2018 International Conference on Chemical, Physical and Biological (ICCPB2018)

Analysis of Electricity Use Behavior with Clustering Method and Classification of Peak and Valley Periods

1st Ke Xu
Hubei electric engineering corporation limited
Wuhan, Hubei province, 430040

2nd Yufeng Xia*
Hubei electric engineering corporation limited
Wuhan, Hubei province,430040
181206972@qq.com

3rd Wenbin Ni
Hubei electric engineering corporation limited
Wuhan, Hubei province, 430040

4th Ting Li
Hubei electric engineering corporation limited
Wuhan, Hubei province,430040

***Abstract*—Based on the analysis of electricity price during peak period of electricity consumption, cluster analysis is carried out on the electricity consumption behavior of users who use electricity overload. The way of analysis is to make data analogy between the curve obtained by clustering analysis and the daily overload curve obtained under normal circumstances and the typical clustering algorithm of overload property, and confirm that the results obtained by clustering analysis algorithm are conclusive and effective. The clustering analysis of power consumption data of a large number of users also determines the algorithm used in clustering analysis on the basis of potential indicators, and its analysis results are more accurate.**

Keywords—load index, regulation potential, time-sharing price, cluster analysis

I. USER BEHAVIOR ANALYSIS BASED ON KMEANS CLUSTERING ALGORITHMS

A. Kmeans Clustering Algorithms

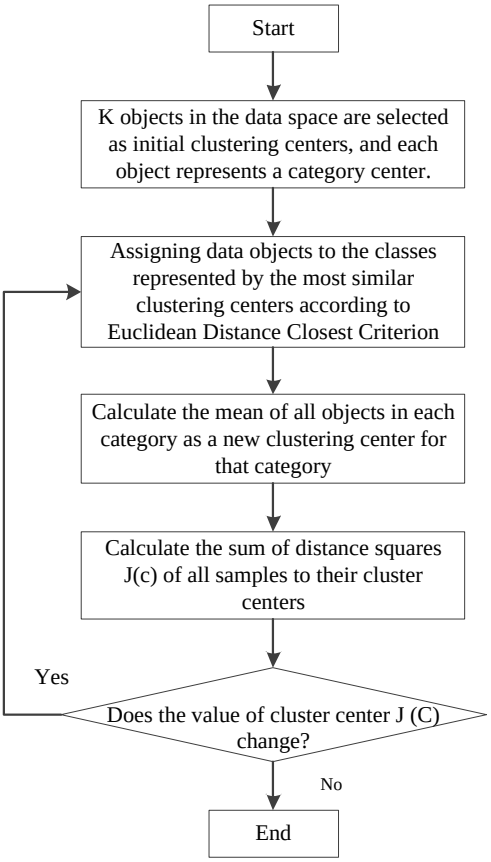

Figure 1 Kmeans clustering algorithm steps

In the analysis of power consumption data, accurate clustering analysis of a large number of data is a necessary part of the research content. Through the analysis of user's overload power consumption, we can detect the user's power consumption standard at all times, which can adjust the user's power consumption. When users use electricity in daily life, the change of

  DOI: 10.25236/iccpb.2018.024

power consumption may be affected by many reasons. This article carries out clustering analysis on the basis of users' daily overload data, and classifies users with the same trend of overload curve, and classifies different trends into different categories, and carries out research on users' electricity consumption data. In this paper, Kmeans clustering algorithm is used to cluster overload data. Kmeans clustering algorithm belongs to hard clustering algorithm. Firstly, the initial starting point of clustering is selected. Secondly, the user's electricity data are classified. Finally, the average value of the classified data is calculated. According to the change of average value, the clustering centers are adjusted continuously, so that the data of the same species are as close as possible, and the objects of the same species are as similar as possible. The Kmeans clustering algorithm needs to analyze the initial data with the cluster number K and the object number N. If the cluster center does not change, it needs to input the cluster number K further. The flow calculation is as follows (Fig. 1):

B. Analysis on the Basis of Overload Data of User'S Electricity Consumption

On the basis of Kmeans clustering algorithm, this paper uses the daily overload data of users as the analysis object, namely DC-Kmeans clustering.

(1) Data collection

In the records of power companies, the daily overload of users is searched for and sampled. The collection of data is recorded as X.

$$X = \left[x_1, x_2, \cdots, x_k, \cdots, x_M\right] \tag{1}$$

As shown in the above formula, the power data of 96 sampling points per day for any one user load of M sample data is set to x_k , as follows:

$$x_k = \left[x_{k,1}, x_{k,2,} \cdots, x_{k,96}\right] \tag{2}$$

(2) Initialization of data

The trend of overload curve needs to be analyzed and the similarity of curve needs to be considered. In order to exclude the influence of overload data on research results, it is necessary to classify overload data by taking the maximum daily overload of users as a benchmark. Considering the change of overload curve caused by sudden change of sampling device at special time, which affects the result of overload clustering analysis, this paper uses smoothing formula to correct the sudden change of data.

$$\dot{x}_{k,t} = \frac{\sum_{a=1}^{a_1} \dot{x}_{k,t=a} + \sum_{b=1}^{b_1} \dot{x}_{k,t+b}}{a_1 + b_1} \tag{3}$$

The values of a and b in the expression take the values in the range of front and back of the midpoint of clustering respectively.

(3) Cluster analysis of data

At last, Kmeans clustering algorithm is used to analyze overload data. Based on the analysis of distance square and clustering effect evaluation parameters, K user groups are obtained. For the same users, there are similarity curves. Therefore, the same kind of overload users are classified as the same kind of users and the cluster center is selected as the corresponding equivalent overload curve.

II. RESEARCH INDICATORS ON THE BASIS OF USER POWER CONSUMPTION

A. Normal Overload Index Criteria

Because only clustering analysis of the nature of the daily overload curve needs large-scale database. Therefore, taking the daily overload curve as the initial value, the important factors for the user's usual indicators are: overload rate, the difference between peak and valley, and the maximum number of hours used for overload. These factors all affect the power consumption of overload users, and profoundly show the power consumption of each market using the usual indicators for analysis (DCKmeans clustering algorithm). It affects many basic properties of overload. After the initial input of data, the maximum value of overload needs to be transformed by the parameters of hours and overload rate. After the initial input of data, the maximum value of overload needs to be transformed by the parameters of hours and overload rate. It can also be concluded that cluster analysis using overload parameters leads to the decrease of sampling parameter coefficients, while cluster analysis cannot accurately describe the trend of overload curve, but it can express the daily power consumption of users. According to the above research, the example of this paper chooses the peak and valley tariffs to calculate the clustering parameters, and chooses the peak overload rate σ * f, the usual overload rate σ *p, the peak and valley overload rate σ * g to calculate the clustering parameters.

$$\overset{*}{\sigma}_f = \frac{\sum_{t_i = t_f} \overset{*}{x}_{k,t_i}}{\sum_{t_i = t_f} t_i} \tag{4}$$

$$\overset{*}{\sigma}_p = \frac{\sum_{t_i = t_p} \overset{*}{x}_{k,t_i}}{\sum_{t_i = t_p} t_i} \tag{5}$$

$$\overset{*}{\sigma}_g = \frac{\sum_{t_i = t_g} \overset{*}{x}_{k,t_i}}{\sum_{t_i = t_g} t_i} \tag{6}$$

III. INTRODUCTION

In this part, 300 users' electricity data collected by a city's power company are used as the analysis object, and IC-Kmeans algorithm and AP-Kmeans algorithm are used to study.

Computation based on IC-Kmeans algorithm: IC-Kmeans algorithm is used to calculate the clustering index of 300 users, and the distance square and parameter index required by IC-Kmeans algorithm for large numbers of users. The best results are shown in Figure 2.

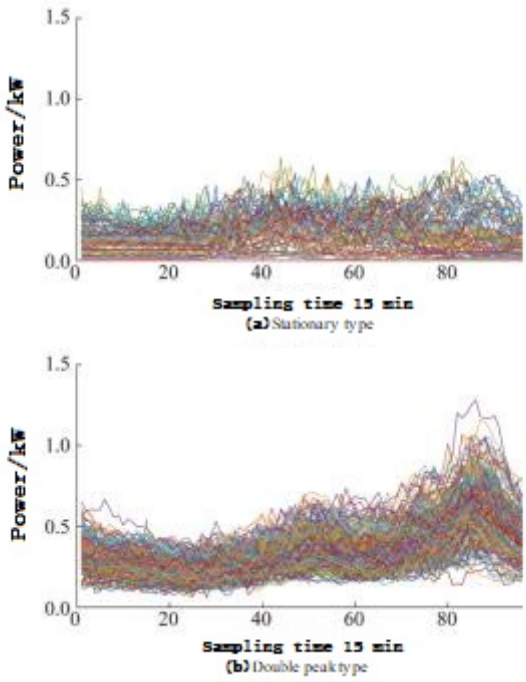

Figure 2 Clustering analysis of a large number of users based on IC-Kmeans algorithm

As can be seen from Figure 2, clustering analysis of large numbers of users on the basis of IC-Kmeans algorithm can be divided into two categories: stationary users and bimodal users. Taking stationary user 1 as an example, the difference of overload data is very small, while bimodal user cluster analysis is based on 12 points and AP-Kmeans algorithm. The best result in Figure 3 is 3, as shown in Figure 4.

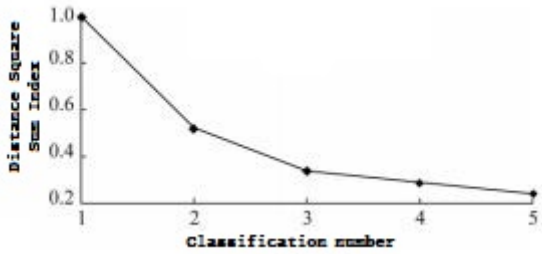

Figure 3 Distance Square Sum Index Based on AP-Kmeans for Large-scale Users

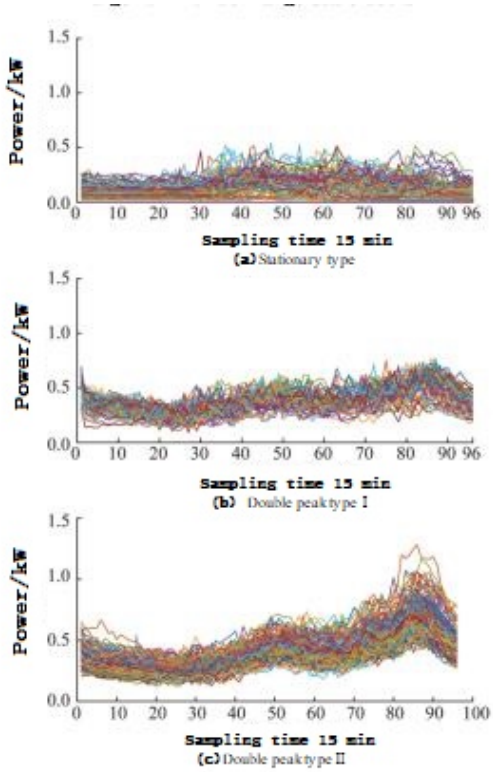

Figure 4 Large-scale user clustering based on AP-Kmeans algorithm

From the analysis of Figure 4, based on the AP-Kmeans algorithm, we can conclude that the types of users can be divided into three types: stationary users, bimodal I users and bimodal II users. The results of ICKmeans algorithm are similar to those of stationary user overload curve, and the number of users' electricity consumption decreases. There are two overload peaks in double-peak type I and double-peak type II, but the difference between two types of double-peak type is larger. The results of IC-Kmeans algorithm analysis are compared, and the clustering results of each user are compared. The average value of distance square and parameter ρm are used to evaluate the clustering results. The higher the value of ρm, the worse the clustering results. This index avoids considering the factor of user number, and the ρm value changes with the clustering effect of user type. The results of clustering analysis are compared in Table I.

TABLE I COMPARISON OF LARGE-SCALE USER CLUSTERING RESULTS

User category	N_m		ρ_m	
	$IC-K_{means}$	AP - K_{means}	$IC-K_{means}$	AP - K_{means}
Stationary type	112	96	0.79	0.75
Double peak type	188		0.69	
Double peak typeI		98		0.58
Double peak typeII		106		0.52

From the analysis results in Table I, we can see that compared with IC-Kmeans algorithm, the number of stable users using AP-Kmeans algorithm decreases, while the number of users with double-peak type I and II increases. According to the analysis

of the two algorithms, the APKmeans algorithm has the smallest value of ρ m, and its effect is the best, and it is the strongest stability in most types. As a result, compared with the typical IC-Kmeans clustering algorithm, the AP-Kmeans clustering algorithm is more accurate for a large number of users. The reason is that the APKmeans clustering algorithm is based on the adjustment potential index. When it is at the peak and valley, the index has a big gap. This is the same as the difference of the power consumption index between the expanded peak and valley. Therefore, for the power users in the peak and valley, the analysis results of AP-Kmeans clustering algorithm in this article are closer to real life.

IV. CONCLUSION

In this paper, the model is used to represent the adjustment potential index of users, and the clustering analysis of overload data of users is carried out based on the reference of this index. Compared with the clustering analysis results of daily overload curve and overload parameters, the clustering analysis algorithm used in this paper is more accurate. When clustering analysis is used for a large number of users, the proposed algorithm can ensure that the clustering efficiency can still be guaranteed within the prescribed range of clustering effect. Taking the regulation potential characteristics of users as the analysis factor, the clustering effect is more suitable for users with significant peak-valley characteristics, and can be used as the criterion of user selection for demand response adjustment.

REFERENCES

[1] Fermín Moreno. Characterization of Peaks and Valleys of Electricity Demand. Application to the Spanish Mainland System in the Period 2000-2020[J]. Energy and Power Engineering, 2011, 3(4):10.

[2] Cheng R, Xu D, Han H , et al. Selection of water delivery schedule based on peak-valley electricity price and annual cost usage[J]. Transactions of the Chinese Society of Agricultural Engineering, 2013, 29(s 10–13):60-64.

[3] Shao-Yun G E, Liu H, Hong L. Optimization of peak-valley TOU power price time-period in ordered charging mode of electric vehicle[J]. Power System Protection & Control, 2012.

[4] Veall M R . Time-of use rates and peak period electricity consumption : An empirical note[J]. Energy Economics, 2006, 8(4):257-262.

[5] Rogers D F, Polak G G. "Optimal Clustering of Time Periods for Electricity Demand-Side Management"[J]. IEEE Transactions on Power Systems, 2013.

[6] Nuchprayoon S . Electricity load classification using K-means clustering algorithm[C]// Brunei International Conference on Engineering & Technology. IET, 2015.

[7] Hui-Ting Q, Yi-Jie L. Research of peak and valley period partition approach on statistics[C]// International Conference on Electric Utility Deregulation & Restructuring & Power Technologies. IEEE, 2011.

[8] Nojavan S, Mohammadi-Ivatloo B , Zare K . Optimal bidding strategy of electricity retailers using robust optimisation approach considering time-of-use rate demand response programs under market price uncertainties[J]. Generation Transmission & Distribution Iet, 2015, 9(4):328-338.

[9] Mcloughlin F, Duffy A , Conlon M . A Clustering Approach to Domestic Electricity Load Profile Characterisation Using Smart, Metering Data[J]. Applied Energy, 2015, 141:190-199.

[10] Li Y, Wang Z, Lu Y, et al. The implementation of peak and valley time price for electricity and the response of large industries[J]. Automation of Electric Power Systems, 2001.

[11] Tang J, Ren Z, Hu X Z . An operable peak-valley time-of-use tariff setting method for demand side management[J]. Power System Technology, 2005, 29(22):71-75.

[12] Haitao H, Jiejing W U, Danzhen G U , et al. Pricing model of time-of-use electricity tariff considering customers classified by load factor[J]. Power System Protection & Control, 2016.