Research on Electricity Characteristic Recognition Method of Clean Heating Based on Big Data Model

Xin-lei Wang[1], Jia-song Luo[1], Tong Xu[1], and Guo-bin Zeng[2(✉)]

[1] State Grid Economic and Technological Research Institute Co. LTD., Beijing 102200, China
wangxinlei5525@163.com
[2] Haikou University of Economics, Haikou 571127, China
zengguobin9965@163.com

Abstract. Because the traditional coal-fired heating mode consumes a lot of energy and is harmful to the environment, it produces a clean heating mode using electric energy, which is realized by energy storage heating equipment. The operation of energy storage heating equipment needs to be planned according to the electricity characteristics of clean heating. Therefore, a method based on large data model is proposed to integrate the electricity characteristics of clean heating using Hadoop platform. Then, according to the integrated data, the electricity load characteristics, electricity consumption characteristics, electricity consumption cycle characteristics and regional characteristics are identified to complete the electricity characteristics of clean heating. Farewell. Through experimental demonstration, it is proved that this method can effectively identify the electrical characteristics of clean heating and accurately predict the future heating data.

Keywords: Clean heating · Electricity forecasting · Seasonal cycle · Electricity characteristics

1 Introduction

Energy is the lifeblood of the national economy in modern society. Any resident and industrial activities can not do without the support of energy supply. Nowadays, China's energy production and consumption are increasing rapidly with the development of economy and society. The problems of backward energy structure, unreasonable energy utilization mode, over-reliance on fossil fuels and low energy utilization rate caused by rapid growth are becoming increasingly prominent [1]. Traditional coal-fired heating mode has a huge demand for coal mining, while coal-fired produces waste gas, which greatly aggravates environmental pollution. Therefore, clean heating has been respected. Electric energy storage heating equipment uses grain power or wind power to heat solid regenerator. The temperature reaches 750 °C. Heat is stored through insulation materials. When the fan is heated, the heat in the regenerator is blown out, and the water is transferred through the heat exchanger, and then through

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 348, pp. 25–35, 2021.
https://doi.org/10.1007/978-3-030-67874-6_3

the heating pipeline to the resident's home [2]. This method has no emissions, no pollution, and achieves clean heating.

Under the background of environmental protection concept and realistic energy problems, this clean heating mode is worth being fully implemented. The operation method of energy storage heating equipment needs to be designed in detail according to the heating and electricity characteristics. Since the concept of big data was put forward, domestic research has been in full swing. In theory, the concept, technology and application of big data have gradually formed a scientific, systematic and industrialized research system. In terms of theoretical system, the analysis of large data is mainly divided into five aspects: visual analysis, data mining algorithm, predictive analysis ability, semantic engine, data quality and data management [3]. Therefore, a large data model platform was born. This paper uses Hadoop platform based on big data model to integrate the characteristic data of clean heating data, and then analyses the characteristics of clean heating data, so as to provide a reasonable scheme for the allocation of energy storage heating equipment, and ensure that the clean heating mode achieves the optimal technology-economy ratio, the best sustainable development, the greatest economic benefit, the most reliable power supply and the best environmental protection [4].

2 Electricity Characteristic Recognition of Clean Heating Based on Large Data Model

2.1 Feature Recognition Data Integration

Before identifying the characteristics of clean heating electricity data, it is necessary to collect the electricity data, which often comes from different data sources, including databases, data warehouses and documents. At this time, it is necessary to integrate the data, that is, data integration. Because an attribute of an object may have different names in different databases, data integration may lead to inconsistencies and a large number of redundant data. So data integration can improve data quality better, reduce the occurrence of redundancy of result data sets and inconsistency of entity names, and improve the accuracy of data in the mining process [5]. At present, data integration still has the following problems.

Entity recognition. The same attribute in the clean heating data may have different attribute names in different databases, which may involve entity recognition in data integration. Generally, metadata can be used to avoid entity recognition problems that may arise in data integration.

Redundant data. Redundancy is a common problem in data integration. The main situation of data redundancy includes repeated occurrences of the same attribute value and duplication caused by inconsistent naming of the same attribute. Relevance analysis can be used to detect whether the attribute is redundant.

Numeric conflicts. The actual power consumption data may be stored in different measurement units because of different storage scenarios. For example, the time in a power data set may be 12 h in one system and 24 h in another system. If these two sets of data are integrated together, there will be a numerical conflict. For such numerical conflicts, normalization and other operations can be used to remove the impact of unit conflicts.

In view of the problems that may arise in the above data, Hadoop platform is selected for data processing. Hadoop is a distributed system infrastructure developed by Apache Foundation. It has high throughput and fault tolerance, and solves the problem of huge amounts of clean heating and power data. The Hadoop software framework is implemented in Java language, which supports cluster operations of thousands of data computing nodes and data processing at PB level [6]. Hadoop is composed of several independent systems, which are interdependent and constitute the whole of Hadoop. The operation of Hadoop platform system is shown in Fig. 1.

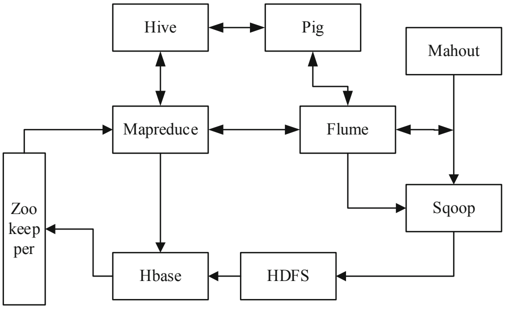

Fig. 1. Hadoop platform system operation diagram

MapReduce is a distributed computing framework with many data analysis components. Hive is a data warehouse that can provide a query language similar to SQL. It converts SQL to MapReduce and executes on Hadoop. It can be used for offline analysis. Pig is a data stream processing framework. Mahout is a data mining algorithm library. It mainly implements some classical algorithms in machine learning field, which can help designers build effective models better. HBase is a column-oriented database, which uses Google's BigTable data model and enhanced coefficient ranking mapping table. In order to facilitate the unified management of clusters, a distributed collaboration service Zookeeper is used. Sqoop is a data synchronization tool. It is mainly used to transfer data between traditional database and Hadoop. Its essence is MapReduce program. Flume is an open source log collection system [7, 8].

The data analysis platform in this paper uses four Aliyun servers to build Hadoop cluster. The Hadoop version is Apache version of Hadoop 2.6.4. The server system is Centos 6.5, 64-bit operating system, single-core CPU. The memory of each server is 1G, and the size of hard disk is 40G. The high availability architecture of Hadoop is used to ensure the stability of the data processing cluster. When the primary node has problems, the standby nodes can be activated to ensure the normal operation of the cluster.

First, the host name is changed to the host name that receives heating and power data, and then Java files are installed on each server to define the configuration environment variables. Then configure password-free login, install SSH for each server, so that password-free communication between each host, and synchronize the time to ensure that the cluster time is consistent. Download hadoop.2.6.4. tar.gz, copy the compressed package to the same location on each server and decompress it. Then we begin to configure Hadoop. The data files in the etc./hadoop directory of the Hadoop folder are XML edited, and the editing process is shown in Fig. 2.

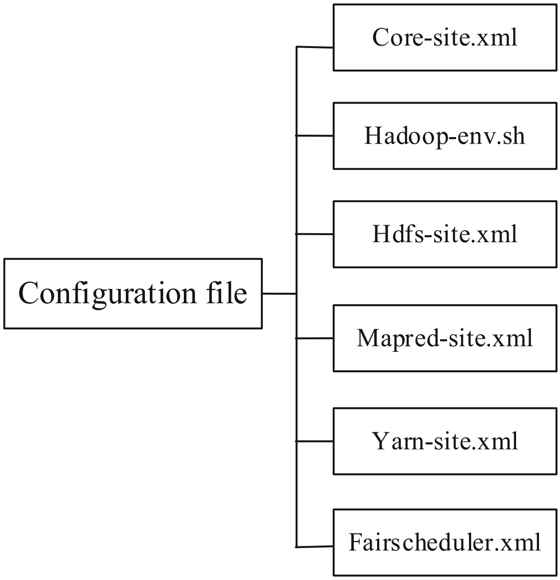

Fig. 2. Electricity data XML chart

Because of the huge amount of data of clean heating power, the time of collecting and processing data by other methods is long and the workload is too large, so the Hadoop method can also save the time needed for data loading and improve the performance of data warehouse construction. In the process of data integration, based on the detailed understanding of power consumption data, index is used to improve the performance of analysis and query, and index the column attributes often involved in query in dimension tables and fact tables, which further improves the performance of data analysis. Considering the data warehouse based on Hive, data analysis is carried out in the form of MapReduce, so the number of Map and Reduce can be set to ensure high performance in the process of data integration.

After integrating the data of clean heating power consumption, the identification of the characteristics of clean heating power consumption is gradually completed by acquiring the corresponding identification data. The identification of electric characteristics of clean heating includes: identification of electric load, identification of electric consumption, periodic identification of electric consumption, identification of regional characteristics and so on. Through the interpretation of data, the identification of electric characteristics of clean heating is carried out.

2.2 Identifying Electric Load

In the process of clean heating and power consumption, there is a certain rule that the power consumption of heating and power changes with time, which can be described by the change of power load with time [9, 10]. According to the different time periods, the load curve can be divided into daily load curve, weekly load curve, monthly load curve and annual load curve. The identification of power load characteristics is an important index for power supply side planning and management. By acquiring load data and identifying load characteristics, the index data of load characteristics of demand are shown in Table 1.

Table 1. Identification indicators of power load characteristics

Description class	Comparison class	Curve class
Daily maximum (minimum) load	Daily load rate	Daily load curve
Daily average load	Daily minimum load rate	Weekly load curve
Peak valley difference	Daily peak-valley difference rate	Annual load curve
Monthly maximum (minimum) load	Monthly average daily load rate	Annual continuous load curve

Among them, the calculation methods of various characteristic indexes of electric load are as follows:

Load rate (ratio of average load to maximum load) = heating power average load (KW)/heating power maximum load (KW)] * 100%; average daily load rate (%) = heating power period daily load rate/heating power period calendar day; peak-valley difference (%) = heating power period peak-valley difference maximum value (KW)/maximum load on the day (KW) * 100%; monthly load balance rate (%) = reporting monthly average daily power consumption (KWH)/Report Maximum daily power output (KWH) * 100%; annual load balance rate (%) = (maximum monthly load/maximum monthly load of 12 * maximum) × 100%; maximum load utilization hour (H) = maximum power output (supply and use) during reporting period (KWH)/maximum power output (supply and use) during reporting period (KW); simultaneous rate (%) = [maximum power system load (KW)/absolute maximum load of each component unit of power system (KW) * 100%.

Through calculation, the load characteristic curve of clean heating electricity is drawn. Usually, the load characteristic can be divided into high load rate type, continuous type, peak-up type, peak-avoidance type, temperature sensitive type and so on. High load rate type means that the load characteristic curve of heating power is stable, the difference between peak and valley is small, and the load rate is over 90%. The load curve of continuous type is between 80% and 90%. The load curve is relatively stable. Because of the limited adjustment ability of the existing heat storage heating equipment, the load curve will be continuous, but there will be a certain peak-valley difference. The peak time point of heating load in a day; the peak avoidance type shows the trough time of heating power.

2.3 Identifying Electricity Consumption

Accurate identification of electricity consumption is one of the keys to ensure the operation of clean heating mode. Therefore, multi-dimensional features of heating data are identified based on large data model.

Based on the massive data of clean heating power consumption, a multi-dimensional evaluation index system of electricity consumption characteristics is established. The time and space of clean heating power consumption are clustered and analyzed, and the characteristics of clean heating power consumption are mined and identified. Develop the process of power consumption identification, and use the evaluation method to avoid the adverse impact of other factors on the accuracy of the identification algorithm. The data identification process of clean heating power consumption based on large data model is shown in Fig. 3.

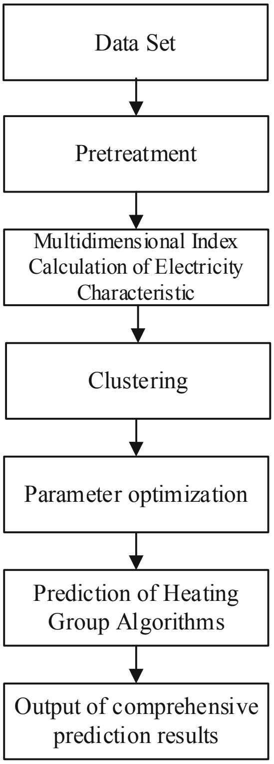

Fig. 3. Identification flow

Due to the difference of electricity consumption characteristics among different heating users, the heating objects are classified according to different electricity consumption modes, and the changing rules of electricity consumption of different users are studied. The key to realize the identification is to use large data analysis. The identification methods and results of electricity consumption characteristics are closely related to the selection of electricity consumption characteristics indicators. Therefore, it is necessary to define reasonable evaluation indicators of electricity consumption characteristics to assist the acquisition and identification of electricity consumption data.

The steps of power consumption identification are as follows: firstly, effective data preprocessing is carried out for m large data sets of heating power consumption, and target analysis data set Vd is generated, in which the first user's power consumption sequence is Vi (xt1, xt2, … , xtn), I = 1, 2, … , m; by calculating the multi-dimensional evaluation index of power consumption of mass users, the data set VD of user's multi-dimensional characteristics is established, in which the first user's characterization sequence is Vi (xt1, xt2, … , xtn), I = 1, 2, … , M Vi (xt1, xt2, … , xtn, 1, 2, W), I = 1, 2, … , m; Then clustering the data sets of heating multi-dimensional electricity characteristics, mining and identifying electricity consumption patterns, and getting the K user groups set Gk of M users. Finally, according to the characteristics of user groups, the parameters are trained separately. On the basis of parameter selection, an algorithm recognition model is established to obtain the identification values of electricity consumption of each heating group and all heating groups, and to identify and evaluate them. The formulas for evaluating the recognition results are as follows:

$$\beta_{\mathrm{i}}=\sum_{j=1}^{w} a_{ij}x_j = a_i^T x \tag{1}$$

The parameters in the above formulas are derived from the data of power consumption groups. When the evaluation result is beta > 8, the evaluation result is effective. This method is suitable for the analysis and processing of large data platforms, and has high recognition accuracy and stability. The sources of big data mainly include internet, physical information system, scientific experiments and so on. Usually, physical information system data and scientific data are obtained through a sensor network composed of sensors or observation devices. The rapid popularity of the Internet and mobile Internet provides the most convenient and lasting carrier for data recording. Because heating is a periodic process, the identification of clean heating power consumption can get the overall data of annual electricity consumption. In order to better understand the time rule of heating power consumption, it is necessary to identify the periodicity of clean heating power consumption.

2.4 Identifying Periodicity of Electricity Consumption

The fluctuation of electricity consumption for clean heating has obvious periodic regularity with time, which is called seasonal effect [11, 12]. According to a certain mathematical method, time series with seasonal effects are decomposed into components that show periodic changes with time series and components that are basically independent of time changes, so that seasonal electricity consumption law can be more clearly presented. The seasonal adjustment algorithm is used to obtain monthly or quarterly time series data, which contains four components, namely, long-term trend component, fluctuation cycle component, seasonal component and irregular component [13, 14]. The long-term trend component represents the long-term trend characteristics of time series. The fluctuation cyclic component is a cyclic change in units of several years. They describe the basic change law of time series. The variation of time series is an average value, and the formula is as follows:

$$\overline{Y}_{T+1} = \frac{1}{N}(Y_T + \ldots + Y_{T-N+1}) \tag{2}$$

Among them, T is the number of seasonal cycles, N is the statistic, t is the initial quantity of seasonal cycles, and its standard error S formula is:

$$S = \sqrt{\frac{\sum_t^T N + 1(\overline{Y} - Y_t)^2}{T - N}} \tag{3}$$

According to the above calculation method, seasonal periodic variation characteristics of clean heating power consumption are obtained with the support of large data.

2.5 Recognition of Regional Characteristics

The purpose of dividing the data of clean heating power consumption by heating area is to qualitatively identify the characteristics of clean heating power consumption in different areas, grasp the trend of heating power consumption in different areas, and provide guidance for accurate heating power supply.

In order to identify the regional characteristics of heating power consumption, it is necessary to subdivide the type of electricity consumption data in the identified area, and to refine the type of electricity consumption in the area as far as possible, so as to ensure the accuracy and validity of the identification results [15]. In the conventional classification of regional electricity consumption data, the industrial characteristics of the region are positioned according to the step quantity of electricity consumption. For example, some industries with high heating demand will inevitably lead to the increase of heating power consumption in the region. In addition, different regions have different climate temperatures and different demands for heating, so the data generated are necessarily different [16]. In view of the above two aspects, the data characteristics of regional clean heating power consumption are identified, as shown in Fig. 4.

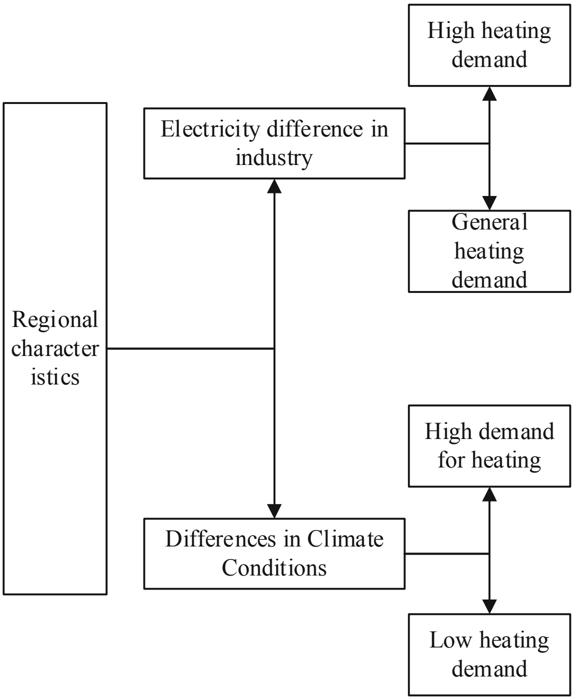

Fig. 4. Regional characteristic recognition classification chart

So far, the identification of electricity characteristics of clean heating based on large data model is basically completed.

3 Analysis of Experimental Demonstration

In order to verify the validity of the method designed in this paper for identifying the electrical characteristics of clean heating, an experimental demonstration is carried out. Taking S city of North China, which is heated by clean heating mode, as the experimental object, the characteristics of its previous heating data are identified, and the electricity characteristics and seasonal periodic characteristics of clean heating power are obtained. The results are shown in Table 2.

Table 2. S identification result of electricity characteristic of clean heating in city

Features	Warm winter condition	Cold winter	Normal behavior
Electricity consumption	2800	3100	3000
Seasonal cycle	3 months	4 months	3.5 months

Then the results of this year's electricity consumption data are analyzed and estimated. The experimental predictions are compared with the actual values. The experimental demonstration and comparison results are shown in Fig. 5.

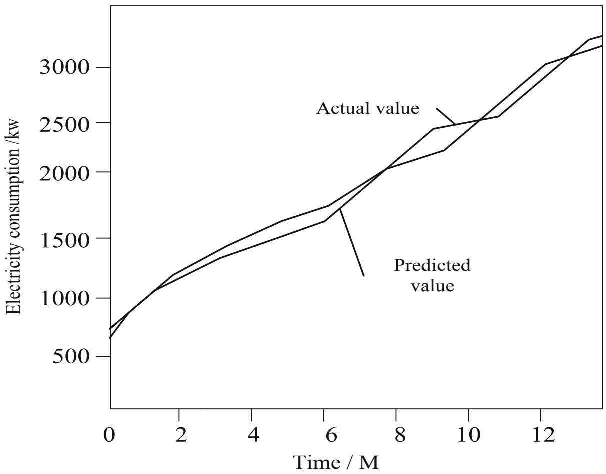

Fig. 5. Experimental demonstration diagram

It can be seen from Fig. 5 that the identification results of electrical characteristics of clean heating by using the identification method proposed in this paper are reliable, and the power consumption of clean heating is predicted according to the identification results. The predicted value is very close to the actual value, which proves the effectiveness of the proposed method.

4 Concluding

Based on the large data model, this paper uses Hadoop platform to integrate the characteristics of the clean heating data, and then identifies the electricity consumption characteristics, periodic characteristics and regional characteristics of the clean heating data, and proves the accuracy and effectiveness of the design through experiments. At the stage of full implementation of clean heating mode, it is hoped that the design can identify the electricity characteristics of clean heating more accurately, and make a more scientific storage and discharge plan for power supply and storage equipment.

References

1. Tang, J.: Precise prediction of monthly electricity consumption based on multi-dimensional feature analysis. Power Syst. Protection Control **45**(16), 145–150 (2017)

2. Peng, S.: Mining method model and application of online learning behavior research in big data era. Audiovis. Educ. Res. **40**(1), 70–79 (2017)
3. Anonymous: Simulation of intelligent data acquisition, operation and maintenance Mining Model for power consumption information under large data. Comput. Simul. **35**(10), 412–415 (2018)
4. Anonymous: Prediction and analysis of electric heating power consumption based on big data platform. Appl. Electron. Technol. **44**(11), 67–69 (2018)
5. Chai, Q., Zheng, W., Pan, J., et al.: Research on condition monitoring and fault handling methods of intelligent distribution network based on large data analysis. Modern Electron. Technol. **41**(4), 105–108 (2018)
6. Anonymous: Performance evaluation of wind turbines based on mutual information association of large wind data. Sci. Technol. Eng. **18**(29), 216–220 (2018)
7. Guo, J., Zheng, X.: A summary of learning analysis based on big data. Audiovis. Educ. China **1**, 121–130 (2017)
8. Vincenti, G., Bucciero, A., Helfert, M., Glowatz, M. (eds.): E-Learning, E-Education, and Online Training. LNICST, vol. 180. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49625-2
9. Li, T., Tan, W., Liu, Z.: Key technologies of smart grid based on big data. Power Supply Technol. **41**(8), 1195–1197 (2017)
10. Yu, Y., Zhu, Z., Huang, C., et al.: Distribution network fault trend judgement based on big data analysis. Power Supply Technol. **42**(1), 132–134 (2018)
11. Cong, H., Zhao, L., Meng, H., et al.: Applicability evaluation of biomass pyrolytic poly-generation technology on clean heating in northern rural of China. Trans. Chinese Soc. Agric. Eng. **34**(1), 8–14 (2018)
12. Ke, G., Tao, D., Qiao, J.-F., et al.: Learning a no-reference quality assessment model of enhanced images with big data. IEEE Trans. Neural Netw. Learn. Syst. **29**(4), 1301–1313 (2018)
13. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902–903 (2019)
14. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mobile Netw. Appl. **24**(1), 1–4 (2019)
15. Fer, I., Kelly, R., Moorcroft, P.R., et al.: Linking big models to big data: efficient ecosystem model calibration through Bayesian model emulation. Biogeosci. Discus. **12**(1), 1–30 (2018)
16. ur Rehman, M.H., Ahmed, E., Yaqoob, I., et al.: Big data analytics in industrial IoT using a concentric computing model. IEEE Commun. Mag. **56**(2), 15–21 (2018)