Analysis of Load Characteristics of Typical Large Industrial Users in Hunan Province Based on K-means Clustering

Ming Wen[1], *Yong* Chen[2], *Fenkai* Chen[3,*], *Lili* An[2], *Bo* Chen[2], *Ting* Zeng[4], *Qing* He[3], and *Chengran* Fu[3]

[1]Economic and Technology Research Institute, State Grid Hunan Electric Power Company Limited, Changsha, 410004 , China
[2]State Grid Hunan Electric Power Company Limited, Changsha, 410000, China
[3]School of Economics and Management, North China Electric Power University, Beijing, 102206, China
[4]Changsha Planning Design and Research Institute of Exploration and Surveying, Changsha, 410007, China

Abstract. The analysis of load characteristics of large industrial users is the basis for understanding the way of using electricity and analyzing its electricity consumption behavior. The power load of 13 typical large industrial users in Hunan Province was selected, and based on K-means clustering method and distance equalization function, the optimal number of clusters for large industrial users was determined, and the user classification was finally realized to analyze the load characteristics of each type of users. The results show that the load characteristics of each type of users have a certain degree of difference, and the trend of power consumption trends is different, but the daily average load curve fluctuations of each type of users are basically the same, which are consistent with the law of electricity consumption.

1 Introduction

With the continuous reform of the electric power system, the volume of power users in various industries has increased, and the behavior of electricity consumption has been diverse. It is of great significance to study the load characteristics of power users [1]. In China's industrial structure, large industrial users account for a large proportion, there would be difficult to analyze the load characteristics of large industrial users one by one. Therefore, users with similar trends in load curves can be grouped into one group by clustering method [2], and then study the power consumption rules of all kinds of users.

Nowadays, the research focus of various experts and scholars is different for the study of load characteristics. First, most studies analyze the grid load in a given area [3-4]. For the study of power consumption characteristics of large power users, the literature [5] uses big data technology and DTW algorithm to study the similarity and difference of various user load curves; the literature [6] first analyzes the electricity consumption behavior of the residential, industrial and commercial users, followed by fuzzy clustering method to achieve industrial user classification, study industrial user load characteristics; the literature [7] studied the daily load curve of different users in the same industry, and use fuzzy C-means method to achieve similar user division and type identification. Secondly, to analyze user load characteristics, the methods used to implement user classification are different, such as fuzzy clustering method [6], K-means clustering method [5, 7] and neural network clustering method [8]. Because the K-means clustering algorithm solves the problem that the

centralized system framework is difficult to solve the massive data processing and analysis, and can optimize the selection of the initial central data [9], This paper selects the K-means clustering method to cluster the typical large industrial users in Hunan Province and study its load characteristics.

In this paper, the K-means clustering method and the distance equalization function are combined to analyze the annual average daily load curves of 13 typical large industrial users in Hunan Province, and analyze the load characteristics of each type of large industrial users, and provide a basis for studying the user's electricity consumption behavior.

2 The classification of large industrial users based on K-means clustering method

The K-means clustering method is a typical partitioning clustering method, and it is also a distance-based partitioning clustering method. The clustering results can maximize the similarity within the group and maximize the difference between groups. This paper uses K-means clustering method to subdivide large industrial users in Hunan Province. The specific idea is shown in Figure 1.

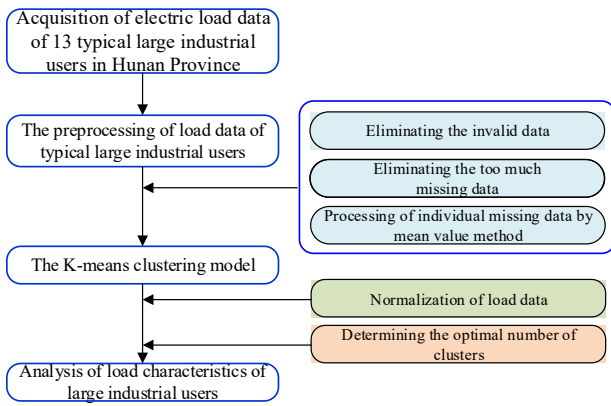---

* Corresponding author: 1090962561@qq.com

Fig.1. The research idea of the clustering analysis for typical large industrial users based on K-means clustering.

2.1. The preprocessing of typical large industrial user load data

The lack of load data of large industrial users will affect the correctness of the clustering results. Therefore, it is necessary to perform the missing value processing on the typical large industrial users' power load data before clustering, which generally includes the following three schemes.

(1) Eliminating the invalid data;

(2) Eliminating the too much missing data;

(3) Processing of individual missing data by mean value method: Each user's electricity consumption behavior has a certain regularity. Therefore, the same load point data before and after a certain load data point can be used, and the data can be filled by the mean method.

2.2 The K-means Clustering Model

2.2.1 Normalization of load data

Because different types of users have different maximum loads, at the same time, in order to eliminate the heteroscedasticity of load data, it is necessary to normalize the acquired data, and the data range after processing is [0,1]. Secondly, it is analyzed by using the standard unitary value, the process is as follows(1):

$$\overline{L_i} = \frac{L_i}{L_{max}} \qquad (1)$$

Where, i denotes the time, and the range of values is [0-23]; $\overline{L_i}$ is the data after the load is processed as a unit value; L_i denotes the user's load at time i; L_{max} denotes the maximum of the user's daily load.

2.2.2 Determination of the optimal number of clusters

The traditional K-means clustering method needs to specify the number of clusters in advance. This method

has certain randomness and the results are not representative. Therefore, the distance equalization function [10] is used to determine the optimal number of clusters. The distance within the class is minimized and the distance between classes is maximized.

2.2.3 K-means clustering model

For a given set of related samples D={$X_1,X_2,..., X_n$}, Firstly, K points are determined as the center of the cluster, secondly, the nearest sample points are found according to the related distance measure, which are grouped into the nearest cluster. At the same time, the center of the cluster is updated, and the above operations are performed iteratively until the data does not change or the center of the cluster does not change. At this time, the divided cluster C={C_1, C_2, ..., C_k} is output [11]. Then the loss function of the cluster is as shown in equation (2).

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \qquad (2)$$

Where, k is the total number of clusters; μ_i is the average of the cluster C_i.

3 Case analysis

3.1 Basic data

This paper selects the electricity load data of 13 representative industrial users with high proportion of electricity consumption in Hunan Province in 2017, and conducts cluster analysis of annual average daily load curve. For the privacy of user information, this article will number the 13 major industrial users from 1-13. Among them, the types of large industrial users involved are paper industry, steel industry, non-metal industry, chemical industry, electrical and electronic equipment manufacturing, and automobile manufacturing. The types of users are shown in Table 1.

Table 1. The percentage of users of various types

Type of business	Number	Proportion of various types/%
Paper industry	3	16.67
Steel industry	3	14.17
Non-metal industry	1	6.67
Chemical industry	3	11.67
Electrical and electronic equipment manufacturing	2	13.33
Automotive Manufacturing	1	6.67
Total	**13**	**100%**

3.2 Basic data

When clustering the annual average daily load curve using the K-means method, it is necessary to first

determine the number of classifications. In order to make the classification of the classified categories more obvious, the optimal number of clusters is determined by the distance equalization function in Section 2.2.2, namely $2 \leq k \leq \sqrt{13}$, $k = 3$. The number of users included in each category is shown in Table 2, and each type of user cluster center is shown in Figure 2.

Table 2. The number of users in each category

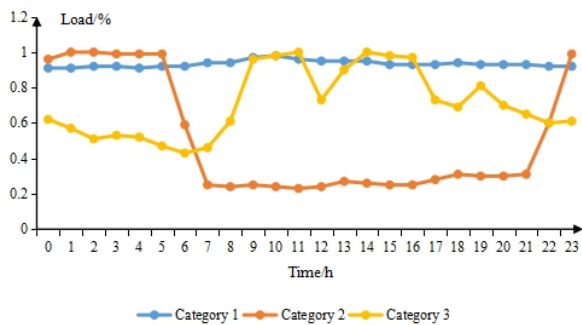| Category	Number
1	10
2	2
3	1

Fig.2. The load characteristics carve of large industrial users after K-means clustering.

As shown in Fig.2, the three types of load characteristics clustered based on the annual average daily load curve of large industrial users have a certain degree of difference. The annual average daily load curve of large industrial users in the first category has a relatively stable trend, with small peak-to-valley difference, indicating that the power equipment utilization efficiency is higher, and the electricity consumption is more economical; the electricity consumption of large industrial users in the second category is mainly concentrated from 22:00 to 7:00 the next day, which is a typical peak-avoidance user; the annual average daily load curve of the large industrial users in the third category shows two peaks and one valley. And the load reached the maximum at 8:00-12:00 and 14:00-17:00, and the shift time was reached at around 12:00, and the load reached the valley.

3.3 Analysis of load characteristic curves of various large industrial users

In this paper, the monthly maximum load curve and the quarterly average daily load curve of each type of large industrial users are plotted and analyzed to study its load characteristics.

3.3.1 K-means clustering model

This section mainly analyzes the monthly maximum load of three types of large industrial users and plots them in Figure 3.
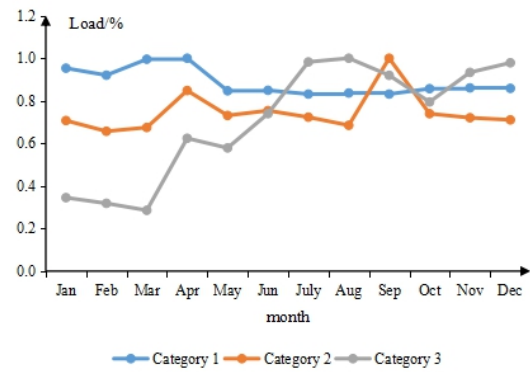

Fig.3. The monthly maximum load curve for various large industrial users.

As shown in Fig.3, the maximum load of the first type of large industrial users has not changed much in one year, and the maximum load in March and April is large; the maximum load of the second type of large industrial users reached the maximum in September. The maximum load in December and January and February is small; the maximum load of the third category of large industrial users varies greatly between months, and the maximum load reaches the maximum in July and August, and reaches the minimum in January-April.

3.3.2 Quarterly average daily load curve of large industrial users

The spring is March, April, and May, the summer is June, July, and August, the autumn is September, October, and November, and the winter is December, January, and February. The average daily load curve for each quarter is analyzed as shown in Figure 4.
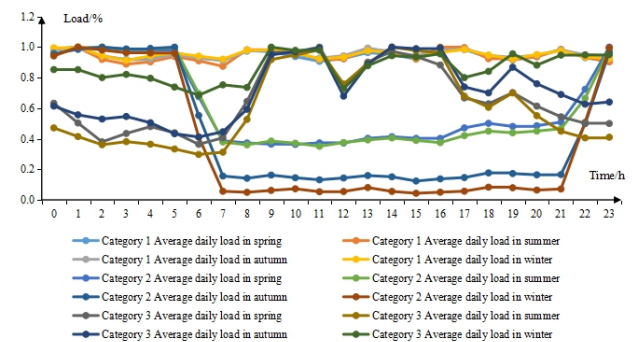

Fig.4. The quarterly average daily load curve of various large industrial users.

As shown in Fig.4, the daily average load curve fluctuations of the three major users in each season are basically the same, which are consistent with the rules of electricity consumption. Among them, in terms of seasonal differences, the average daily load of the first-class large industrial users is basically the same; the average load of the second-class users in autumn and winter is lower, and the average daily load in spring and summer is basically the same; the average daily load of the third-class large industrial users in winter is higher,

and the average daily load in the rest of the season is basically the same.

4 Conclusions

This paper uses the K-means clustering method to cluster analysis of large industrial users, and obtains the following conclusions:

(1) With the distance equalization function, 13 typical large industrial users are divided into three categories. The first type of users use electricity more smoothly, which is a smooth power user; the second type of users use electricity at 22:00 - 7:00 the next day, which is a typical peak-avoidance user; the third type of user has the 2 peak value, and reached the maximum during the peak period of the system, called the peak user.

(2) Due to the periodicity and regularity of the load characteristics, it is possible to find out the inherent rules through load characteristic analysis, so as to grasp the power consumption of various users, such as using curve analysis to analyze the load characteristics of various users. It is found that the trend of the quarterly average daily load curve of various users is consistent with the rules of electricity consumption.

Acknowledgment

References

1. W. Chen, L.Q. Yue, K. Cui, et al. Analysis on Power Load and Consumption Characteristics of Representative Consumers[J]. Electric Power Technologic Economics, 2011, 23(9): 44-49.

2. G.P. Zhou, Z.J. Sun, J. Fu, et al. Load Characteristics Analysis of Power Industry Based on Fuzzy C-means Clustering Algorithm[J]. Electrotechnical Application, 2016, 35(16): 86-90.

3. X. Sun, H.Y. Zeng, W. Liang, et al. On an Analytical Approach to Load Characteristics of Urban Power Grids[J]. Journal of Hunan University of Technology, 2016, 30(3): 43-48.

4. Y. Lan, J.Z. Tan, X. Li, et al. Study on Regional Power Load Characteristics[J]. Electric Power Technologic Economics , 2010, 22(7): 30-35.

5. J.L. Shen, C.G. Lu, J. Yuan, et al. Application of Big Data Technology in Power Consumption Feature Classification of Large Customers[J]. Zhejiang Electric Power, 2017, 36(12): 37-41.

6. M. Cai, P.F. Shen, H. Zhu, et al. Load Characteristic Analysis of Distribution Grid Based on Fuzzy Clustering[J]. Smart Power, 2015, 43(4): 24-27+39.

7. Z.H. Li, X.B. Yin, H.J. Yang, et al. Seasonal Load Characteristics Analysis Based on Improved K-means Clustering Algorithm[J]. Power System and Clean Energy, 2018, 34(02): 53-59+64.

8. Y.J. Xu, X.Q. Huang, Y.J Cao, et al. Aggregation of Air Conditioner Load Based on Self-organizing Feature Map Neural Network[J]. Proceedings of the CSU-EPSA, 2015, 27(11): 26-33.

9. Y.F. Yi. Data Analysis Based on K-Means Clustering[J]. Modern Manufacturing Technology and Equipment, 2017(4): 8-9.

10. Q. Fu. A Study on K-means Algorithm for Determining Optimal Number of Clusters[J]. Journal of Nanchang Junior College, 2011,26(02):158-159.

11. Y.X. He, Z. Jiao, F.K. Chen, et al. Residential power user segmentation based on k-means clustering method in the context of big data[C]//2018 3rd International Conference on Advances in Energy and Environment Research , 2018.