International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0098

CLUSTERING ANALYSIS OF USER POWER INTERACTION BEHAVIOR BASED ON IMPROVED K-MEANS ALGORITHMS

Dan Wang[1,2*], Bingyu Zhou [1,2], Bo Liu [1], Pengfei Su[1,2], Qi Yang[1,2]
1 Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China (Correspond Author)
2 Qindao Institute for Ocean Technology of Tianjin University, Qingdao, Shandong Province 266235, China

ABSTRACT

The interactive grid is the basic model of the modern global power grids, analysis of users' interaction power behavior is a core task. This paper firstly uses self-organizing map SOM neural network training and artificially separated methods to optimize the initial clustering center of K-means algorithm. Then, under the background of peak-to-valley time-of-use electricity price, the adjustment potential index based on user psychology is constructed, and the users' electricity consumption behavior based on load data and adjustment potential index is analyzed. Finally, the clustering results of the two improved algorithms and the clustering results of classical K-means algorithm are compared. By comparing the advantages of K-means++ algorithm in accurate identification and clustering of users' electricity consumption behavior, the effectiveness of K-means++ clustering center selection process in significantly shortening clustering time is analyzed.

Keywords: electricity consumption behavior, cluster center optimization, load data, adjustment potential index, clustering analysis.

1. INTRODUCTION

Under the background of hierarchical, digitized and informatized interactive power consumption, mining and clustering analysis of power information such as load power and adjustment potential of a large number of power users is an important part of power demand side management [1-2]. At present, there are some researches on user load clustering analysis. In literature [3], a method of dimensionality reduction clustering of daily load curve based on singular value decomposition is proposed. K-means algorithm is used to cluster daily load curve. In reference [4], a conditional filter is designed for load feature extraction, and the whole day is divided into five periods to select peak and valley values. Six-dimensional feature data are selected for load clustering. Reference [5] proposing a new distributed clustering algorithm based on adaptive K-means for large distribution data.

The above research improves the K-means clustering process in the data processing perspective, and does not improve from the algorithm itself to solve the problems of the K-means algorithm. This paper will start from the classical K-means algorithm, improve its clustering center primary selection process, cluster analysis based on user load and adjustment potential indicators, and analyze the results of improved algorithm in user electricity behavior cluster analysis.

2. DESCRIPTION OF IMPROVED K-MEANS CLUSTERING ALGORITHM

K-means algorithm has a wide range of applications in large-scale data mining and processing. However, the clustering results are greatly affected by the initial clustering center.

In order to improve the selection of initial clustering centers of K-means algorithm, two algorithms are adopted in this paper: 1) Input data into self-organizing maps (SOM) network for initial clustering, as shown in Figure 1, and obtain preliminary accurate clustering results. These points are used as initial clustering centers of K-means algorithm, i.e. the SOM+K-means algorithm, the clustering process of the algorithm is shown in Figure 3. 2) In the process of selecting the initial clustering center of the K-means algorithm, as shown in Figure 2, these centers are artificially separated from each other, and the point farther from the current cluster center has a higher probability of being selected as the new cluster center, namely K-means++ algorithm, the clustering process of the algorithm is shown in Figure 4.

Selection and peer-review under responsibility of the scientific committee of the 11th Int. Conf. on Applied Energy (ICAE2019).

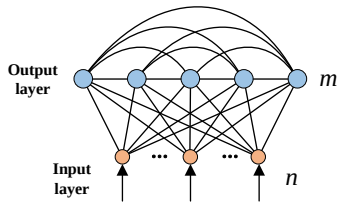

Fig. 2. Primary selection principle of SOM network clustering center

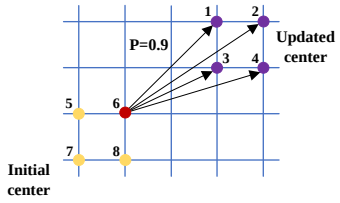

Fig. 2. The principle of initial clustering center optimization of K-means++ algorithm

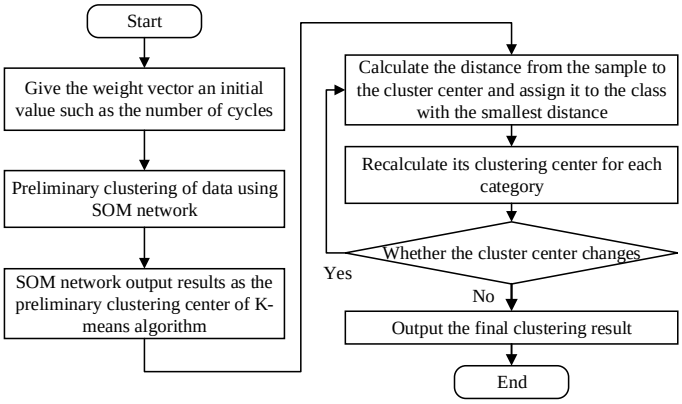

Fig. 3. SOM+K-means clustering algorithm flow chart

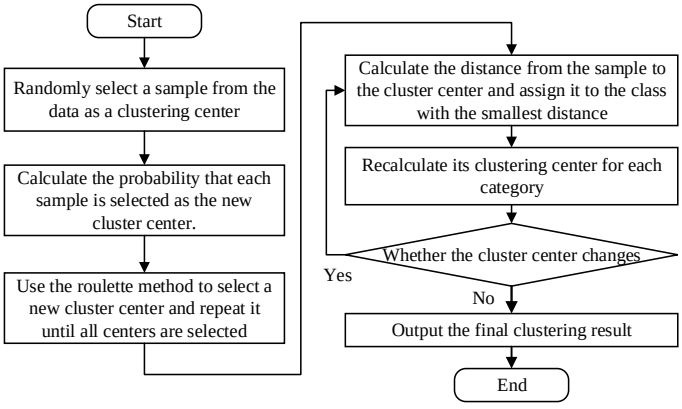

Fig. 4. K-means++ clustering algorithm flow chart

3. CLUSTERING ANALYSIS METHOD OF USER ELECTRICITY BEHAVIOR

There are two kinds of clustering analysis methods used in this paper: direct clustering analysis based on daily load data of users and clustering analysis based on regulation potential index.

3.1 Clustering method based on user daily load data

The clustering method based on user's daily load data mainly includes three steps: data acquisition, data processing and data clustering. Data acquisition requires the selection of sample feature vectors to fully represent the essential characteristics of the sample. In this paper, the power of power user n at 96 uniform sampling points in a day is p_n, and the power data set of group m is P.

$$p_n=[p_{n1},p_{n2},\cdots,p_{n96}] \tag{2}$$

$$P=[p_1,p_2,\ldots,p_n,\ldots,p_m] \tag{3}$$

In order to avoid inaccurate classification when the load level difference is large, a new standard value matrix P is obtained by standardizing the power n at each time point.

$$p_n^0=[\frac{p_{n1}}{p_{n\max}},\frac{p_{n2}}{p_{n\max}},\ldots,\frac{p_{nk}}{p_{n\max}},\ldots,\frac{p_{nm}}{p_{n\max}}] \tag{4}$$

In the formula, p_{nk} is the sampling power of user n at the k th sampling point, and p_{nmax} is the maximum real-time power of user n at 96 sampling points per day.

After data acquisition and pre-processing, it is necessary to apply clustering algorithm to cluster the sampled data.

3.2 Clustering method based on regulating potential index

The consumer's electricity consumption behavior is mainly affected by the production and living demand, and also by electricity price. Feedback will be generated only after the user's incentive reaches a certain threshold, and then will increase with the increase of the incentive degree until a saturation value is reached. The fitting curve considering load transfer rate can be expressed as follows:

$$p_{kt_2}^{\tau}=p_{kt_2}+\sum_{t_i\neq t_2}\lambda_{kt_2t_i}p_{kavt_i} \tag{5}$$

In the formula: $p_{kt_2}^{\tau}$ is the normalized fitting power of the k th user after considering the load transfer rate at t_2 time, $\lambda_{kt_2t_i}$ is the load transfer rate from t_2 time to t_i time, and p_{kavt_i} is the normalized average load at t_i time.

It is concluded that the clustering index θ_{kt_i} of user k considering user load transfer rate in t_i time is as follows:

$$\theta_{kt_i}=p_{kt_i}^{\tau}-p_{kt_i}=\sum_{t\neq t_i}\lambda_{kt_it}p_{kavt} \tag{6}$$

A new regulation potential index is obtained by using peak-valley time-of-use tariff, and the similarity of the index is used as the clustering evaluation criterion.

$$\theta_{k,t_i}=\begin{cases}\lambda_{\mathrm{fg}}\varphi_{\mathrm{f}}^{*}+\lambda_{\mathrm{pg}}\varphi_{\mathrm{p}}^{*},t_i\in t_{\mathrm{g}}\\ \lambda_{\mathrm{fp}}\varphi_{\mathrm{f}}^{*}-\lambda_{\mathrm{pg}}\varphi_{\mathrm{g}}^{*},t_i\in t_{\mathrm{p}}\\ -\lambda_{\mathrm{fg}}\varphi_{\mathrm{g}}^{*}-\lambda_{\mathrm{fp}}\varphi_{\mathrm{p}}^{*},t_i\in t_{\mathrm{f}}\end{cases} \tag{7}$$

In the formula: λ_{fg}, λ_{pg} and λ_{fp} are load transfer rates from peak time to valley time, from normal time to valley time and from peak time to normal time respectively. φ_f^*, φ_g^* and φ_p^* are peak, valley and hourly load rates respectively.

4. EXAMPLE ANALYSIS

The information of electricity price in this paper is based on the data of peak-valley time-of-use price and load transfer rate in reference [6].K-means algorithm, SOM+K-means algorithm and K-means++ algorithm are used to analyze and compare based on load data and adjustment potential index. The data is based on the daily load of users in a jurisdiction of a power company.

4.1 Clustering analysis of user's electricity use behavior using K-means algorithm

Based on the load data, firstly, the K-means algorithm is used to cluster the user's electricity consumption behavior data. Set the initial clustering number $N=3$, where the clustering results based on load data and regulation potential index are the same, and the clustering results are shown in Figure 5.

It can be seen the users classified into three categories, the first category is a typical bimodal user (user 6-8), which is characterized by a large gap between the two peaks, the peaks are concentrated at 12:30 and 20:30. The second category includes 7 users, the vast majority of which (users 1-3, users 10-11) are typical unimodal users. The characteristics of such users are that the peak starts from 06:30 to 09:00 and ends from 17:00 to 20:30. However, there are users 4 and 5 with different curve shapes in the second clustering. These two user curves have the characteristics of bimodal users, but compared with the first bimodal users, there is a small gap between the two wave peaks, which are concentrated at 9:30-13:00 and 17:30-20:00. The third category contains one user (9), which is also a typical first bimodal user, and has the same characteristics as described in the first category.

4.2 Clustering analysis of user's electricity use behavior using SOM+K-means algorithm

Based on load data, the self-organizing center K-means algorithm is used to cluster and analyze the user's electricity consumption behavior data. The clustering results based on load data and regulation potential index are the same. The clustering results is shown in Figure 6. It can be seen that all the users are divided into three categories. The first category accurately contains all the typical bimodal users of the first category (users 6-9). Compared with the K-means algorithm, the user 9 is accurately identified and classified. The second category includes five users, all of which (users 1-3, users 10-11) are typical unimodal users. There are two users in the third category. For the second type of bimodal user (user 4, user 5) whose clustering results based on K-means algorithm fail to recognize and classify successfully, there is a small difference between two values, focusing on the peaks of 9:30-13:00 and 17:30-20:00.

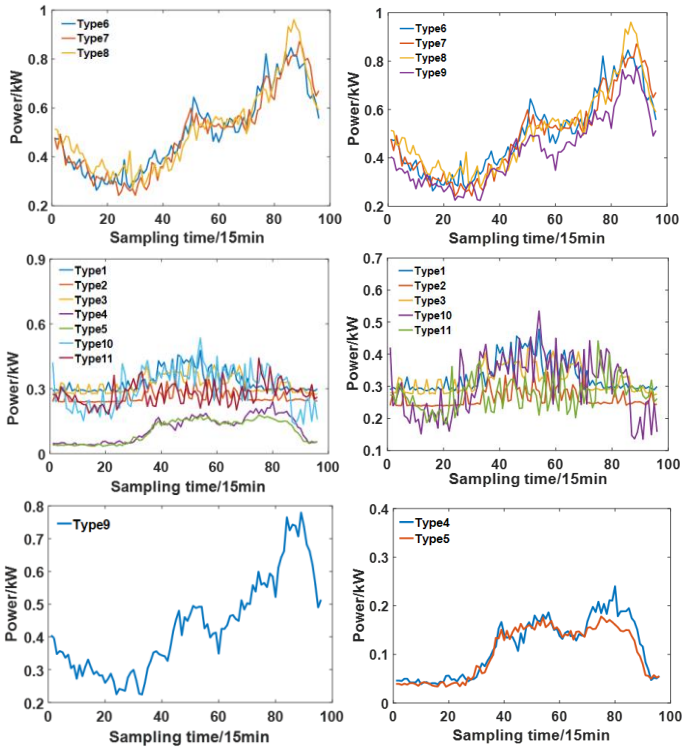

Fig. 5. Clustering curve of load data based on K-means

Fig. 6. Clustering curve of load data based on SOM+K-means or K-means++

4.3 Clustering analysis of user's electricity use behavior using k-means++ algorithm

Based on the load data, K-means++ algorithm is used to cluster and analyze the user's electricity consumption behavior data. The initial number of clusters is $N=3$. The clustering results based on load data and regulation potential index are the same. K-means++ algorithm obtains the same clustering results as SOM+K-means algorithm. The clustering results is shown in Figure 6. Compared with K-means algorithm, it can accurately identify and distinguish user types. The specific analysis is the same as section 4.2.

4.4 Contrastive analysis of clustering results of algorithms

From the clustering results, it can be seen that for direct clustering based on load and clustering based on regulation potential index, the two clustering methods have obtained consistent results in this example. However, from the principle of the two methods, the clustering method based on adjustment potential index can be applied to larger sample clustering analysis, and the amount of data needed to be acquired is smaller and the efficiency is higher when dealing with large-scale data.

The clustering results using K-means algorithm can ensure that similar users' electricity consumption behavior has basic consistency in the sampling point range, but there are also errors. The reason is that the second type of bimodal user has a smaller peak difference and does not have the same obvious bimodal feature as the first type. In addition, the limitation of K-means algorithm is that the clustering result of K-means algorithm mentioned above is greatly influenced by the initial clustering center. It is necessary to seek a better optimization algorithm to get more accurate clustering result.

SOM+K-means algorithm(S+K) and K-means++ algorithm(K++) solve the problem of inaccurate clustering results of K-means algorithm by optimizing the initial clustering centers. They distinguish each type effectively and screen out the load types mixed with other categories, which have very good properties. The clustering results of the two algorithms are the same, but there are differences in running time. The K-means++ and SOM+K-means algorithms are compared in 7 runs, as shown in Table 1.

Table 1. Two algorithms clustering time

	1	2	3	4	5	6	7	mean
K++/s	0.66	0.68	0.68	0.68	0.67	0.64	0.65	0.66
S+K/s	1.06	1.12	1.09	1.11	1.09	1.07	1.09	1.09

Thus, in direct clustering based on load data and adjustment potential index clustering, K-means++ algorithm significantly shortens the clustering time compared with SOM+K-means algorithm. The reason is that the introduction of artificial neural network into SOM algorithm increases the cost of learning time, while K-means++ algorithm only chooses the cluster centers of classical K-means algorithm. It divides randomly selected centers artificially as far as possible without complicated learning and computing process, so it has more advantages in clustering speed.

5. CONCLUSIONS

In this paper, SOM+K-means algorithm and K-means++ algorithm are used to analyze the daily load data of users in a power company's jurisdiction area based on user load data and regulation potential indicators. The two algorithms solve the problem of inaccurate clustering caused by the selection of initial clustering centers in the classical K-means algorithm, and effectively realize the recognition of user categories and the distinction of similar categories. The two algorithms achieve the preliminary optimization of clustering centers, which has a significant effect on improving the accuracy of clustering results. At the same time, the K-means++ algorithm separates the clustering centers artificially with a simple process, which significantly shortens the clustering time compared with SOM+K-means algorithm which needs training of neural network.

ACKNOWLEDGEMENT

This project was supported by National Key R&D Program of China (No. 2018YFB0905000); This project was supported by Science and Technology Project of SGCC(SGTJDK00DWJS1800232). This study was conducted in cooperation of APPLIED ENERGY UNiLAB-DEM: Distributed Energy & Microgrid. UNiLAB is an international virtual lab of collective intelligence in Applied Energy.

REFERENCE

[1] BINH P T T, TUONG L D. Clustering the behaviour of electricity consumption[C]// IPEC, 2012 Conference on Power & Energy. IEEE,2013: 402-406.

[2] Jun L U, Yanping Z, Wenhao P, et al. Interactive Demand Response Method of Smart Community Considering Clustering of Electricity Consumption Behavior[J]. Automation of Electric Power Systems, 2017, 41(17):113-120.

[3] Chen Y, Hao W U, Shi J, et al. Application of Singular Value Decomposition Algorithm to Dimension-reduced Clustering Analysis of Daily Load Profiles[J]. Automation of Electric Power Systems, 2018.

[4] AL-OTAIBI R, JIN N, WILCOX T, et al. Feature construction and calibration for clustering daily load curves from smart-meter data[J]. IEEE Transactions on Industrial Informatics, 2017, 12(2): 645-654.

[5] Zhu W, Wang Y, Luo M, et al. Distributed Clustering Algorithm for Awareness of Electricity Consumption Characteristics of Massive Consumers[J]. Automation of Electric Power Systems, 2016.

[6] REN Bingli, ZHANG Zhengao, WANG Xuejun, et al. Assessment method of demand response peak shaving potential based on metered load data[J]. Electric Power Construction, 2016, 37(11): 64-70.