Power Consumption Behavior Analysis Method Based on Improved Clustering Algorithm of Big Data Technology

Zheng Zhu[1(✉)], Haibin Chen[1], Shuang Xiao[1], Jingrui Yan[2], and Lei Wu[2]

[1] Electric Power Research Institute, State Grid Shanghai Municipal Electric Power Company, Shanghai 200120, China
cshouq20220606@163.com

[2] LongShine Technology Group Co., Ltd., Shanghai 201600, China

Abstract. With the increase of the types and quantities of electrical equipment, the magnitude of user power consumption data has increased exponentially. Deep mining and analysis of it is the key to help the power grid understand customer needs. Therefore, the research on user power consumption analysis method based on improved clustering algorithm based on big data technology is proposed. Analyze the influencing factors of users' electricity use behavior (economic factors, time factors, climate factors and other factors), on this basis, explore the characteristics of users' electricity load, reduce and process users' electricity data samples based on PCA algorithm, improve clustering algorithm by using big data technology – extreme learning machine principle, cluster and process users' electricity use behavior, use GSP algorithm to mine sequential patterns of users' electricity use behavior, and obtain users' electricity use rule model, Thus, the analysis of users' electricity consumption behavior is realized. The experimental data show that the minimum value of MIA index obtained by applying the proposed method is 0.09, and the maximum accuracy of the user power consumption law model is 92%, which fully confirms that the proposed method has better application performance.

Keywords: Big Data Technology · Power Consumption Data · Power Consumption Behavior of Users · Improvement of Clustering Algorithm · Behavior Analysis

1 Introduction

There are various types of power big data, and the power consumption data of power users is one of the fastest growing data sources [1]. With the gradual elimination of traditional mechanical meters and the popularization and application of intelligent terminals such as smart meters, power grid enterprises have mastered a large amount of user electricity data. For a long time, users' electricity consumption data has been ignored and only used for the measurement of electricity bills [2]. With the gradual maturity of data

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 534, pp. 318–331, 2024.
https://doi.org/10.1007/978-3-031-50577-5_23

mining technology, researchers in power grid enterprises and universities have begun to pay attention to the potential value of user electricity data. The analysis of user's electricity consumption behavior based on user's daily load curve has gradually become a popular application scenario of power big data mining, so the research of user's electricity consumption behavior analysis method based on big data technology and improved clustering algorithm is proposed. Based on the analysis of the influencing factors of user's electricity consumption behavior, the user's electricity consumption data samples are reduced based on PCA algorithm. Based on big data technology, the improved clustering algorithm is used to mine the sequence pattern of users' electricity consumption behavior, obtain the laws of users' electricity consumption, and realize the analysis of users' electricity consumption behavior.

2 Research on Analysis Method of User's Electricity Consumption Behavior

2.1 Analysis on Influencing Factors of Users' Electricity Consumption Behavior

In order to improve the accuracy of the analysis of users' electricity consumption behavior, the first step is to analyze the influencing factors of users' electricity consumption behavior, so as to lay a solid foundation for the subsequent research on users' electricity load characteristics.

The electric load of power users is composed of industrial load and residential load, and the characteristics of electric load of users are affected by many factors. The analysis of the influencing factors of the user's electricity use behavior plays an important role in the analysis of the user's electricity use behavior. The degree of influence of each factor on the user's behavior varies. The main factors include economic factors, time factors, climate factors, and other factors, as shown below:

(1) Economic factors

Generally speaking, economic development determines the growth rate of power load. The improvement of economic level is accompanied by the increase of electricity consumption. This is because electricity is an indispensable energy for social development today. The economic output value of most industrial users is in direct proportion to the use of power energy; For residential users, with the improvement of economic development level, household appliances have been popularized, so the power consumption of ordinary residential users has also shown a year-on-year growth trend with the number of local economic development;

(2) Time factor

The influence of time factor on the electricity consumption behavior of users has a certain regularity. Here, the time factor is divided into holiday factor and weekend factor. Holiday factors refer to the impact of national statutory holidays such as Spring Festival, May Day, National Day and Dragon Boat Festival on users' electricity use behavior. During holidays, especially during the "Spring Festival" and other major festivals in China, a certain amount of industrial load is in shutdown or only maintained at a low constant level. No industrial operation with high energy consumption is carried out, but only the online operation of equipment is

maintained. Therefore, the industrial load decreases significantly during holidays [3]. For residential users and commercial users, the peak of electricity consumption in the evening generally rises during holidays, and there is no obvious difference between peak and valley, and their electricity load also changes greatly. Weekend factor refers to the difference between the user's electricity use behavior on weekdays and weekdays within a week. Because users work in "week" as the time unit in most of the time, this effect is cyclical, and because of the different electricity demand of different types of users, their electricity use behavior on weekdays and weekdays has personalized differences;

(3) Climatic factors

Climate is also one of the main factors that affect the electricity consumption behavior of users. Generally speaking, climate factors include air temperature, precipitation, wind power and humidity, among which the correlation between air temperature and power consumption behavior of users is the most obvious. The impact of climate factors on users' electricity consumption behavior can be shown in many aspects: for example, the rising temperature in summer will force people to cool down by opening air conditioners, fans, etc., and then users' electricity consumption will increase significantly; In winter, when the temperature is low, people use electric heaters and other heating equipment for heating, which will also lead to increased power consumption; When the weather is cloudy or rainy, people will increase the use of lighting equipment and many other aspects. In the current research, the researchers consider the comprehensive impact of various meteorological factors on the user's electricity consumption behavior, so as to more accurately identify the change rule of the user's electricity load.

(4) Other factors

Other factors that affect users' electricity use behavior mainly include the influence of electricity price, demand side management measures, power supply side, people's income level and change of consumption concept. This section briefly introduces the influence of electricity price and demand side management measures on users' electricity use behavior.

The impact of electricity price on users' electricity consumption behavior is mainly shown in the following two aspects:

(1) The level of electricity price

The level of electricity price is closely related to social economic development level, price level and other factors. Because different users have different affordability to electricity price, the adjustment of electricity price level has different influence on users. However, generally speaking, the rise in electricity prices will lead to a downward trend in electricity consumption of power users, but the overall trend of electricity consumption is also closely related to other factors such as the level of economic development;

(2) Electricity price structure

The electricity price structure also has a significant impact on users' electricity consumption behavior. For example, the implementation of peak and valley electricity prices means that power companies determine the peak and valley periods of electricity consumption based on the average electricity load of users in the region,

and then implement different electricity price mechanisms in the peak and valley of electricity consumption, increase the peak electricity price, reduce the low valley electricity price, and then encourage users to use electricity in the valley, so as to achieve the purpose of cutting the peak and filling the valley, balance the contradiction between power supply and demand, and give full play to the leverage role of price, Optimize the allocation of power resources [4].

The development of reasonable DSM measures will effectively improve the characteristics of power load on the user side, effectively avoid peak power consumption and ease the tension of insufficient power supply. For example, implement peak and valley electricity price load management measures to encourage users to use electricity in the low valley, avoid the concentration of users' electricity demand in a fixed period of time, and achieve peak shaving and valley filling. For another example, by formulating a reasonable and orderly power consumption strategy and according to the characteristics of users' power consumption behavior, reducing or limiting users' power load in different periods of time can effectively reduce the peak valley difference of the grid and make the grid load curve tend to balance.

The above process has completed the analysis of the influencing factors of users' electricity consumption behavior, mainly including economic factors, time factors, climate factors and other factors, providing basic support for the subsequent exploration of users' electricity load characteristics.

2.2 Research on Characteristics of User Power Load

Based on the above analysis results of influencing factors of users' electricity use behavior, explore the characteristics of users' electricity load, and reduce the data samples of users' electricity use, so as to facilitate the subsequent clustering operation of users' electricity use behavior [5].

The power users are divided into three categories: industrial, commercial and residential, and their typical daily load curves are obtained to analyze the characteristics of power load of users, as shown below:

Industry is the largest power consumption industry at present, mainly including electricity for high energy consumption industries such as steel, chemical industry, nonferrous metals, cement, and light industries such as textile, paper, and non-staple food processing. There are many kinds of industrial industries, and different industries have different power consumption characteristics, which are very different. Therefore, we can cluster the industrial load curve to determine different industrial industries, judge the prosperity and decline of some industries, and provide targeted energy efficiency services, power saving consulting, reliable power supply, information notification, high-quality energy use, high-quality services, customized services, access services, equipment leasing, and power supply channel initiatives. The daily load curve characteristics of typical industrial users are shown in Fig. 1.

The commercial load mainly occurs in large commercial buildings and senior office buildings, and is mainly composed of electrical loads in department stores, entertainment venues, supermarkets, food plazas and senior office buildings. It can be roughly divided into three parts: retail, entertainment and office. Commercial power load mainly includes computer equipment, daily lighting, central air conditioner, window air conditioner, fan

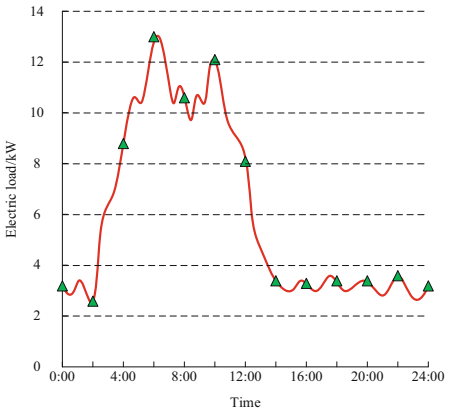

Fig. 1. Characteristic Diagram of Typical Industrial User's Daily Load Curve

and elevator. The load characteristics of commercial and office buildings are: 1. The load of commercial and office buildings shows obvious timeliness and seasonality. The load curve of the office building is relatively stable in a day, with little fluctuation, and rises sharply at night; Due to the composition of the business system and different operating conditions, the business load curve will also be slightly different [6]. Therefore, according to different types of load curves, we can judge the rise and fall of certain industries and the geographical location of business districts. The daily load curve characteristics of typical commercial users are shown in Fig. 2.

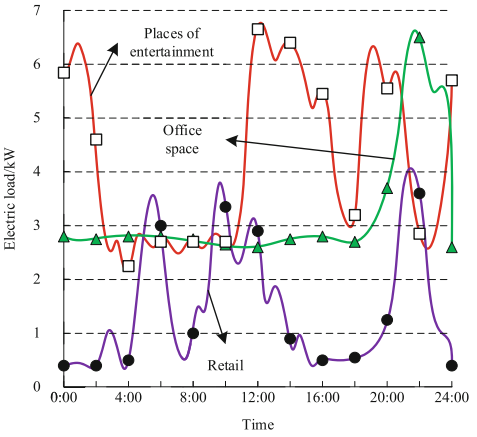

Fig. 2. Characteristic Diagram of Daily Load Curve of Typical Commercial Users

The residential load mainly comes from the use of ordinary appliances such as lighting, washing machines and refrigerators, as well as high energy consumption appliances such as air conditioners, electric heaters, water heaters, cooking appliances and electric water heaters. In ordinary electrical appliances, lighting power consumption has changed

the most in a day, but there are few time differences. Lighting load often occurs at the same time of the day. According to the power utilization habits of residential users, lighting mainly occurs at night, and the time period is usually from 18:00 to 22:00, also known as "lighting peak"; Refrigerators are household appliances that have been used for a long time, and their service time is relatively less adjustable; The washing machine is usually used in the morning or evening of weekdays, or on Saturdays and holidays; Generally, TV sets are used most frequently in the evening, followed by morning and noon. Among high energy consumption household appliances, electric water heaters and air conditioners are mostly used at night, while cooking appliances such as rice cookers and induction cookers are used at three meals a day. But because each family has different appliances, which are scheduled by different people and have different use preferences, each family will have different daily load curves. The daily electricity load curve of residents will reflect different electricity consumption behaviors. According to these behaviors, different groups can be clustered to analyze different types of users' electricity consumption behaviors. The daily load curve characteristics of typical residential users are shown in Fig. 3.

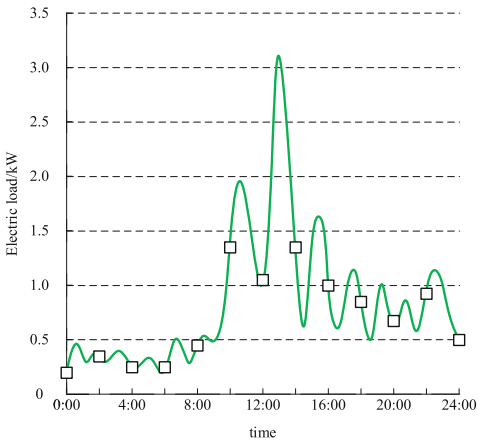

Fig. 3. Characteristic Diagram of Typical Residential User's Daily Load Curve

Based on the above exploration results of user power load characteristics, user power consumption data samples are obtained. The data volume is huge, and there are many redundant data samples, which has brought some obstacles to the subsequent clustering of user power consumption behavior. Therefore, PCA algorithm is applied in this study to reduce the user power consumption data samples.

PCA algorithm is a multivariate statistical technique commonly used to reduce data dimensions. It can not only easily process a large number of data, but also avoid a large number of complex calculations. Its basic idea is to reconstruct the original variables with certain correlation into a new set of integrated variables that are not related to each other. PCA dimensionality reduction is to map m samples with n variables that can be described (i.e. $m \times n$ data matrix) to a matrix with m variables that can be described (i.e. $r \leq n$) through orthogonal transformation. Not only should the position subspace composed of

base vectors optimally consider the correlation of data, but also the corresponding base vectors should meet the orthogonality. These generated principal components can reflect most of the information of the original matrix, which is usually expressed as a linear combination of the original variables, which can reduce the user power consumption data samples to the maximum extent and facilitate the clustering of subsequent user power consumption behavior [7].

In this paper, the $m \times n$ dimension user power load data matrix is reduced to $m \times r$ dimension user power load data matrix through PCA algorithm. The specific reduction processing is as follows: Formula (1):

$$\begin{cases} \alpha = \frac{1}{m} * X_t^T * X_t \\ [U, S, V] = svd(H) \\ X_{tr} = U(:, 1:r)^T * X_t \end{cases} \tag{1}$$

In formula (1), α represents the reduction variable of user power load data matrix; X_t represents the user power load data matrix at time t; T represents transposed symbol; U and V represent a simple matrix satisfying $\alpha = USV^T$; S represents a diagonal matrix with non negative diagonal elements, and the diagonal elements are sorted in descending order; $svd(H)$ represents a random constant, which needs to be set according to the actual situation.

The element S_{ij} in diagonal matrix S represents the j-th feature corresponding to the i-th user load data. Therefore, the larger the S_{ij} value, the more information it carries. On the contrary, the smaller the S_{ij} value, the less information it carries. In fact, there are only a few features in the user load data of power system that are extremely important. The minimum r is usually selected to meet the conditions of formula (2):

$$\frac{\sum_{i,j=1}^{r} S_{ij}}{\sum_{i,j=1}^{m} S_{ij}} \times 100\% \geq \beta^o \tag{2}$$

In formula (2), β^o represents the variation percentage of the reduced dimension data. After mapping the user power load data to the low dimensional space, it can be verified that PCA will select the main features that can maximize the variation, which also shows that the characteristics of the user power load data of the entire power grid can be represented by a small amount of information.

The above process has completed the exploration of user power load characteristics, and the reduction of user power data samples has laid a solid foundation for subsequent clustering operations.

2.3 Clustering of Users' Electricity Consumption Behavior

Based on the above simplified user power consumption data samples, the clustering algorithm is improved by using the big data technology – extreme learning machine principle. Based on this, the user power consumption behavior is clustered to make

sufficient preparation for the realization of the end user power consumption behavior analysis.

By comparing and analyzing the current popular clustering algorithms, we understand the main problems and defects of each clustering algorithm. Next, we are going to adopt the idea of spectral clustering, and quickly embed the original data through the advantages of fast learning speed of the extreme learning machine, and then cluster them with K-means algorithm [8]. Limit learning machine is a simple, easy to use and effective learning algorithm for solving single hidden layer feedforward neural network SLFNs, and it is also a key component of big data technology. In the traditional neural network learning algorithm, a large number of network training parameters need to be set manually, and it is too easy to cause local optimal solution. However, the limit learning machine only needs to set the number of hidden layer nodes of the network, does not need to adjust the bias and input weight of the hidden element of the network during the algorithm implementation, and will only produce the unique optimal solution, so it has the advantages of good generalization performance and fast learning speed. As a single-layer feedforward neural network, ELM randomly generates hidden layer neural bias and input weight, and obtains output weight through analysis and calculation. ELM chooses non differentiable and even discontinuous functions as its activation functions. In the hidden layer, "sigmoid", "sine", "Gaussian" and "hard limiting" functions can be selected as their activation functions, while in the output layer, linear activation functions can be selected as their activation functions.

The mathematical model of limit learning machine classification with L hidden layer nodes and $h(\cdot)$ activation function can be expressed as formula (3):

$$F_o(X_i) = \sum_{j=1}^{L} \chi_j h\left(\omega_j x_i + \delta_j\right) \tag{3}$$

In formula (3), $F_o(X_i)$ represents the mathematical model of the classification of the limit learning machine; χ_j represents the auxiliary parameter of the limit learning machine; ω_j represents the weight coefficient corresponding to the user power consumption data sample; δ_j represents the width of Gaussian hidden layer neurons.

Based on the mathematical model of extreme learning machine classification shown in Formula (3), improve the K-means clustering algorithm concept, as shown in Formula (4):

$$\min \frac{1}{2}\|\chi\|^2 + \frac{C}{2}\sum_{i=1}^{m}\|E_i\|^2 \tag{4}$$

In formula (4), C represents the final category number of the improved K-means clustering algorithm; E_i represents the matrix error value after embedding.

Although power users are generally divided into residential users, commercial users and industrial users, the load characteristics of different types of users may still be very different, and some load characteristics of different users are similar. In this paper, 300 residential users' electricity data are selected from the smart meter data. The original data is 336 dimensions. By supplementing the missing data and eliminating the wrong data, the PCA algorithm is used to preprocess the data dimension reduction, which is

finally reduced to 48 dimensions. Finally, the extreme learning machine algorithm is used to embed the original data. The improved K-means algorithm is used to cluster the original data, which is finally grouped into five categories. The specific process is shown in Fig. 4.

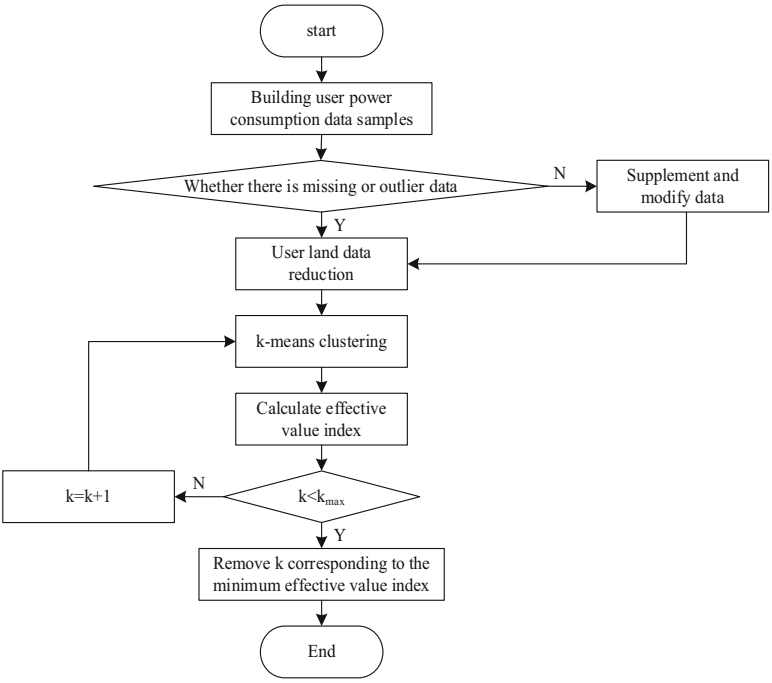

Fig. 4. Clustering flow chart of user power consumption behavior based on improved K-means clustering algorithm

In order to prove the effectiveness of the algorithm, this paper uses correctly defined indicators to evaluate the clustering effectiveness. MIA is the average distance between each data point assigned to the same cluster and its centroid. Defined as Formula (5):

$$MIA = \sqrt{\frac{1}{k}\sum_{i=1}^{k} d(X_k, C_k)} \tag{5}$$

In Formula (5), $d(x_k, C_k)$ represents the distance between data point X_k and its centroid C_k.

Obviously, the lower the value of MIA index is, the better the clustering effect is; On the contrary, the higher the MIA index is, the worse the clustering effect is [9].

The above process completes the improvement of clustering algorithm and the clustering of users' electricity use behavior, and describes the evaluation indicators of clustering results, which provides basis for subsequent analysis experiments of users' electricity use behavior.

2.4 Analysis and Realization of Users' Electricity Consumption Behavior

Based on the above clustering results of users' electricity use behavior, the GSP algorithm is used to mine the sequence patterns of users' electricity use behavior, and based on this, the user electricity use rule model is obtained, thus realizing the analysis of users' electricity use behavior.

GSP algorithm is an evolution and extension of Apriori algorithm, and its running process in computer is similar to Apriori algorithm. The GSP algorithm uses a hash tree to store candidate sequences to improve the speed of the algorithm. Hash tree has root node, internal node and leaf node, and candidate sequence is stored in leaf node. The basic steps of GSP algorithm are as follows:

Step 1: By scanning the clustering result set Y of user power consumption behavior, all items in Y and their support are counted, and frequent 1 sequence pattern candidate set ζ_1 is obtained. Delete the 1 sequence pattern that does not meet the minimum support, and obtain candidate set λ_1 as the initial set of the cycle;
Step 2: Based on the sequential pattern set λ_k, obtain the candidate pattern set ζ_{k+1}, whose length is $k + 1$, and build a hash tree with the depth of $k + 1$ to facilitate the calculation of degrees. Then the Y scans to calculate the support of ζ_{k+1}, find the frequent sequence with $\sup(\vartheta)$ not less than $\min\_\sup$, form the frequent pattern set λ_{k+1}, and delete the non frequent items from the hash tree by pruning step. Finally, the sequential pattern set ζ_{k+2} is obtained through the join operation, and a new hash tree is constructed;
Step 3: Loop the above process until the longest frequent sequence is found or no new frequent sequence is generated, and then return the set λ of all pattern sets λ_k, that is, frequent sequence pattern set.

Among them, join step and pruning step are important processes of GSP algorithm in mining candidate sequence patterns. Connection step: candidate frequent sequences are connected by frequent sequences λ_{k-1} and λ_k. If the subsequence obtained by discarding the first term of Ψ_1 is the same as the subsequence obtained by discarding the last term of Ψ_2, then sequence Ψ_1 is connected with sequence Ψ_2. The candidate sequence generated by connecting sequence Ψ_1 with sequence Ψ_2 is sequence Ψ_1 expanded with the last item in sequence Ψ_2. There are two cases: if the added item is a separate element in Ψ_2, the added item will be added to the last item of Ψ_1 as part of the item; If the added item is a separate element in Ψ_2, it will form a separate element and be appended at the end of Ψ_1 in the merge sequence. It should be noted that when connecting λ_{k-1} and λ_k, when adding Ψ_2 in this paper, it should not only be a part of the item, but also be an independent element.

Based on the frequent sequence pattern of users' electricity use behavior mined by GSP algorithm, it is the habitual electricity use behavior of this kind of users. By combining the symbols of electrical equipment with the frequent sequence, this paper can obtain the electricity use rule model of some users. Obtaining frequent sequences is equivalent to obtaining frequent item sets for such users. This paper can not only analyze the sequence of occurrence between the use of electrical equipment, but also further mine the correlation of the use of electrical equipment based on frequent item sets [10].

Association rules are expressions in the form of $A \rightarrow B$, where A is the leader and B is the successor. In this paper, confidence is used to express the confidence of association

rule $A \rightarrow B$. Support is the percentage of the number m of A and B sequences in the sequence database to the total number n of sequences in the set. Confidence is the ratio of the number of sequences containing A and B, m, to the number of sequences containing A. Strong association rules refer to the association rules that meet the minimum support and confidence. The calculation formula of support and confidence is Formula (6):

$$\begin{cases} \varsigma(A \rightarrow B) = \frac{m}{n} \\ \tau(A \rightarrow B) = \frac{m}{g} \end{cases} \tag{6}$$

In formula (6), d represents the degree of support; $\tau(A \rightarrow B)$ represents the confidence level, and g represents the number of sequences.

The rule with high confidence may not be correct, because the rule may just meet the minimum support, and the phenomenon expressed by the strong association rule may be accidental. To solve this problem, this paper can judge by the lifting value μ to measure whether strong association rules are available. When μ is greater than 1, A is positively correlated with B, the rules are available. When μ is not greater than 1, A and B are not strongly correlated, the rules are not referenced. The strong association rule is shown in Formula (7):

$$\mu(A \Rightarrow B) = \frac{P(AB)}{P(A) \times P(B)} \tag{7}$$

In Formula (7), $P(A)$ and $P(B)$ represent the probability distribution function of frequent item sets and electrical equipment symbols; $P(AB)$ represents the correlation distribution function of $P(A)$ and $P(B)$.

The generated strong association rules can be combined with the simultaneous occurrence sequence in the frequent sequence to obtain the power consumption law model of users. Based on this, the power consumption law model of a specific type of users is analyzed, and the time-sharing price is combined to optimize the energy consumption of users. To sum up, the analysis of users' electricity consumption behavior is realized, which provides assistance for the improvement of distribution system management level, work efficiency and user service level.

3 Experiment and Result Analysis

3.1 Experiment Preparation Stage

In order to verify the application performance of the proposed method, it is necessary to determine the number of user power consumption behavior clusters before the experiment, so as to facilitate the smooth implementation of subsequent experiments.

In essence, the clustering method selected in this study does not need to determine the number of clusters in advance. It can determine the number of clusters according to different needs after the formation of the "tree graph" and segment the tree graph accordingly, but this process is too subjective in determining the number of clusters. In order to make the determination of cluster number more objective, this paper uses three validity indicators to determine the cluster number, and selects three cluster indicators

DB indicator, HS indicator and CH indicator to comprehensively determine the best cluster number. The specific calculation of the three indicators is Formula (8):

$$\begin{cases} \Gamma_{DB} = \frac{1}{Q^*} \sum_{i=1}^{Q^*} \max\left(\frac{d_i - d_j}{d_{ij}}\right) \\ \Gamma_{HS} = \left|\Gamma_{Hom}\left(Q^* - \Gamma_{Sep}(Q^*)\right)\right| \\ \Gamma_{CH} = \frac{tr[Sb(Q^*)]/(Q^*-1)}{tr[Sw(Q^*)]/(Q^*-1)} \end{cases} \tag{8}$$

In Formula (8), Γ_{DB}, Γ_{HS} and Γ_{CH} represent the values of DB indicator, HS indicator and CH indicator; Q^* represents the number of clusters; d_i and d_j represent the average distance from category i, j to the cluster center; d_{ij} is the distance between categories i, j; Γ_{Hom} and Γ_{Sep} represent auxiliary parameters; $tr[\cdot]$ represents the trace of the matrix; $Sb(Q^*)$ represents the inter class dispersion matrix under the cluster number Q^*; $Sw(Q^*)$ represents the intra class deviation matrix with cluster number Q^*.

HS index and CH index take the cluster number corresponding to the maximum point in the curve as the best cluster number, while DB index takes the cluster number corresponding to the minimum point in the curve as the best cluster number.

The above process completes the preparation of the experiment and provides the basis and support for the subsequent experimental results analysis.

3.2 Analysis of Experimental Results

Reference [5] K-means clustering algorithm and reference [8] evolutionary strategy and clustering algorithm are selected as comparison method 1 and comparison method 2. Based on the above experimental preparation, the user's electricity consumption behavior analysis experiment is conducted. In order to intuitively display the application performance of the proposed method, the MIA index and the accuracy of the user's electricity consumption law model are selected as evaluation indicators. The specific experimental results are analyzed as follows:

The MIA index obtained through experiments is shown in Table 1.

As shown in Table 1, the minimum MIA index of the proposed method is 0.09, the minimum MIA index of comparison method 1 is 0.61, and the minimum MIA index of comparison method 2 is 0.51. Compared with the comparison methods 1 and 2, the MIA index value obtained by the proposed method is smaller, which indicates that the proposed method has better clustering effect on users' electricity consumption behavior.

The accuracy of the user's electricity consumption law model is obtained through experiments, as shown in Fig. 5.

As shown in Fig. 5, the maximum accuracy of the proposed method's user power consumption law analysis is 92%, the maximum accuracy of the comparison method 1's user power consumption law analysis is 86%, and the maximum accuracy of the comparison method 2's user power consumption law analysis is 74%. Compared with the comparison methods 1 and 2, the accuracy of the user's electricity consumption law model obtained by the proposed method is higher and has better analysis effect of electricity consumption behavior.

The above experimental data show that, compared with the two comparison methods, the MIA index value obtained by the proposed method is smaller, and the accuracy of

Table 1. MIA Index Table

Experimental group	Propose method	Comparison method 1	Comparison method 2
1	0.56	0.95	0.85
2	0.41	0.88	0.74
3	0.30	0.80	0.96
4	0.10	0.78	0.68
5	0.12	0.84	0.51
6	0.09	0.78	0.75
7	0.41	0.83	0.71
8	0.35	0.69	0.90
9	0.21	0.61	0.83
10	0.27	0.80	0.72

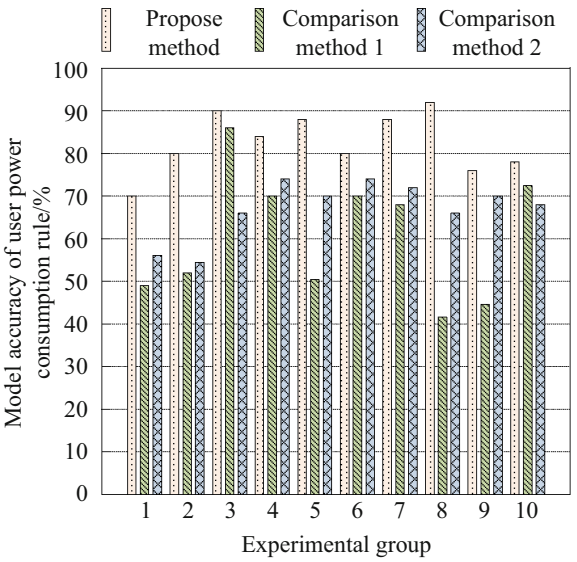

Fig. 5. Schematic Diagram of Model Accuracy of User Power Consumption Law

the user power consumption law model is higher, which fully proves the feasibility and effectiveness of the proposed method.

4 Conclusion

According to the analysis results of influencing factors of users' electricity consumption behavior, this study explored the characteristics of users' electricity consumption load. Based on the simplified user electricity consumption data sample, the big data

technology-limit learning machine principle is used to improve the clustering algorithm, cluster the user electricity consumption behavior, use GSP algorithm to mine the user electricity consumption behavior sequence pattern, and obtain the user electricity consumption law model, thus realizing the analysis of user electricity consumption behavior. The experimental results show that the proposed method greatly reduces the MIA index value, improves the accuracy of the user electricity consumption law model, provides more effective method support for the user electricity consumption behavior analysis, and also provides some reference for related research. The next research will be applied to the analysis of electricity consumption in different fields to analyze and optimize the areas that need to be improved in this research and further expand the application scope of the proposed method.

References

1. Kaddour, S.M., Lehsaini, M.: Electricity consumption data analysis using various outlier detection methods. Int. J. Softw. Sci. Comput. Intell. **13**(3), 12–27 (2021)
2. Cheng, J., He, Ya., Bao, G., et al.: Cluster analysis of consumer electricity load based on CK-means algorithm. Comput. Simul. **38**(7), 63–67, 133 (2021)
3. Guerrero-Prado, J.S., Alfonso-Morales, W., Caicedo-Bravo, E., et al.: The power of big data and data analytics for AMI data: a case study. Sensors **20**(11), 3289 (2020)
4. Qin, X., Li, J., Hu, W., et al.: Machine learning k-means clustering algorithm for interpolative separable density fitting to accelerate hybrid functional calculations with numerical atomic orbitals. J. Phys. Chem. A **124**(48), 10066–10074 (2020)
5. Peng, C.Y., Raihany, U., Kuo, S.W., et al.: Sound detection monitoring tool in CNC milling sounds by K-means clustering algorithm. Sensors **21**(13), 4288 (2021)
6. Wang, J., Huang, S., Wu, D., Lu, N.: Operating a commercial building HVAC load as a virtual battery through airflow control. IEEE Trans. Sustain. Energ. **12**(1), 158–168 (2021)
7. Fan, Y., Liu, Y., Qi, H., et al.: Anti-interference technology of surface acoustic wave sensor based on K-means clustering algorithm. IEEE Sens. J. **21**(7), 8998–9007 (2021)
8. Li, H., Spencer, B.F., Chen, H., et al.: An intelligent algorithm based on evolutionary strategy and clustering algorithm for Lamb wave defect location. Struct. Health Monit. **20**(4), 2088–2109 (2021)
9. Yang, C., Weng, Y., Huang, B., et al.: Development and optimization of CAD system based on big data technology. Comput.-Aided Des. Appl. **19**(S2), 112–123 (2021)
10. Jamsheela, O., Gopalakrishna, R.: Parallelization of frequent itemset mining methods with FP-tree: an experiment with PrePost + Algorithm. Int. Arab J. Inform. Technol. **18**(2), 208–213 (2021)