Research on Power Data Value Mining Technology in the Energy Internet Era

Kunpeng Liu, Ziqian Li, and Yuchen Song(✉)

Customer Service Center of State Grid Corporation of China, Tianjin 300300, China
luyuhan0402@163.com

Abstract. In recent years, electric power enterprises have gradually accumulated a lot of practical experience in data-based applications, but with the development of the energy Internet, the features of interconnection, openness, peer-to-peer and sharing of the energy Internet have given a new connotation to electric power data, which presents characteristics of multiple sources, heterogeneity, large volume, accuracy and real time, etc. With this comes the challenge of data analysis technology and data application development, which also means that more diverse data applications become possible. This paper systematically summarizes several types of existing electric power data applications, analyzes the research direction and application scenarios of electric power data application technologies in connection with the characteristics of the energy Internet, in order to further promote the value mining of electric power data within and outside the enterprise; at the same time, it optimizes and improves the k-means algorithm for the defect that it is easy to fall into the local optimal solution, and uses the improved k-means algorithm to calculate the clustering The results will help the implementation of value-added applications of power data and promote the healthy development of energy Internet.

Keywords: power data · internal and external values · k-means algorithm · data mining · cluster analysis

1 Introduction

After years of development, the power grid has built distribution automation and equipment lean management system, enterprise operations, grid operations and customer service and other business areas and applications at all levels have taken shape, the application of the power Internet of Things has a certain foundation, its business system still has a certain potential to build. The continuous development of smart grids and the in-depth construction of smart distribution grids has led to a sharp increase in the number of collection terminals, a significant increase in the frequency of collection, and an exponential growth in power user-side data and an increase in complexity. At the same time, with the increasing maturity of 5G technology, the integration of 5G technology and big data technology will help the application of big data in many fields such as real-time monitoring of grid status, new energy prediction, user-side portrait depiction

© The Author(s) 2023
J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 1162–1170, 2023.
https://doi.org/10.2991/978-94-6463-198-2_120

and demand-side response, generating new business models [1]. It is very important to tap the deep value of power data and provide data support for public services, policy making and enterprise management.

How to understand more deeply the attributes and characteristics of electricity users, build user portraits, and provide personalized and differentiated services based on electricity user portraits is a hot issue in current user-side big data research. Clustering analysis based on electricity consumption information data can effectively understand and extract the electricity consumption characteristics of users and the commonality of group electricity consumption, which can help realize the segmentation of electricity users, improve the service breadth and depth of power grid companies, and lay the foundation for new intelligent electricity solutions. To address this issue, some scholars have carried out research on the behavior analysis and user classification of electricity users. In [2], a fuzzy integrated evaluation method was used to study residents' preferences and smart electricity consumption behavior based on typical urban residents' smart electricity questionnaire data, but electricity collection data, which can objectively reflect electricity consumption behavior, was not used; in [3], a network analysis method was applied to study the correlation of electricity consumption behavior on short time scale, and a network analysis of electricity consumption behavior correlation clustering and hierarchy analysis was constructed. The literature [4] studied the clustering method of customers' electricity consumption behavior and used it for interactive demand response for the multi-user daily load demand response problem in the complex smart electricity consumption environment; the literature [5] studied the regional scale residential daily electricity consumption load model and model testing method based on cluster analysis, using the average daily electricity consumption of individual households and the maximum daily load of the year as indicators for cluster analysis, but did not consider the data mining performance problem caused by massive data.

With the vigorous development of the information system of power grid enterprises and the continuous improvement of the information level, the amount of power data is also growing dramatically, how to use data management to improve the management level of power grid enterprises, data resources as the company's strategic assets, strengthen centralized management, achieve company-wide information sharing, strengthen data analysis, "all business data, all data business", improve the level of data application and business value, is a challenge at this stage and a long time in the future.

2 The Current State of Application of Power Data

Electricity data involves all aspects of the power system in power generation, transmission, distribution, use and dispatch [6]. The data in the power industry mainly comes from the various aspects of power generation, transmission, transformation, distribution, consumption and dispatch in the production and use of electrical energy, and can be broadly divided into three categories: first, power grid operation and equipment testing or monitoring data; second, power enterprise marketing data, such as data on trading tariffs, electricity sales and electricity consumption customers; and third, power enterprise management data.

Combined with the above data, application scenarios can be divided into internal application scenarios and external application scenarios according to the service targets, and typical application scenarios are described below.

2.1 Typical Application Scenarios Within an Enterprise

With green energy, energy saving and sustainable development becoming the focus of attention worldwide, smart grids have become the inevitable result of economic and technological development, manifested in the use of advanced technologies to improve the performance of power systems in terms of energy conversion efficiency, power utilisation, power supply quality and reliability, etc. Relevant application scenarios include power system transient stability analysis and control, grid load forecasting, modelling and monitoring, as well as new energy consumption analysis.

The finance department of power enterprises has the advantage of accumulating various types of data. Common application scenarios include planning and budget monitoring and analysis, cost-benefit analysis and forecasting, investment decision aid analysis, internal control management analysis, financial risk early warning, etc.

There are many data generated in the field of materials, and the main application scenarios are efficiency and effectiveness analysis of the materials supply chain, materials demand forecasting, materials quality warning, materials supplier analysis, and materials contract default risk identification.

In addition, common application scenarios for research on customers' electricity consumption behaviour include demand-side management/demand response, customer energy efficiency analysis, customer service quality analysis, marketing business assistance analysis such as business expansion and installation, and the deployment of electric vehicle charging facilities.

2.2 Typical Application Scenarios Outside the Enterprise

For external application scenarios, this paper carries out in-depth value mining analysis of power data. The external application scenarios of the company's power data mainly include power user services, energy enterprise services, social public services and commercial value-added services, among which the value of power data on power user services is mainly in demand-side response, energy-saving transformation, intelligent operation and maintenance, intelligent charging, auxiliary analysis of industrial expansion and installation, fault outage management and user interaction, and analysis of power consumption behavior [7, 8]. The value in energy enterprise service is mainly in renewable energy consumption, energy storage operation, power trading, etc. The value in social public service is mainly in industry resumption rate, industrial adjustment decision, economic regulation decision, urban planning, urban management, smart city, etc. The value in commercial value-added services is mainly in data space rental, financial fraud monitoring, advertising targeting, commercial layout planning, and equipment entry detection.

3 Analysis of k-means Algorithm Based Clustering Method for Electric Load Data

With the promotion of smart grid, the electric power industry has accumulated a huge amount of load data. It is called load clustering to divide different types of loads according to the relevant attributes of users and to explore the composition of loads and the relationship between them in different time periods. Scientific and reasonable load clustering can not only help power supply enterprises to realize peak management, but also identify shortcomings and provide data basis for adjusting service strategies, reducing costs and improving energy utilization.

3.1 Temporal Characteristics of Electricity Loads

In order to analyze and calculate the load with respect to the temporal characteristics of the electric load, some characteristic indicators are used.

(1) Load ratio. Daily load rate = total daily electricity consumption / maximum daily load × 24h × 100%; daily minimum load rate = minimum daily load / maximum daily load × 100%.
(2) Load curve. The daily load curve changes with time and is affected by temperature, region and time factors. The curve can reflect the proportional relationship between the total electricity consumption of the day and the product of 24h maximum load, and thus derive the electricity consumption characteristics of different users. According to the peak-valley difference of the load curve, it can realize peak management, grid scheduling, time-of-use tariff and power allocation, etc.
(3) Peak-to-valley load difference. Daily peak-valley difference = daily maximum load - daily minimum load.
(4) Annual maximum load utilization hours. Annual maximum load utilization hours = (annual electricity consumption / annual maximum load) = 8760 × annual load rate.
(5) Simultaneous rate of load. Simultaneous rate = regional maximum load / the sum of the maximum load of each partition × 100%

3.2 Steps of k-means Algorithm

The k-means algorithm is a method that divides n data sets into k-means (k-clusters) according to parameter k, and finally minimizes the sum of squares of the distances from the data points of each cluster to the center of the cluster.

The specific steps of the k-means algorithm are as follows: (1) arbitrarily select k points as the center or mean of the initial clusters; (2) calculate the distances from other data points to the cluster centers; (3) assign the data points to the nearest centers according to the nearest distance principle; (4) calculate the new cluster centers using the mean algorithm; (5) if there is no change in the neighboring centers or the criterion function E has converged, the algorithm ends, otherwise continue iteration; (6) the final k clustering centroids generated and the clustering division centered on it are the final results.

3.3 K-means Algorithm Model

Suppose that n m-dimensional samples are clustered to obtain a sample set $X = \{X_1, X_2 ..., X_n\}$, where $X_i = (X_{i1}, X_{i2} ..., X_{im})$, k classes are denoted as $C = \{C_1, C_2 ..., C_K\}$, and the center of mass $z_j = \frac{1}{n_j} \sum_{x \in c_j} X_i, i = 1, 2, ..., k$ (where n_j is the number of data points in c_j), the goal of clustering is to make k classes satisfy the following equation.

$$\sum\nolimits_{j=1}^{k} \sum\nolimits_{x \in c_j} d_{ij}(x_i, z_j) \rightarrow min \tag{1}$$

where: $d_{ij}(x_i, z_j)$ is the distance calculation function, and the Euclidean distance is chosen to calculate; k is the number of clusters; z_j is the clustering center of sample j.

3.4 Improved k-means Algorithm

For the drawback that k-means algorithm is easy to fall into local optimum, this paper proposes an improved k-means algorithm based on log-adaptive GSA.

(1) Initialize the population size n, the maximum number of iterations T, and the gravitational coefficient decay factor parameters λ, t.

(2) Calculate the value of the adaptation degree.

(3) Update the functions $best(t)$, $worst(t)$ and $M_i\ (t)$.

$$best(t) = \min_{i \in \{1,2,\cdots,N\}} fit_i(t) \tag{2}$$

$$worst(t) = \max_{i \in \{1,2,\cdots,N\}} fit_i(t) \tag{3}$$

$$M_i\ (t) = \frac{M_i(t)}{\sum_{j=1}^{N} m_i(t)} \tag{4}$$

$$m_i\ (t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{5}$$

where: $fit_i(t)$ and $M_i\ (t)$ are the adaptation value and mass of particle i at the t th iteration, respectively; $best(t)$ and $worst(t)$ are the minimum and maximum values of the adaptation value, respectively.

(4) The parameter α is improved to a logarithmic function of t.

$$\alpha(t) = \lambda \times \ln \frac{t + T}{T} \tag{6}$$

where: $\alpha(t)$ is the decay factor of gravitational coefficient; λ is the parameter of α function; t is the number of current iterations; T is the maximum number of iterations.

(5) Calculate the gravitational force, velocity and acceleration of the particle.

$$F_i^d\ (t) = \sum\nolimits_{j \in kbest, j \neq i}^{N} rand_j F_{ij}^d\ (t) \tag{7}$$

$$v_i^d\ (t + 1) = rand \times v_i^d\ (t) + a_i^d\ (t) \tag{8}$$

Table 1. Number of users as a percentage

Load curve clustering category	Number of users/ Households	Percentage
1	18478	54%
2	7186	21%
3	6501	19%
4	2053	6%

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \tag{9}$$

where: $F_{ij}^d(t)$ is the force of particle j on particle i at the t th iteration in d dimensions; $F_i^d(t)$ is the total force of particle i at the t th iteration in d dimensions; $v_i^d(t+1)$ and $v_i^d(t)$ are the velocities of particle i at the t th and $t+1$ th iterations in d dimensions; $a_i^d(t)$ is the acceleration of particle i at the t th iteration in d dimensions; and *rand* is a random variable between 0 and 1.

(6) Update the particle position, the value of fitness and the global optimal solution.

(7) Repeat the iterations until the termination condition is satisfied.

(8) The optimal solution is obtained and used as the initial clustering center for k-means algorithm clustering.

3.5 Algorithm Analysis

To verify the effectiveness of the algorithm proposed in the article, the daily load curve of a city in China is selected and studied by clustering through the traditional k-means algorithm and the improved k-means algorithm in this paper. The number of clusters k = 4 is set, and the load curve is clustered and analyzed, and the actual number of users in each category of the study is shown in Table 1. Each category of users can be further divided into five types according to the industry attributes, and the specific electricity consumption types are shown in Table 2.

3.5.1 Traditional k-Means Algorithm Electric Load Clustering Results

Let the number of clusters k = 4, and use the traditional k-means algorithm to calculate the clustering results as follows.

(1) Three-peaked curve. In the three time intervals of 8:00–9:00, 11:00–12:00, and 17:00–21:00, the curve has a rising trend, and 1 ~ 2h after each interval, the curve shows a downward trend, and the curve is basically in a stable state in the remaining time. 20:00 load The maximum power is 10.2kW.

(2) Evening peak curve. In the three time periods of 7:00–9:00, 10:00–12:00 and 17:00–22:00, the curve has an upward trend, and 1 ~ 2 h after each interval, the curve shows a downward trend, and the remaining time curve is basically in a stable state. 20:00 load The power is the largest, 12kW [9].

Table 2. Statistics of users' electricity consumption types

Load curve clustering categories	Number of customers with different types of electricity use/ Households					Grand total
	Non-industrial electricity	Residential electricity	Industrial electricity	Commercial electricity	Other	
1	7245	3948	3954	5012	1441	18478
2	2169	2498	1082	1160	277	7186
3	1875	1237	1943	1401	145	6501
4	1022	408	306	302	15	2053

(3) Smooth curve. In 10:00–13:00 and 17:00–24:00, the curve has slight fluctuations, and the remaining time curve is relatively smooth. 11:00 has the maximum load power of 13.8 kW.

(4) Staggered curve. At 10:00–11:00 and 17:00–21:00, the curve has a rising trend, and at 1:00–9:00 and 21:00–24:00, the curve shows a downward trend. 21:00 has the maximum load power of 8.6 kW.

3.5.2 Improved k-means Algorithm Electric Load Clustering Results

The clustering results are calculated using the improved k-means algorithm as follows.

(1) Smooth curve. 10:00–14:00, 17:00–24:00, the curve has slight fluctuation, the remaining time curve is relatively smooth. 11:00 load power is the largest, 13.4kW, users mainly use electricity for commercial use.

(2) Evening peak curve. 1:00–9:00, the load power is low; 11:00–13:00, the curve rises slightly; 17:00–20:00, the curve rises sharply. 20:00, the load power is the largest, 11.8kW, and the users are mainly residents.

(3) Three-peak curve. In the three periods of 6:00–7:00, 10:00–11:00 and 17:00–18:00, the curve has a rising trend, and the maximum value appears at 11:00; in the periods of 7:00–9:00, 11:00–15:00 and 21:00, the curve has a rising trend. -15:00, 21:00–24:00, the curve has a downward trend, with a minimum value at 15:00. The users' electricity consumption is dominated by industrial electricity consumption.

(4) Staggered peak curve. 0:00–7:00, 18:00–24:00 two intervals of power; 7:00–18:00 interval power is smaller. 21:00 load power is the largest, 8.5kW, user power and electricity consumption staggered peak period, users are mainly non-industrial users. Industrial users are the main users.

3.5.3 Conclusion

The experiments show that the improved k-means algorithm has good search ability, the initial clustering center is closer to the actual one, the number of iterations is less, the convergence speed is faster, and the final clustering results are more accurate.

4 Summarization

Big data is not only a comprehensive technology, but also a science. At present, the analysis and application of electric power data involves various fields within the enterprise, but the rapid development of technology and the growing demand have posed higher and higher challenges to the application of electric power data, and it is still necessary to continuously research and explore the application of electric power data in the fields of enterprise management, economic efficiency and value-added services in the future, and at the same time to quickly improve the level of data management within the enterprise and enhance the data service capability.

This paper mainly focuses on the application value mining of electric power data. Under the background of integration and sharing of large amount of electric power data, this paper investigates the current demand for electric power data from various types of enterprises, introduces the current more mature internal and external application scenarios, and carries out in-depth value mining analysis of electric power data, summarizes the technical development direction worthy of future research, and hopes to further promote the value of electric power data in the energy Internet era. The traditional k-means algorithm is also improved, and the experiments prove that the improved algorithm has shorter time-consuming optimization search, smaller deviation between initial clustering and actual, and better clustering effect on electricity load data mining, but further research is needed on the practical application of clustering results on subsequent decision-making impact, load prediction, and management approaches.

References

1. Wang Yi, Chen Qixin, Zhang Ning, etc.. The integration of 5G communication and ubiquitous power internet of things: Application Analysis and research prospect [J] . Grid technology, 2019,43(5) : 1575–1585.
2. He Yongxiu, Wang Bing, Xiong Wei, Zhang Ting, Liu Yangyang. Analysis of residential intelligent power consumption behavior based on Fuzzy Comprehensive Evaluation and design of interactive mechanism [J] . Grid technology, 2012,36(10) : 247–252. DOI: 10.13335/J. 1000–3673. October 07,2012.
3. Chen Pengwei, Tao Shun, Xiao Xiangning, Li Lu, Zhang Jian. A network model for short-time scale correlation analysis of electricity consumption behavior [J] . Power system automation, 2017,41(03) : 61-69
4. Lu Jun, Zhu Yanping, Peng Wenhao, Qi Bing, Cui Gaoying. Interactive demand response method for intelligent community considering power consumption behavior clustering [J] . Power system automation, 2017,41(17) : 113-120
5. Xu Jieyan, Xu Wenyang, Chu Yuan, Jin Yuan, Kang Xuyuan, Chen Zheng. Study on the construction method of daily electricity load model of residential buildings at regional scale [J] . China electric power, 2020,53(08) : 29-39
6. Wu Kaifeng, Liu Wantao, Li Yanhu, etc.. Power Big Data Analysis Technology and application based on cloud computing [J] . China electric power, 2015,48(2) : 111–116.
7. Xin Miaomiao, Zhang Yanchi, Xie Da. A review of consumer behavior analysis based on big data [J] . Electrical automation, 2019,41(1) : 1-4.

8. Liu Yufang, Gao Qian, Xu Chao, etc. . Power big data value and application demand analysis [J] . China management informatization, 2018,21(20) : 52 -54.
9. Jin Zhiyu, Wang Maomao, Shi Huilei. Research on power load clustering based on DBSCAN and improved K-means clustering algorithm [J] . Northeast electric power technology, 2019,40(06) : 10-14.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.