



Research on Residential Power Consumption Behavior Based on Typical Load Pattern

Anmeng Mao^(✉), Jia Qiao, and Yong Zhang^{ID}

School of Electrical Engineering, University of Jinan, Jinan 250022, Shandong, China
cse_zhangy@ujn.edu.cn

Abstract. According to the current analysis of residents' electricity consumption behavior, with the popularization of smart meters, to a certain extent, residents' electricity consumption data can be collected more efficiently and accurately to ensure the accuracy of subsequent electricity consumption behavior analysis. Based on the traditional fuzzy C-means clustering, clustering analysis can be performed on residential electricity consumption behavior. However, due to the large volume of data, more noise points will be generated in traditional clustering analysis, which will affect the clustering results. When studying the electricity consumption behavior of residents, based on a large amount of electricity consumption data, traditional clustering analysis will generate more noise points, which will affect the clustering results. In the study of electricity consumption behavior, the artificial neural network is introduced in the data preprocessing to classify the data. It can be found that the fuzzy C-means clustering combined with the neural network can effectively eliminate the noise points and have a good clustering effect.

Keyword: Data mining · BP neural network · Fuzzy C-means clustering

1 Introduction

Non-intrusive load monitoring analyzes the operating status of various electrical equipment in the tested households through data monitoring of the electrical load port signals of the household users. The correct monitoring of the electrical load of the household users has an impact on the health of the users home. It is of great significance to understand and effectively control energy consumption [1, 2]. The non-intrusive load monitoring power control system is a diversified management data system, which combines communication and computer technology, and incorporates a large amount of data processing and artificial intelligence technology. Using it not only improves the utilization rate of energy consumption, but also ensures the healthy life of household users by monitoring electricity consumption [1].

From the perspective of residential users, through the analysis of electricity consumption data [3], it is possible to monitor whether the electricity consumption behavior of household users is safe, and whether the equipment in the unit family is aging or

damaged, and through the signal data monitoring and analysis of various equipment of household users, it is found Abnormal signals can be processed in time [4, 5]. By observing the electricity consumption behavior status of the monitored households [6, 7], it is possible to observe and record the behavior patterns of the users in the households for long-term monitoring. Once it finds that the user’s life status is abnormal, it can promptly feedback to the person in charge of the user’s life safety status, thereby ensuring the user’s life safety [8, 9].

In terms of power supply demand, through the analysis of user power consumption behavior data, the power system can be more reasonably divided into time periods and power consumption to set electricity prices, and improve the utilization of power system assets [10]. Moreover, peak periods of electricity consumption often occur in different seasons in our country [11]. Through the use of power equipment, the peak period can be consciously avoided and the utilization rate of the power grid can be improved [12, 13].

2 Application of Traditional Fuzzy C-Means Clustering in the Study of Electricity Consumption Section

2.1 Data Collection

In the model analysis, the UK_DALE data set used is an open-access data set from the United Kingdom, which records the household appliance-level electricity consumption of the entire family, with a sampling rate of 16 kHz and a single device at 1/6 Hz. This is the first open-access UK data set at this time resolution. The data set records five houses. The active power consumed by a single device and the apparent power demand of the entire house are recorded every six seconds. After completing the data collection function, the data in this study were mainly concentrated in 48 power load points, and 10 different time points were collected for different load points, and a total of 480 power consumption data were analyzed.

According to the analysis of different power load points at different times, we can find that the data shows a more obvious trend. In Fig. 1, the abscissa represents the user load point, the ordinate represents the power consumption, and the different curves represent the different moments. Electricity consumption can be found to have the same trend over time, so it is universal to analyze and process data.

2.2 Fuzzy C-Means Clustering

Fuzzy C-means clustering divides n vectors ($i = 1, 2, \dots, n$) into c fuzzy groups, and finds the clustering center of each group to minimize the value function of the dissimilarity index. The main difference between FCM and HCM is that FCM uses fuzzy division, so that each given data point uses a degree of membership between 0 and 1 to determine the degree to which it belongs to each group.

In FCM the value function (is objective function) is shown in Eq. 1:

$$J(U, C_1, C_2 \dots C_C) = \sum_{i=1}^C J_i = \sum_{i=1}^C \sum_j^n u_{ij}^m d_{ij}^m \tag{1}$$

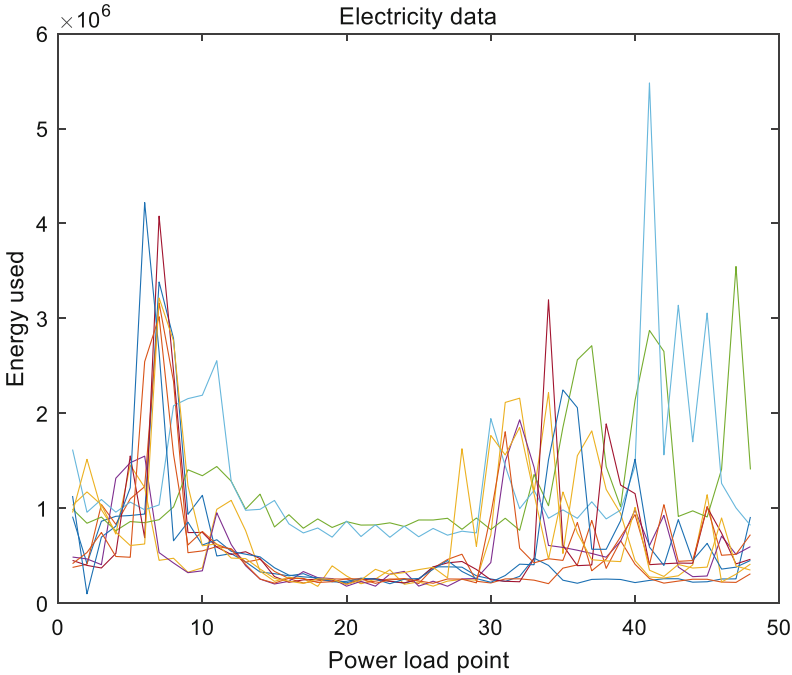


Fig. 1. Electricity data

Here u_{ij} is between 0 and 1, c_i is the cluster center of fuzzy group I , $d_{ij} = \|c_i - x_j\|$ is the Euclidean distance between the I -th cluster center and the J -th data point; and is a weighted index. Calculation of Center c_i :

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \tag{2}$$

Membership calculation:

$$u_{ij} = \frac{1}{\sum_{k=1}^m \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{3}$$

Perform cluster analysis on the collected electricity consumption data. At the same time, in the case of less electricity consumption data, at present, for analysis, too many classifications will lead to excessive calculations, and classification problems cannot be effectively solved when there are fewer classifications. Divide the data results into 3 categories. In the case of avoiding incomplete classification, effectively analyze the user's electricity consumption behavior. Based on the above analysis and data, the residents'

electricity consumption data are awakened by fuzzy C-means clustering. The specific clustering results are as follows (Fig. 2):

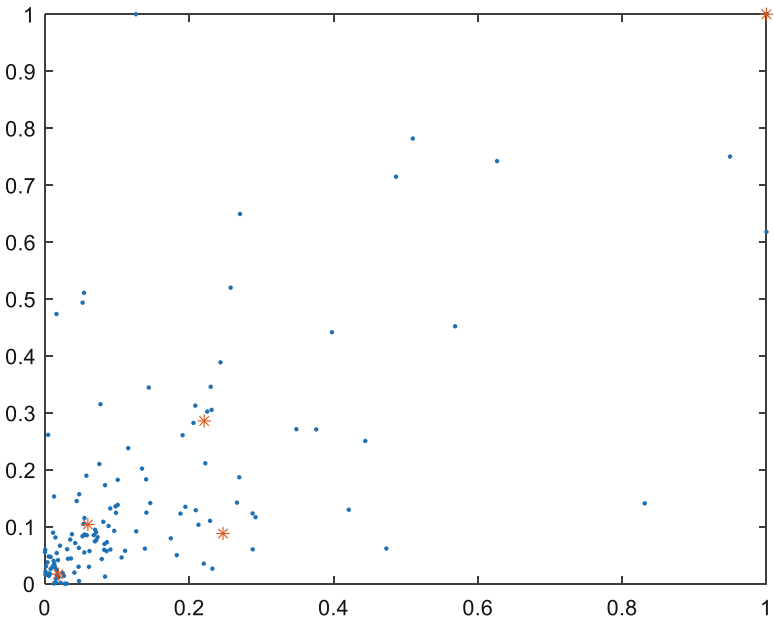


Fig. 2. Electricity data

From the analysis of the clustering results, it can be found that in traditional clustering, there are more noise points and the classification and clustering results are not obvious in combination with the current electricity nodes and loads.

Therefore, artificial neural network is introduced to pro-process the data, and the traditional fuzzy C-means clustering is optimized by combining artificial neural network (BP neural network).

3 Application of Improved Fuzzy C-Means Clustering in the Study of Electricity Consumption

3.1 Data Preprocessing Base on Artificial Neural Network

In the traditional fuzzy C-means clustering, there will be more noise points. that is, for some electricity consumption data, Therefore, data preprocessing is required in this part, combined with artificial neural network (BP neural network) to classify and preprocess the data.

At the first time, from the point of view of node analysis, compare with different collection time points (Fig. 3):

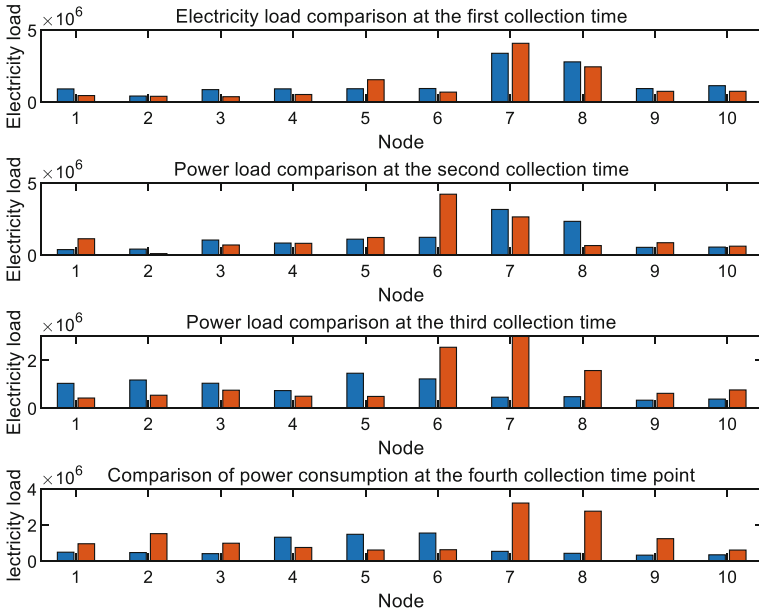


Fig. 3. Comparison of acquisition time

From the above comparison, some node data has a large gap. At the same time, for the node data, due to the particularity of the electricity consumption data, it can be found that the electricity consumption data is relatively large, which will be affected to a certain extent in the subsequent calculation and analysis. Affect the subsequent calculation efficiency and calculation accuracy, so in this step, the power load data is normalized (Fig. 4):

Base on the current data, we will construct a sample database and use the current 480 electricity consumption data for statistics. In order to better calculate the difference between the samples, use the working day data as a and the rest day data as b, and use the following formula to calculate the difference value as an evaluation indicator. Use the formula as follows:

$$u = \frac{|a - b|}{a + b} \tag{4}$$

The calculation result table for 48 nodes through the above formula is as follows: (the previous 10 nodes as an example) (Table 1).

After performing the same processing on 48 nodes, it can be found that the value distribution law of the average index U is as follows (Fig. 5):

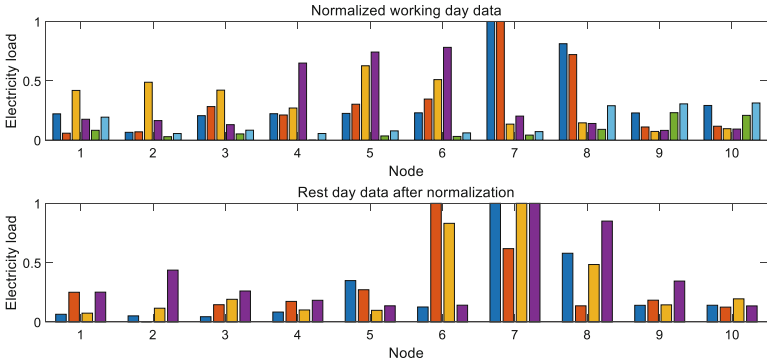


Fig. 4. Time comparison after normalization

Table 1. Data difference value.

Nodes	1st	2nd	3rd	4th	Average
1	0.5485	0.6239	0.7001	0.1771	0.5124
2	0.1235	1	0.6168	0.4534	0.5484
3	0.6483	0.3211	0.3766	0.3352	0.4203
4	0.4549	0.1008	0.4570	0.5614	0.3934
5	0.2155	0.0540	0.7302	0.6907	0.4226
6	0.2918	0.4862	0.2397	0.6943	0.4280
7	1	0.2362	0.7630	0.6638	0.4157
8	0.1678	0.6832	0.5384	0.7173	0.5267
9	0.2403	0.2481	0.3276	0.6166	0.3582
10	0.3498	0.0225	0.3395	0.1853	0.2270

The point data with a small difference between the rest days are subjected to a unified clustering operation, and the nodes greater than 0.5 are clustered separately. Through the above analysis, the sample library is constructed and the BP neural network is trained.

3.2 Improve Fuzzy C-Means Clustering

Perform C-means clustering on the preprocessed data. This part of the data can be found to be divided into three categories, namely: Electricity data is not greatly affected by time, and it is recorded as a; Electricity consumption data is greatly affected by time, and data on working days is recorded as b; The electricity consumption data is greatly affected by time, and the data on the rest day is recorded as c; The data is re-clustered, and the clustering results are as follows (Fig. 6):

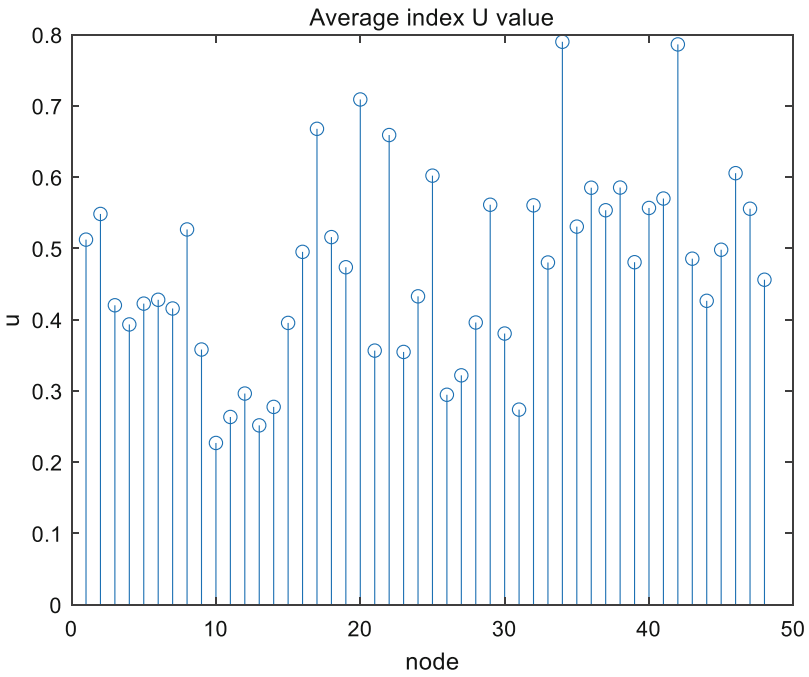


Fig. 5. Average index U

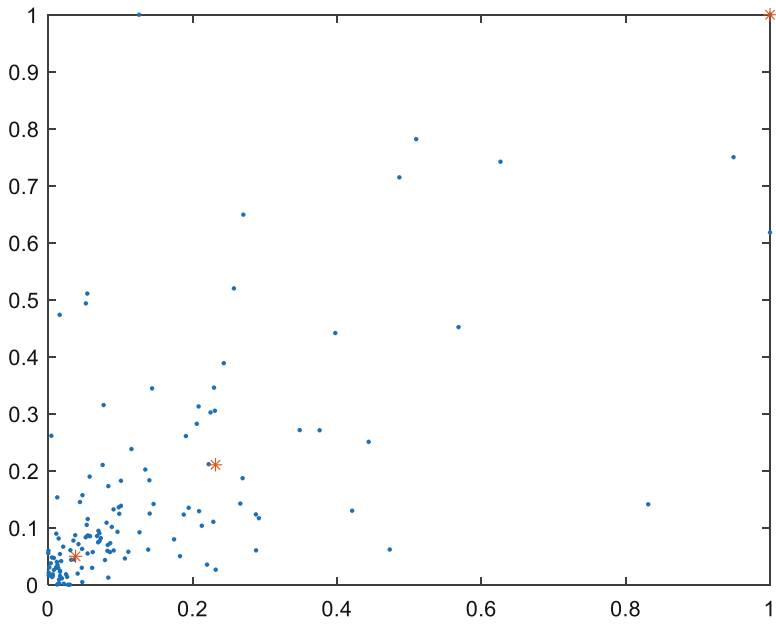


Fig. 6. Clustering results

From the above clustering results, there is a clear gap between the three types of results. At the same time, for the traditional fuzzy C-means clustering, the data clustering effect processed by the BP neural network is better.

4 Conclusion

In the traditional analysis of user electricity consumption behavior, due to the current large amount of data, to a large extent, the increase in the amount of data will greatly increase the noise points in the clustering results, and the large amount of data will the cumulative effect of errors in the clustering center of the clustering results appears everywhere. Therefore, the BP neural network is introduced to analyze the actual situation of data collection and the current electricity consumption, and the data is processed and classified by the BP neural network. Clustering analysis of the data processed by the neural network can find that the noise points and errors in the clustering are largely reduced, which has obvious improvement effects.

Acknowledgment. Funding from the key research and development project of Shandong province (2012CX30302) is gratefully acknowledged.

References

1. Pan, S., et al.: Cluster analysis for occupant-behavior based electricity load patterns in buildings: a case study in Shanghai residences. *Build. Simul.* **10**(6), 889–898 (2017)
2. Valderrama, J.F.B., Valderrama, D.J.L.B.: Two cluster validity indices for the LAMDA clustering method. *Appl. Soft Comput. J.* **89**, 106102 (2020)
3. Chen, Q., Ma, Y.M.: The research on cloud platform considered privacy household load data processing. *Adv. Mater. Res.* **1049–1050**, 1929–1933 (2014)
4. Cheng, Q., Min, C., Ciwei, G., Huixing, L., Tugang, S.: Research on the analysis of user's electricity behavior and the application of demand response based on global energy inter-connection. In: 2016 China International Conference on Electricity Distribution (CICED) (2016)
5. Jackson, D.B.: System and method for managing energy consumption in a compute environment (2012)
6. Ke, X., Yufeng, X., Wenbin, N., Ting, L.: Analysis of electricity use behavior with clustering method and classification of peak and valley periods (2019)
7. Ozawa, A., Furusato, R., Yoshida, Y.: Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. *Energy Build.* **119**, 200–210 (2016)
8. Panapakidis, I.P., Alexiadis, M.C., Papagiannis, G.K.: Deriving the optimal number of clusters in the electricity consumer segmentation procedure. In: *European Energy Market* (2013)
9. Perez, J., Velasquez, J.D., Franco, C.J.: Characterization of the hourly load curve in the colombian electricity market. *IEEE Lat. Am. Trans.* **13**(12), 3826–3831 (2015)
10. Song, C., Wang, C., Ahuja, N., Zhou, X., Daniel, A.: Optimize datacenter management with multi-tier thermal-intelligent workload placement. In: *Thermal Measurement, Modeling and Management Symposium* (2015)

11. Tajeuna, E.G., Bouguessa, M., Wang, S.: A network-based approach to enhance electricity load forecasting. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (2018)
12. Trotta, G., Gram-Hanssen, K., Jrgensen, P.L.: Heterogeneity of electricity consumption patterns in vulnerable households. *Energies* 13 (2020)
13. Yang, Z., Lin, X., Jiang, W., Li, G.: An electricity data cluster analysis method based on saga-fcm algorithm. In: IEEE International Conference on Networking (2017)