sum_{j=1}^{S_j} (\alpha_{j,t} \cdot \Delta P_{I_{k,l}^{r+1}.t}) \\
&= P_{I_{k,l}^{r+1}.t.base} + \sum_{j=1}^{S_j} (\alpha_{j,t} \cdot (P_{I_{k,l}^{r+2}.t} - P_{I_{k,l}^{r+2}.t.base})) \quad (8)
\end{aligned}$$

Integration of these two kinds of sub-partitions finally helps get the forecasting result of forecasting partition I_k^r :

C. Model Analysis

According to the load data of a certain region, partition model can be divided into two levels of forecasting spaces, called underlying forecasting partition I_k^2 and upper forecasting partition I^1 , which denote the load capacity of the region and the load capacity of counties and urban districts respectively, then,

$$I^1 = \sum_{k=1}^5 I_k^2 \quad (10)$$

It is divided into 7 underlying forecasting spaces I_k^2 based on the provided data, and these 7 sub-partitions are divided into five types of power supply loads, that is, $m_1 = m_2 = \dots = m_7 = 5$, which represent power utilization in the primary industry, power utilization in the secondary industry, power utilization in the tertiary industry, urban and rural resident living power utility and wholesale electricity respectively. The expression for each forecasting partition:

D. Forecasting Result

The reference value of the power utilization and maximum load in each forecasting partition, as well as the correction coefficient of the power utilization in different types of partitions, is calculated according to the data analysis in the second part. The 7 underlying forecasting intervals are predicted respectively in accordance with the

model structure in the third part. The result of forecasting of the total load in a certain region is shown in Tab.1.

E. Forecasting Result

The basic data used for the model calculation is the load data of a certain region in 2013 and the years before 2013. From a perspective of forecasting result, the

forecasting result in 2014 was 8.4631 billion kilowatt-hours, while the actual value of the total electricity consumption in 2014 was 8.5317 billion kilowatt-hours. The predicted total load and different types of loads are relatively accurate, which suggests that the method based on big data is practicable.

TABLE I. THE FORECASTING RESULT OF TOTAL ELECTRICITY THROUGHOUT THE CITY IN A CERTAIN REGION

	2014	2015	2016	2017	2018	2019	2020
Total Electricity Consumption	84.63105	89.02612	93.99264	99.24549	104.8004	110.6742	116.8845
Primary Industry	0.447944	0.40798	0.371582	0.33843	0.308237	0.280737	0.25569
Secondary Industry	43.97611	46.39505	48.94705	51.63942	54.47989	57.47661	60.63816
Tertiary Industry	13.60228	14.24107	15.13236	16.07943	17.08577	18.1551	19.29135
Urban and Rural Resident Living Power Utility	13.57559	14.33225	15.13109	15.97445	16.86481	17.80481	18.79719
Wholesale Electricity	13.02913	13.64976	14.41056	15.21376	16.06173	16.95696	17.90209
Maximum Load	1480.163	1562.663	1649.761	1741.713	1838.791	1941.28	2049.48

IV. CONCLUSION

This paper first proposed a medium and long-term load forecasting method based on the thought of big data by analyzing and summarizing the new problems faced by medium and long-term load forecasting and the shortcomings of the existing medium and long-term forecasting methods, set up a frame structural system for the big data-based load forecasting method by combining the features of the big data used for load forecasting, and then deeply discussed the characteristics of load data, such as partition structuring and correlation, as well as its features in such aspects as critical eigenvalue, limitation from data timeliness and interference from small probability events. On this basis, the paper built a refined forecasting model, and applied this big data-based forecasting method to load forecasting in a certain region; then, it made a detailed analysis on the load data of the region and obtained a forecasting result, verifying the validity and practicability of this forecasting method.

REFERENCE

- [1] [1].Cheng Zhenbao.Power load forecasting method .China High-Tech Enterprises. ,2013,(32):114-115.
- [2] [2].Li yongxia,Su Ruiwen. The influencing factors of power system load forecasting GUANGDONG ELECTRIC POWER,1995,(4):44-46.
- [3] [3]. TU Xin-li,LIU Bo,LIN Wei-wei. Survey of big data. Application Research of Computers, 2014,31(6):142-146.
- [4] [4].Li Guibing,Luo Hong. The Research on intelligent data analysis Under the large data. Science & Technology Information. 2013,(30):11-12.
- [5] [5]. KONG De-zhi,LIU Qun-xing,WANG Ying-kai,ZHANG Ying-ying. The Research on Technology and Application of Big Data. Electronic Product Reliability and Environmental Testing,2013,11(31).
- [6] [6].Zhao Yunshan,Liu Huanhuan. Research on Application of Big Data Technique in Electricity Power Industry. Telecommunications Science,2013,8:57-62.
- [7] [7] WANG Dewen,SUN Zhiwei . Big Data Analysis and Parallel Load Forecasting of Electric Power User Side.Proceedings of the CSEE. 2015, (3)
- [8] [8] YE Feng HE Hua GU Quan ZHANG Gaofeng. Bad Data Identification and Correction for Load Forecasting in Energy Management System. AUTOMATION OF ELECTRIC POWER ., 30(15)
- [9] [9] ZHANG Suxiang , ZHAO Bingzhen, WANG Fengyu, ZHANG Dong. Short-term Power Load Forecasting Based on Big Data Proceedings of the CSEE. 2015, (1)
- [10] [10] Wen Jinbin, Wang XinLi, Lixue, Zheng Yihui, Zhou Lidian ,Shao Fengpeng. Short-Term Wind Power Load Forecasting Based on Frequency Domain Decomposition.Transactions of China Electrotechnical Society. 2013, 28(5)