
Analysis of Power Grid User Behavior Based on Data Mining Algorithms – System Design and Implementation

Yan Wang^{1,*}, Jiawei Xu¹, Xiaowen Chen² and Ying Huang³

¹*Information and Communication Branch of Hainan Power Grid, Haikou, Hainan, China*

²*Hainan Power Exchange Center, Haikou, Hainan, China*

³*Hainan Power Grid Company Limited, Haikou, Hainan, China*

E-mail: 16011010202@stu.suse.edu.cn

**Corresponding Author*

Received 11 December 2023; Accepted 31 January 2024

Abstract

A data mining based power grid user behavior analysis system has been designed to address the issues of insufficient stability and accuracy in existing power grid user behavior analysis systems. Design the overall structure of the power grid user behavior analysis system; In terms of system hardware design, select a core controller, build and install a server as the foundation for system information transmission and logical operation; Based on ZigBee wireless communication technology, a ZigBee wireless communication protocol stack and communication expansion board were designed; In terms of system software design, Python is used to crawl user behavior data in the system data collection layer, and Python language is used to maintain the crawling program; Use the K-means algorithm to perform secondary mining and clustering on power grid user behavior data, obtain the analysis results of power grid user behavior, and transmit them to the system visualization

Distributed Generation & Alternative Energy Journal, Vol. 39_3, 531–558.

doi: 10.13052/dgaej2156-3306.3937

© 2024 River Publishers

display layer. The weight and Rand coefficient of data analysis were used as indicators to test the application effect of the method in this paper. The experimental results showed that the system can stably and accurately analyze the behavior of power grid users, and has good application effect. This research achievement has important reference significance for the research in the field of power grid user behavior analysis in the world scientific community.

Keywords: Data mining, grid user behavior, analysis system, ZigBee, K-means algorithm.

1 Introduction

In order to improve their economic benefits, electric power enterprises have gradually focused on the behavior of power grid users [1–3], and through the behavior of power grid users to develop targeted power product marketing plans [4–6]. In recent years, electric power enterprises have used information technology to optimize production management, energy management and customer service. In the daily work inside the enterprise [7], power enterprises generally apply the planning management system, user management system and load management system. When the user logs in to the above system to interact with the electric power enterprise, the user behavior [8] data is recorded in the database, and the user behavior data in the database increases exponentially with time. However, when electric power enterprises [9] search and modify user information in the data, they can only extract simple information from the big data of power grid users, and cannot obtain the association relationship between the data of power grid user behavior. The above situation has led to data explosion in power enterprises, but knowledge is poor.

In the case of data resource informatization, how to fully develop grid user data and analyze grid user behavior according to the internal relationship of data has become an urgent problem for power enterprises. At present, many scholars have designed relevant systems for obtaining the behavior data of power grid users, such as Kaur R et al. [10] studied the electricity consumption behavior of power users based on data-driven analysis. Based on unsupervised learning schemes, users are divided into different clusters, so that users in the same cluster have similar consumption patterns, manifested as similar daily life and peak demand periods. Time series data mining based on meaningful statistics and behavioral features analyzes data and analyzes these behavioral features. This method only conducted user behavior

data mining once, and the mining results were not accurate enough. The analysis results of power grid user behavior were not ideal enough. Yin L et al. [11] quickly and accurately identified and analyzed the behavior of a large number of power users based on the GoogLeResNet3 network module. Design the GoogLeResNet3 network module, which includes a fully connected layer, Inception module, and residual module. Through the operation of different modules, complete user data mining and achieve power user behavior analysis. However, this system collects insufficient amount of user behavior data, and the basic data cannot fully reflect the user behavior, which ultimately affects the effect of user analysis. Baker, M [12] designed the power grid household power analysis system, which combines K-means clustering, K-nearest neighbor classification and ARIMA algorithm in data mining technology to analyze user behavior based on the behavior data of power grid users with prepaid meters, effectively integrating the electricity consumption behavior data of power grid users, improving the analysis efficiency of electricity consumption behavior data. However, in the application process, the system did not preprocess the data of power grid user behavior, which led to the inaccuracy of the final analysis of power grid user behavior. Qi, Z, et al. [13] designed an integrated clustering analysis system for users' electricity consumption behavior. The system first used principal component analysis (PCA) to reduce the dimensions of data by examining the complexity, randomness and uncertainty of users' electricity consumption. Then, a single clustering method is used, and most of them are selected for comprehensive clustering to obtain the classification of users' electricity consumption behavior and realize the analysis of grid users' behavior, effectively avoiding the impact of the complexity and uncertainty of user electricity behavior data on the analysis results, and effectively improving the analysis efficiency. However, the system can not obtain comprehensive data of power grid user behavior in the application process, so the analysis accuracy is insufficient. Hoendervanger, J and others [14], based on the power grid user behavior data, preprocessed the data into different groups and analyze the data according to the similarity of user electricity behavior data, effectively aggregate the characteristics of electricity consumption behavior data, improve the comprehensiveness of data analysis, however, in the process of clustering the electricity consumption behavior data, the selection parameters are influenced by the subjectivity, which leads to the poor analysis accuracy of this system.

Data mining [15] is a word that appears more frequently in different fields in recent years. It refers to the process of obtaining implicit information

from massive big data through statistics, machine learning, online analysis and other algorithms. Data mining technology [16] is widely used in various fields. It can help users process massive data and provide powerful data support for enterprise decision-making, business planning, etc. In order to solve the existing research methods of electric power user behavior analysis result is not ideal, based on data mining technology, this paper designs a grid user behavior analysis system based on data mining, innovatively using Python to crawl user behavior data, accurately obtaining sufficient and comprehensive grid user behavior information from the database, and performing information preprocessing, laying a solid data foundation for accurately analyzing grid user behavior. Using Fromto to filter the data categories of grid user behavior, conducting initial clustering, and using K-means clustering algorithm to perform secondary clustering on grid user behavior data, avoiding the influence of subjective factors on parameter selection, thereby improving the accuracy of mining results. The experimental results prove that the contribution of this method is to effectively obtain the electricity consumption behavior of different types of power user groups, ensure that the power companies further provide differentiated services, and provide accurate user behavior analysis data for the marketing planning of power enterprises.

2 Design of Power Grid User Behavior Analysis System

2.1 Overall Structure of the System

According to the idea of “layering”, the overall structure of the power grid user behavior analysis system is designed, as shown in Figure 1.

As shown in Figure 1, the power grid user behavior analysis system consists of big data acquisition layer, data storage layer, ZigBee communication layer, logical operation layer and visual display layer. Among them, the big data acquisition layer uses Python to crawl the power enterprise model management, monitoring management, quality management, service management and other grid user behavior data, and then obtains the grid user behavior data with real-time characteristics and stores it in the SQL database. Then, after accessing the grid user behavior data in the SQL database, the parsed and encapsulated grid user behavior data is transmitted to the logic operation layer through the server, gateway, data interface and sending module in the ZigBee communication layer; This layer uses the power grid user behavior analysis method of K-means data mining algorithm

to obtain the power grid user behavior analysis results, and then transmits them to the visual display layer. The visual display layer uses PC and mobile terminals to present the power grid user behavior analysis results and achieve human-computer interaction.

2.2 System Hardware Design

2.2.1 Selection of core controller

The core controller [17] of the power grid user behavior analysis system is the main hardware of the system. The pentium processor is selected here as the core controller of the power grid user behavior analysis system, and the structure of the core controller is shown in Figure 2.

As shown in Figure 2, the Pentium processor uses register components to complete numerical logic operations such as addition, division, multiplication, etc. It is connected to the integer register component through the controller component. The control ROM is used to store and register the data of the system, and the control unit is used to connect with other modules, and the DP logic, bus unit and APIC module are connected through the 8KB code high-speed register [18], integer register, etc., so as to realize the transmission and control of the data of the power grid user behavior analysis system.

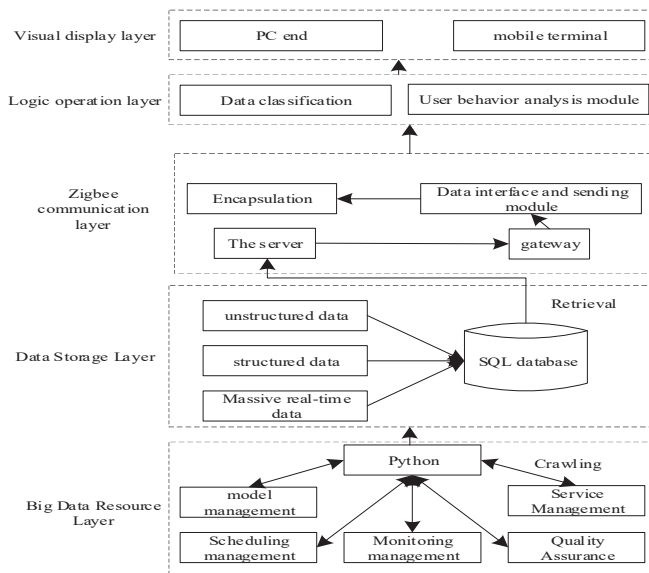


Figure 1 Overall structure of the power grid user behavior analysis system.

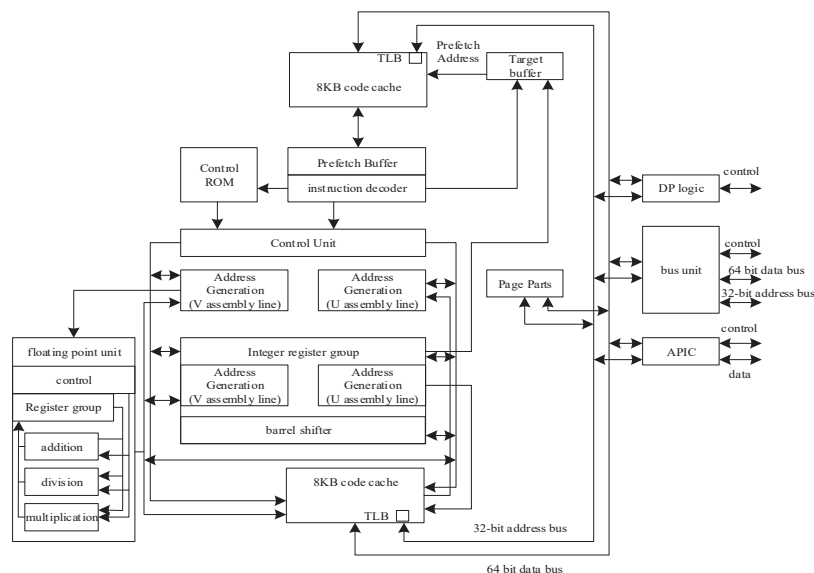


Figure 2 Schematic diagram of pentium processor structure.

2.2.2 Server erection

The erection of the server of the power grid user behavior analysis system is the basis of system information transmission and logical operation [19]. The structure of the server of the power grid user behavior analysis system is shown in Figure 3.

As shown in Figure 3, the server of the power grid user behavior analysis system is composed of acquisition thread, compression coding thread, sending/receiving thread and decoding display thread. The acquisition thread is responsible for receiving the power grid user behavior data, compressing and coding the data with the compression coding thread, and transmitting it to the sending thread [20]; The sending thread uses RDT protocol to package the power grid user behavior data, store it in the sending cache management, and then send the data package to ZigBee wireless sensor network. The ZigBee wireless communication technology is used to decode the user behavior data packet of the power grid according to the RDT protocol, display the data, and realize the data flow transmission between the entire system.

2.2.3 ZigBee wireless communication function design

The communication expansion board is designed based on ZigBee wireless communication technology to realize communication transmission between

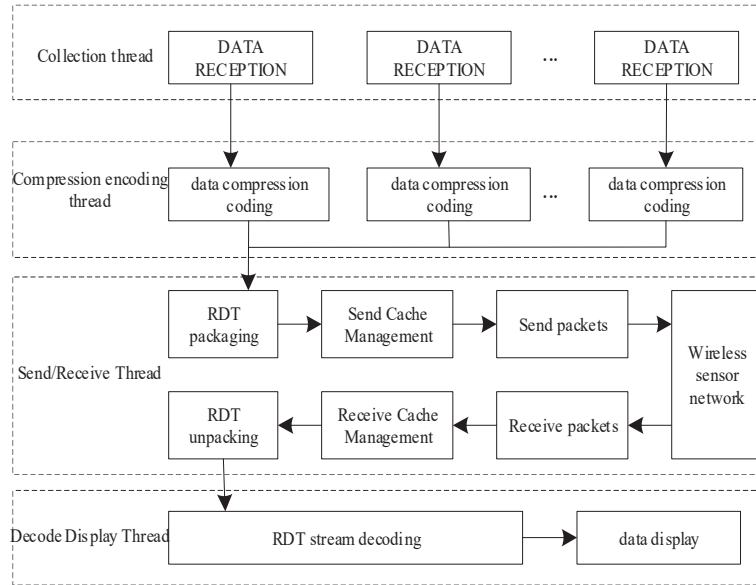


Figure 3 System server installation structure.

Table 1 ZigBee wireless communication technical parameters

Name	Parameter
Networking method	ZigBee Gateway
Communication frequency band	2.4 GHz
Transfer rate	100 Kbps
Communication distance	10–75 m
Safety	Strong
Consumption	Lower

different modules of the system. ZigBee wireless communication technical parameters are shown in Table 1.

Based on IEEE802.15.4 wireless personal LAN protocol, ZigBee wireless communication protocol stack structure is designed, as shown in Figure 4.

It can be seen from Figure 4 that the application layer consists of endpoints, APS security management, ZigBee device [21–23] objects and application objects, which respectively provide the application layer with advanced protocol stack management, device object interface and management service access point functions. The network layer uses the security mechanism to transmit data frame information, and can establish, maintain and close the current network at the same time. In the network layer, the

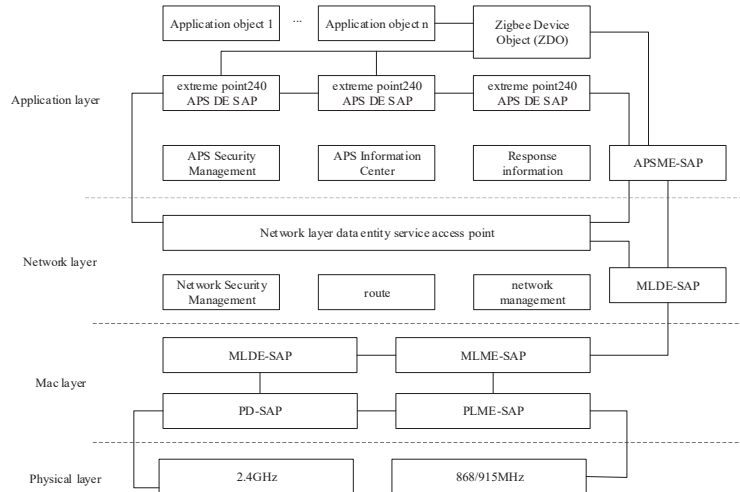


Figure 4 Schematic diagram of ZigBee wireless communication protocol stack structure.

data frame of the layer is composed of the frame header and the payload, and the frame control information, destination, etc. are stored in the data frame header. The network layer is connected to the MAC layer through interfaces MCPS-ASP and MLME-SAP, and is responsible for end-to-end communication transmission and network initialization. In the MAC layer of the protocol stack, it includes media access control and data links. Its main function is to establish communication links [24], and support the interruption and connection of individual LAN links. MAC layer establishes network transmission security mechanism by sending beacon or monitoring beacon. The physical layer of ZigBee wireless communication protocol stack [25] is responsible for receiving and transmitting electromagnetic waves, channel selection, and energy and signal interception. It provides 2.4 GHz and 868/915 MHz frequency band communication services for the entire network.

Design the overall structure of the peripheral power system operation of the communication expansion board, as shown in Figure 5.

In Figure 5, a small number of electronic components are installed in the peripheral circuit of ZigBee communication expansion board. The electronic components are led out of the input/output ports of the control chip, and the peripheral circuit is led out through each port [26] and connected with the ground wire to ensure the normal operation of the circuit. After completing the ground connection, it is necessary to set a circuit for generating a stable clock signal inside the peripheral circuit, and synchronously adjust the initial

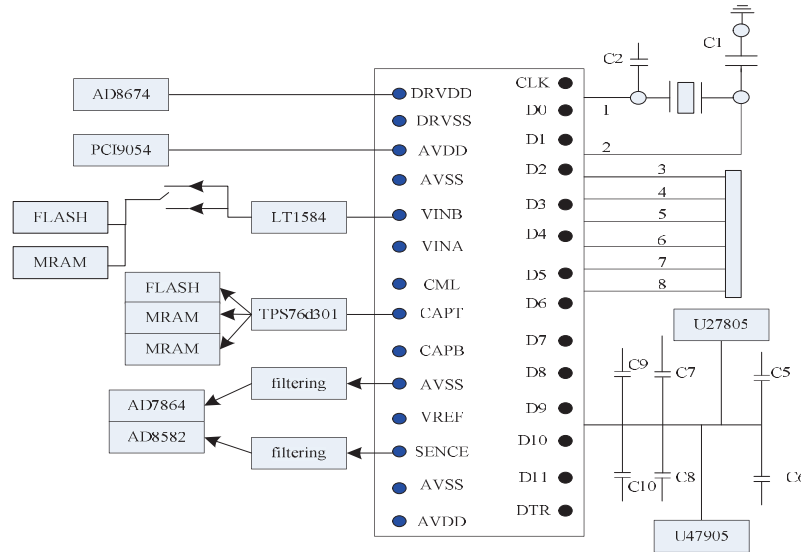


Figure 5 Schematic diagram of peripheral circuit structure of ZigBee communication expansion board.

parameters of the circuit with an oscillator. For external links [27], the oscillator is used to adjust the link parameters at the 32 MHz fixed frequency. The expansion board applies a reset circuit in the form of double reset. The circuit involves resistance, capacitance, reset switch and other components, which can convert high and low level status signals and protect the entire peripheral circuit. ZigBee communication receiving thread connects to the wireless sensor network, receives the power grid user behavior data packet from the network, and then caches the data packet; After unpacking the power grid user behavior data packet according to the RDT protocol [28], it is sent to the decoding display thread, which decodes the power grid user behavior data packet, displays the data, and realizes the data flow transmission between the entire system.

2.3 System Software Design

2.3.1 Crawling user behavior data based on Python

Based on the core controller and server configuration of the system, the big data collection layer of the system uses Python language to crawl user behavior data in the power grid, and maintains it through a crawling framework and crawling program to effectively crawl user behavior data resources in the

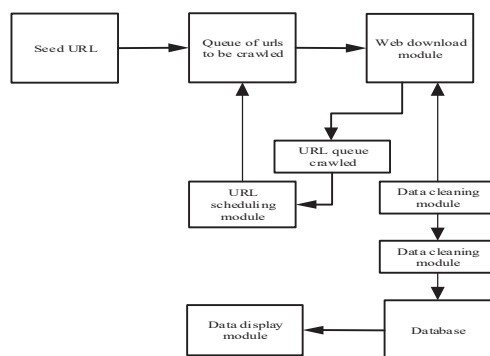


Figure 6 Frame structure diagram of climbing.

power grid. The crawling of power grid websites is based on user behavior classification, website crawling resources, and domain name positioning. It confirms the domain name of the crawled user behavior information, analyzes the correlation between user behavior information crawled by webpage themes, and determines the available user behavior information according to the set threshold setting. Then, the data crawling collection results are received by the ZigBee communication receiving thread in the hardware to receive power grid user behavior data packets, and the data packets are cached.

(1) Climbing frame process

The big data acquisition layer of the system crawls the grid user behavior data through Python, which is described in Figure 6.

From the analysis of Figure 6, it can be seen that the process of crawling the user behavior data of the power grid is as follows:

- ① Retrieve the valuable seed URL;
- ② Put the URL retrieved by the process (1) into the URL queue to be captured;
- ③ Get the URL to be crawled from the URL queue to be crawled, download its corresponding power grid web page, feed back the information of the web page to the data analysis module, and add these URLs again in the URL queue;
- ④ The data analysis module analyzes the grid user behavior data obtained from the download module, obtains valuable data from these data through regular expression, and cleans the data through the data cleaning module, while parsing other URL modules;

- ⑤ After step ④ is completed, feedback the module data to the scheduling module, take the feedback data as the priority queue, and analyze the data in combination with the captured URL queue. If the data parsing module in the scheduling module feeds back the URL data to the URL scheduling module, compare and analyze the obtained URL data and the captured URL queue. If it is found that these data are URLs that have been included and indexed by the search engine, filter out these data [29, 30], otherwise, integrate these data into the URL queue;
- ⑥ Repeat the execution process (3) to (5). When all URLs in the queue of URLs to be captured are captured, the crawling of user behavior information is terminated;
- ⑦ Clean up the data and store the data in standard format into the database;
- ⑧ Get the crawling results of grid user behavior information from the database, and these results are presented in text, pictures or other forms to accurately analyze the grid user behavior.

(2) Python based crawler maintenance

① Collect abnormal data of crawling program

The object of crawler detection and maintenance is abnormal data in the process of crawling. If a string of complete grid user behavior information characters are crawled, and n represent the number of characters in the complete information, then use the Python language to carry out the forward maximum matching, and get the character results of user behavior information according to the crawling arrangement results n_0 indicates the number of characters in the arrangement result [31, 32]. Take these characters as a character and compare them with the grid user behavior information crawled by the crawler program. If they match, the crawler program is stable; otherwise, the crawler program has abnormal data. Formula (1) indicates the domain name location where the crawler crawls user behavior information p :

$$p = \sum_{n \geq 0}^T n \lambda_T \tag{1}$$

Among them, n represents the number of crawling times of the crawler, T as well as λ represents the update time of power grid information and user behavior information crawling task queue respectively.

Based on the crawling sequence of the crawling program, the webpage can be crawled by matching the crawler technology with formula (1). If the

domain name location matches the crawl target, the next domain name will be matched; If there is no match, the abnormal data will be obtained and the power grid web page will be compiled in Python. Continuously cycle this process to complete the matching of all power grid web pages.

② Crawl program maintenance

When maintaining the crawler, it is necessary to build a maintenance objective function and bring the obtained exception data into the objective function. For the power grid website (URL), the user behavior classification is taken as the standard, and the website crawling resources and domain names are located p as a basis, confirm the domain name of the crawled user behavior information, so as to analyze the correlation between the user behavior information crawled by the web page subject. Set the relevance threshold to 1, o and μ represents machine learning and user behavior effectiveness parameters respectively [33–35], X represents random data in the crawler, s represents the crawling program model, and the correlation degree is analyzed by formula (2):

$$P_X(s|o) = \frac{p}{s(\mu_o|Z)}c \quad (2)$$

Where, c represents the probability of change in X in a random state; Z represents data validity parameters; P represent the domain name of the power grid user behavior resource, and the domain name is included in the crawl page. $P_X(s|o)$ represents the correlation degree between X and the crawled user behavior information, when $P_X(s|o)1$, it means the data crawled by the crawler is abnormal, use X' to represents the abnormal data.

Get all abnormal data X' , then, based on the power grid user behavior information website domain name identification results, positioning X' :

$$f(X') = \frac{1}{L'}L \left(\frac{1}{\theta} \right)^c X' \quad (3)$$

Among them, \Re represent the data target of positioning; θ represent the spatial dimension of user behavior information in the power grid; L as well as L' respectively represent the length of the exception data field and the corresponding user behavior resource field.

Realize the abnormal data location target through formula (3), clear, maintain the crawling program, and obtain complete and accurate data on power consumption behavior of power grid users X , and data mining.

2.3.2 Analysis of power consumption behavior of power grid users based on data mining

At the logical operation layer of the power grid user behavior analysis system, based on the core controller and server hardware, the self-organizing center K-means algorithm in big data mining algorithms is used to classify the power grid user behavior data received by the ZigBee communication receiving thread, and to mine the results of power grid user behavior analysis. The process of mining power grid user behavior analysis results by this algorithm is as follows:

Step 1: Initialize the initial weight of the self-organizing neural network. Select j Input nodes as the reference nodes, set W_j as the weight vector of the node, and then set the initial cycle times.

Step 2: From X to filter the data category of grid user behavior in X_K , and initial clustering was performed, where $K = 1, 2, \dots, m$ is the data category of grid user behavior, m is the total number of categories. Calculate the distance between the winner weight vector of each user behavior data type and this category. The expression formula is as follows:

$$[X_K - W_g] = f(X') \sum_{j=1}^p \|X_{Kj}\| - \sum_{j=1}^p \|W_j\| \quad (4)$$

In the above formula, g is the calculation unit representing the winner data; W_g represents the winner weight vector of the data class of grid user behavior.

In the result of formula (4), the Euclidean distance expression formula of any two grid user behavior vectors is as follows:

$$\|A - B\| = [X_K - W_g] \sqrt{\sum_{j=1}^p (e_j - f_j)^2} \quad (5)$$

In the above formula, A, B represent any two grid user behavior vectors; e_j, f_j respectively European space point coordinates of A and B .

Make W_g the winner neighborhood of the data class, approximately acts to the power grid of users N_g , after the input node responds when the self-organizing neural network operates stably, the output node outputs the initial clustering results of the grid user behavior analysis. The learning equation expression formula of this process is as follows:

$$\Delta w_{AB} = N_g \frac{|\eta(t) - \eta(t)w_{AB}|}{\|A - B\|} \quad (6)$$

In the above formula, t is the number of iterations; w_{AB} representation the connection weight between vector A and B ; $\eta(t)$ indicates the learning rate of t .

Step 3: Obtain the cluster center according to the initial clustering results generated by the output node, and use it as the center of each cluster in the K-means clustering algorithm.

Step 4: Use K-means clustering algorithm to cluster the user behavior data of power grid again. First, calculate the sum of squares of the power grid user data sample and the cluster center within the category $J(C)$, the calculation formula is as follows:

$$J(C) = \min \sum_{m=1}^K \Delta w_{AB} Z_n d^2(C_m, x_n) \quad (7)$$

In the above formula, Z_n represents a binary variable; C_m represented the cluster center of sample C ; n represents the clustering influence coefficient; x_n represents other sample data within the class; d represents the distance between the cluster centers of sample categories.

According to the results of formula (7), the power grid user behavior is classified, and the clustering center is updated using the least square method and Lagrange method. The obtained power grid user behavior analysis results are expressed as follows:

$$C_m = \frac{\sum_{n=1}^m Z_n x_n}{\sum_{n=1}^m J(C) x_n} \quad (8)$$

Step 5: After updating the cluster center according to formula (8), output the new cluster result, which is the analysis result of power grid user behavior.

After the above steps, the analysis results of power grid user behavior are obtained and sent to the visual display layer of the power grid user behavior analysis system for display to users.

The above method classifies the different types of electricity use of the same user and the electricity use behavior of different users, so as to facilitate the discovery of user electricity use patterns, analyze the electricity use behavior of specific users, and provide reliable analysis basis for power companies to better conduct demand side management and users to effectively arrange their own electricity use behavior. In addition, when clustering analysis is completed, different types of power user groups can

be effectively obtained to analyze the power consumption behavior of these user groups, which can ensure that power companies further provide differentiated services, guide users to reasonably regulate their own power consumption behavior, and provide good support for scientific and orderly power consumption.

3 Experimental Analysis

Taking an electric power service limited company as the experimental object, the electric power service limited company is mainly engaged in electric power marketing projects such as electric power sales, thermal power production and sales, and grid operation and management. The company was founded in 1999, and since its operation, the company has generated a large amount of grid user behavior data on its official website, and use it as a data source. The method in this paper is used to analyze the grid user behavior data of the electric power service limited company. It provides powerful data support for its marketing planning, and verifies the practical application effect of the method in this paper. The experimental environment is: Win 10 operating system, Intel i7-9400F processor and NVIDIA GeForce GTX1070 graphics card, and the data interface is the HDFS. The experiment uses Python programming language and MATLAB kit to train and test the logic operation layer K-means clustering algorithm on the Tensorflow deep learning framework. The experimental parameters of K mean clustering algorithm are: cluster number: 100; iteration number: 500; maximum evolutionary generation number: 200.

To verify the ability of the system in dealing with the data of power grid users' behavior, the amount of stored data is taken as the measurement index, and the system stores the amount of data under different operating times. In order to make the experimental results more sufficient, the literature [12] system, literature [13] system and literature [14] system are used to carry out comparative experiments with the system designed in this paper. The comparison method and the present system conduct experiments in an experimental environment. The experimental results are shown in Figure 7.

It can be seen from the analysis of Figure 7 that when the four systems store data, their storage data volume curves show a fluctuating trend with time. In the same running time, the storage data volume curve of this system has the smallest fluctuation, and its storage data volume value is the highest. The storage data volume curve from the literature [12] system to the literature [14] system has a large fluctuation, and under the same running time,

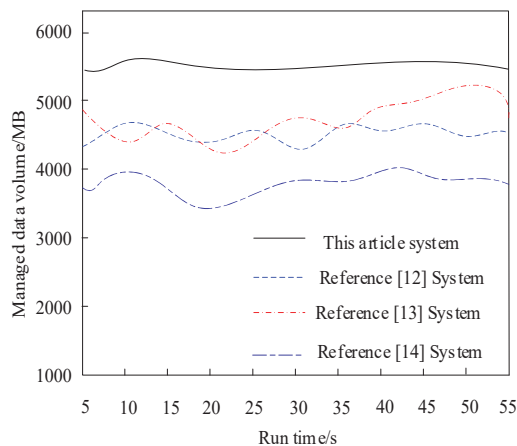


Figure 7 Storage data volume.

the amount of stored data is lower than that of this system. The above results show that during the operation of this system, the amount of user behavior data stored in the system is relatively large, and the amount of stored data is relatively stable in different operation times. The system has good operation stability and strong data processing capability.

Verify the communication transmission capability of the system in this paper, take the 200GB power grid user behavior data as the experimental object, test the communication fluctuation frequency change of the system in this paper when transmitting the power grid user behavior data, and analyze the wireless communication capability of the system in this paper. The test results are shown in Figure 8.

It can be seen from the analysis of Figure 8 that when the system in this paper transmits the behavior data of power grid users, the fluctuation amplitude of its communication eye fluctuates between 1 and -1 . Its communication eye wave shape is relatively regular, the fluctuation frequency difference is small, and the communication eye window is large. The above results show that the system in this paper has fast speed, good communication quality and good communication stability in the communication process.

Verify the ability of this method to crawl the data of grid user behavior, and use this method to crawl the monthly electricity consumption data in the grid user behavior data. Take three users as experimental objects, collect their monthly power consumption data within 12 months, and the collection results are shown in Table 2.

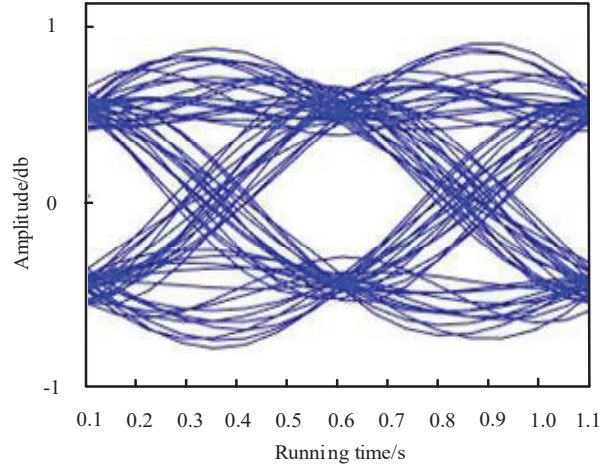


Figure 8 System wireless communication test results.

Table 2 Collection results of power grid user behavior data (kW.h)

Month	User 1	User 2	User 3
1	48296.1	589.4	83.2
2	38722.3	328.5	91.6
3	48841.9	205.9	118.5
4	42103.6	137.5	125.7
5	48207.5	132.2	66.4
6	45621.5	409.1	89.5
7	40335.8	228.5	91.7
8	38966.5	307.4	105.5
9	45063.8	421.1	133.6
10	48917.2	451.5	141.2
11	45066.3	406.5	108.2
12	42517.7	411.7	98.3

It can be seen from the analysis of Table 2 that the method in this paper can effectively collect the power consumption behavior data of power grid users in different months. Taking the above data as the basis for the analysis of power grid user behavior, can get the user’s power consumption behavior at different times, which provides a basis for formulating the power product marketing plan.

Validate the results of the system analysis of grid user behavior in this paper, take the weight index of grid user analysis as the measurement index,

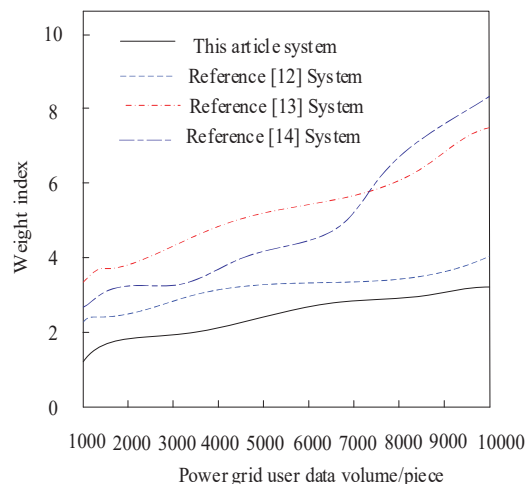


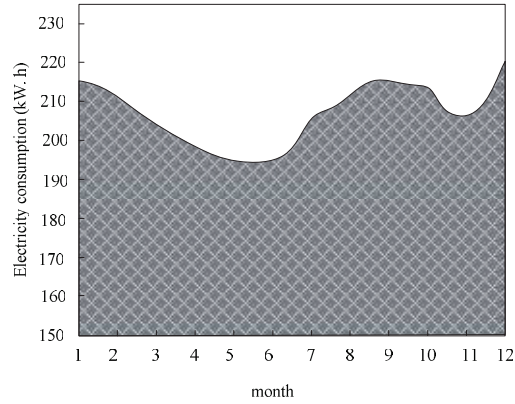
Figure 9 Weight index for analysis of power grid user behavior data.

test the change of weight index of four systems when analyzing different grid user behavior data, and the results are shown in Figure 9.

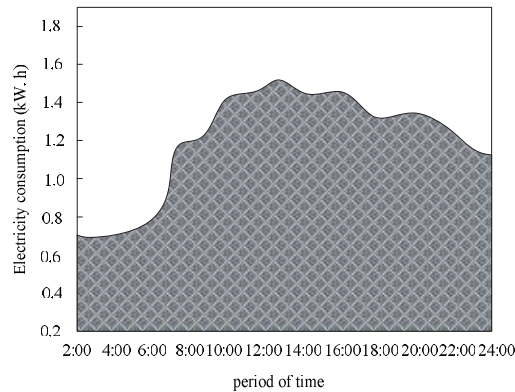
It can be seen from the analysis of Figure 9 that when the four systems analyze the behavior data of grid users, their weight indexes increase with the increase of the amount of grid user data. In this method, after the amount of grid user data exceeds 2000, their weight values are the lowest, and show a slow upward trend with the increase of the amount of grid user data. The weight index of other systems shows a significant upward trend with the increase of power grid user data. This is because other methods only conducted user behavior data mining once, and the mining results were not accurate enough. The method used in this article uses Fromto to filter the data categories of grid user behavior, perform initial clustering, and use K-means clustering algorithm to perform secondary clustering on grid user behavior data, avoiding the influence of subjective factors on parameter selection and keeping the weight value as low as possible. The above results show that the system in this paper has good mining stability when mining power grid user behavior data.

Taking a power grid user as an experimental object, this system is used to mine the user behavior data, and the mining results are shown in Figure 10.

From the analysis of Figure 10(a), it can be seen that the user's electricity consumption in January March and July October is relatively high, compared with that in March June and November. By December, its electricity



(a) Monthly electricity consumption of users



(b) Power consumption during user periods

Figure 10 User behavior data mining results.

consumption showed an upward trend. According to the annual power consumption curve of the grid user, the user is an ordinary user of the company, and its power consumption is for daily life. From January to March, the power consumption of the user shows a downward trend, because the use time of air conditioners in China can gradually decrease from January to March; From March to June, the temperature conditions are good, and users do not need to turn on the air conditioning equipment; From July to October, the temperature is high, and users use air-conditioning equipment frequently, so their power consumption shows an upward trend; The same is true of the increase in electricity consumption in December. It can be seen from the

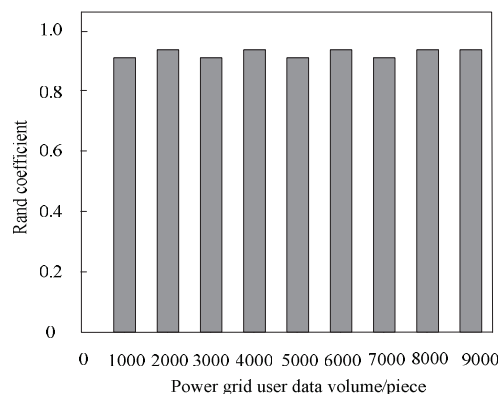


Figure 11 Rand coefficient.

analysis of Figure 10(b) that the user's power consumption shows a peak form from 2:00 to 24:00. The user's power consumption is low from 2:00 to 6:00, and the user's power consumption is high from 8:00 to 22:00. The user uses the range hood, washing machine, air conditioner and other equipment in this period, so the user uses more power from 8:00 to 22:00. Based on the above results, the application of this system can effectively analyze the daily electricity use behavior of power grid users from their behavior data. According to the electricity use behavior of users, the electricity price ladder can be formulated to promote the marketing of power products and services.

Validate the accuracy of the analysis results of the system mining power grid user behavior in this paper, and test the performance of the analysis results of the mining power grid user behavior under different data volumes of power grid user behavior with the RAND coefficient as the measurement index. The Rand coefficient is a commonly used clustering algorithm evaluation indicator, which does not require modification or special processing of the original data. It only needs to compare the consistency of two clustering results to complete the analysis. At the same time, the Rand coefficient can handle situations where the size of different clustering results is inconsistent, and even if the number of clusters in two clustering results is different, it can be compared. Therefore, it is suitable for power grid data mining evaluation with unmodifiable and high number of clusters in clustering results. The test results are shown in Figure 11.

It can be seen from the analysis of Figure 11 that the RAND coefficient of this system fluctuates with the data volume of power grid users when mining the analysis results of power grid user behavior data, but the fluctuation

range is small, and the RAND coefficient always remains at about 0.9. The results show that the system in this paper has a good consistency of clustering categories when mining the analysis results of power grid user behavior data by clustering, and the output analysis results of power grid user behavior data have a high accuracy.

4 Conclusion

This paper designs a power grid user behavior analysis system based on data mining, and uses the self-organizing center K-means algorithm in big data mining algorithm to mine the results of power grid user behavior analysis, combining the advantages of self-organizing neural network and K-means algorithm. When mining the analysis results of power grid user behavior, it first uses self-organizing neural network to preliminarily cluster the data of user power consumption behavior, and then uses K-means algorithm to conduct clustering mining again according to the preliminary clustering results. By mining the data of power grid user behavior twice, the mining results are more accurate and refined, and the output analysis results of power grid user behavior are more accurate. After multi angle experimental verification, this system has a large storage capacity for user behavior data during operation, and the memory storage is relatively stable at different running times, indicating that the system runs stably and has good data processing capabilities. This system has the advantages of fast speed, good communication quality, and good communication stability during the communication process. This system can effectively collect data on the electricity consumption behavior of grid users in different months, thereby obtaining the electricity consumption behavior of users in different periods and providing a basis for formulating power product marketing plans. This system has good mining stability when mining behavior data of power grid users, and can effectively analyze their daily electricity consumption behavior from the behavior data of power grid users. The fluctuation range of the Rand coefficient of this system is relatively small, always maintaining around 0.9, and the clustering analysis results of power grid user behavior data have high accuracy.

5 Future Works

However, in the process of data mining, it is necessary to collect, transmit, and store power grid user behavior data, but sufficient security measures have not been taken to ensure data security and privacy. In the future, deep learning

and data desensitization technologies will be applied to optimize the system, while deeply mining the hidden patterns and features of user behavior while protecting user privacy and security.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Funding Statement

There is no specific funding to support this research.

References

- [1] Real-time Market-to-Market Oscillation Management. DOI: 10.1109/PESGM46819.2021.9638189.
- [2] Karthikeyan, S. Prabhakar, Saravanan, B., Jain, Aman, et al. A comparative study on transmission network cost allocation methodologies[C]. //2013 International conference on power, energy and control. Institute of Electrical and Electronics Engineers, 2013:145–152.
- [3] Deng S, Cai Q, Zhang Z, et al. User Behavior Analysis Based on Stacked Autoencoder and Clustering in Complex Power Grid Environment. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 3(15):1–15.
- [4] Guo Y, Wu L, Wang C, et al. Optimal Configuration of Multi-energy Storage for Load Aggregators Considering User Behavior[C]//2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia). IEEE, 2020. DOI: 10.1109/ICPSAsia48933.2020.9208441.
- [5] Equilibrium Analysis of Electricity Market With Demand Response Exchange to Counterbalance Bid Deviations of Renewable Generators. DOI: 10.1109/JSYST.2019.2928042

- [6] Kapici, E., Kutluay, E., and Izadi-Zamanabadi, R. (2022). A novel intelligent control method for domestic refrigerators based on user behavior. *International Journal of Refrigeration* (136–), 136.
- [7] karthikeyan, Prabhakar S, Palanisamy, et al. Comparison of Intelligent Techniques to Solve Economic Load Dispatch Problem with Line Flow Constraints[C]//Advance Computing Conference, IACC, 2009 IEEE International. 2009.
- [8] Li, H., Chen, Q., Zhong, Z., Gong, R., and Han, G. (2022). E-word of mouth sentiment analysis for user behavior studies. *Information Processing & Management: Libraries and Information Retrieval Systems and Communication Networks: An International Journal*. 5(1), 59–63.
- [9] Lu, R., Liu, N., Li, D., Luo, X., and Fan, Y. (2021). Intelligent monitoring analysis of power grid monitoring information based on big data mining. *Journal of Physics: Conference Series*, 1992(3), 32132.1–032132.8.
- [10] Kaur R, Gabrijelcic D. (2022). Behavior segmentation of electricity consumption patterns: A cluster analytical approach. *Knowledge-based systems*, 251(5):109236.1–109236.16.
- [11] Yin L, Zhong Q. (2023). GoogLeResNet3 network for detecting the abnormal electricity consumption behavior of users. *International journal of electrical power and energy systems*, 145(2):108733.1–108733.11.
- [12] Baker, M. A. (2021). Household electricity load forecasting toward demand response program using data mining techniques in a traditional power grid. *International Journal of Energy Economics and Policy*, 11(4), 132–148.
- [13] Qi, Z. A., HI, B., Xw, A., Tp, A., and Jw, A. (2019). Analysis of users' electricity consumption behavior based on ensemble clustering. *Global Energy Interconnection*, 2(6), 479–488.
- [14] Hoendervanger, J. G., Yperen, N., Mobach, M. P., and Albers, C. J. (2022). Perceived fit and user behavior in activity-based work environments. *Environment & behavior* 54(1):143–169.
- [15] Suo, N., and Zhou, Z. (2021). Computer assistance analysis of power grid relay protection based on data mining. *Computer-Aided Design and Applications*, 18(S4), 61–71.
- [16] Hong, Z., Wei, Z., Li, J., and Han, X. (2021). A novel capacity demand analysis method of energy storage system for peak shaving based on data-driven. *The Journal of Energy Storage*, 39(7), 102617.

- [17] Noami, A., Kumar, B. P., and Chandrasekhar, P. (2020). Design and Implementation of a United Multi-Core Memory Controller using AXI4. Lite Interface Protocol, 4(7), 45–51.
- [18] Ravi, N., Rao, T. S., and Prasad, T. J. (2022). Pipelined c 2 mos register high speed modified, *Int. J. Advanced Networking and Applications*, 3(1), 1031–1034.
- [19] Helms, P., and Limmer, D. T. (2022). Stochastic thermodynamic bounds on logical circuit operation, *arXiv*, 11(1), 670–676.
- [20] Blasco-Arcas, L., Kastanakis, M. N., Alcaiz, M., and Reyes-Menendez, A. (2023). Leveraging user behavior and data science technologies for management: an overview. *Journal of Business Research*, 154(3), 3–7.
- [21] Liu Z, Wang Y, Zeng Q, et al. (2021). Research on Optimization Measures of Zigbee Network Connection in an Imitated Mine Fading Channel. *Electronics*, 10(2), 171–178.
- [22] Yao S, Feng L, Zhao J, et al. (2021). PatternBee: Enabling ZigBee-to-BLE Direct Communication by Offset Resistant Patterns. *IEEE Wireless Communications*, 28(3), 130–137.
- [23] Palate, B. O., and Vera, E. G. (2020). Optimizing Aggregators Placement in Distribution Networks Using ZigBee Technology through PSO and Graph Theory. 2020 IEEE PES Transmission & Distribution Conference and Exhibition – Latin America (T&D LA). 28(9), 2039–2045.
- [24] Oveisi, M., and Heydari, P. (2022). A study of ber and evm degradation in digital modulation schemes due to pll jitter and communication-link noise. *IEEE transactions on circuits and systems, I. Regular papers: a publication of the IEEE Circuits and Systems Society*, 1(8), 69–75.
- [25] Desnanjaya, I., Nugraha, I., Pranata, I., and Harianto, W. (2021). Stability data xbee s2b zigbee communication on arduino based sumo robot. *International Journal of Robotics and Control*, 2(3), 153–160.
- [26] Cayre, R., Galtier, F., Auriol, G., Nicomette, V., Kaaniche, M., & Marconato, G. (2021). WazaBee: attacking Zigbee networks by diverting Bluetooth Low Energy chips. *Dependable Systems and Networks. IEEE*, 6(8), 21–24.
- [27] Bousbaa, Z., Nowak-Brzezińska, A. (2023). Knowledge Engineering and Data Mining *Electronics*, 12(4), 927.
- [28] Konys, A., Sanchez-Medina, J., Bencharef, O.(2023). Financial Time Series Forecasting: A Data Stream Mining-Based System. *Electronics*, 12(9),2039.

- [29] Abideen, Z. U., Mazhar, T., Razzaq, A., Haq, I., Ullah, I., Alasmay, H., and Mohamed, H. G. (2023). Analysis of Enrollment Criteria in Secondary Schools Using Machine Learning and Data Mining Approach. *Electronics*, 12(3), 694.
- [30] Wang, X. Z., Ruan, J. J., Zhou, T. T., Peng, X. L., Deng, Y. Q., and Yang, Q. Y. (2022). Data Mining in the Vibration Signal of the Trip Mechanism in Circuit Breakers Based on VMD-PSR. *Electronics*, 11(22), 3700.
- [31] Chen, G. Y., Zhu, Z. Y., Yang, L., Huang, W. H., Zhang, Y. Z., Lin, G., and Zhang, S. J. (2022). Intelligent Identification and Order-Sensitive Correction Method of Outliers from Multi-Data Source Based on Historical Data Mining. *Electronics*, 11(18), 2819.
- [32] Envelope T F P. (2021). Research on automatic user identification system of leaked electricity based on Data Mining Technology – ScienceDirect. *Energy Reports*, 7(11): 1092–1100.
- [33] Sajwan K, Sharma M, Shukla A K. (2021). Performance Evaluation of Two Medium-Grade Power Generation Systems with CO₂ Based Transcritical Rankine Cycle (CTRC). *Distributed Generation and Alternative Energy Journal*, 35(2):111–138.
- [34] Kumar P N, Chengaiah C, Rajesh P, et al. (2021). A Hybrid Technique for the Performance Optimization in the Combustion Process of a Power Plant Boiler: An Efficient ANNSSA Technique. *Distributed Generation and Alternative Energy Journal*, 28 (5):1561–1572.
- [35] Qadeer A, Khan M E, Alam S. (2021). Estimation of Solar Radiation on Tilted Surface by Using Regression Analysis at Different Locations in India. *Distributed Generation and Alternative Energy Journal*, 35(1): 1–18.

Biographies



Yan Wang was born in March 1995 and graduated from Shanxi University of Science and Technology in 2018. Currently, he works at the Information and Communication Branch of Hainan Power Grid Co., Ltd. She research interests include computer application technology.



Jiawei Xu was born in November 1974 and graduated from Jiangxi University of Technology in 2010. Currently, he works at the Information and Communication Branch of Hainan Power Grid Co., Ltd. He research interests include computer science and technology.



Xiaowen Chen was born in July 1987 and graduated from South China University of Technology in 2013. Currently, he works at Hainan Power Trading Center Co., Ltd. His research interests include computer application technology.



Ying Huang was born in August 1989 and graduated from China University of Mining and Technology in 2011. Currently, she works at Hainan Power Grid Co., Ltd. Her research interests include electricity marketing and electricity fee management.

