Multi-dimensional Park Portrait Model Based On Clustering

Linjuan Zhang1, Changqing Xu1, Lili Wang1, Chenxi Li2,*

1State Grid Henan Economic Research Institute, Zhengzhou, China
2North China Electric Power University, Beijing, China

*Corresponding author: chenxili@ncepu.edu.cn

Abstract. With the rapid growth of the social economy, the number of various types of parks in China has increased significantly. Faced with the increasing difficulty of managing the power consumption of parks, qualitative analysis labels such as industrial parks and technology parks are no longer applicable, and utility companies need to develop differentiated and personalized power supply strategies for parks with different characteristics. Therefore, it is necessary to study the power consumption characteristics of the park more detailed. A multi-dimensional park portrait model is proposed in this paper. Firstly, based on biclustering algorithm, the user’s power consumption data are analyzed. Then the combined forecasting model is established to analyze the future power consumption of the park. Besides, an index evaluation system is established to complete the power configuration and demand response analysis of the users’. Finally, K-means cluster analysis is performed on the analysis results to obtain the park portrait. 90,000 pieces of data and related electricity business records from Henan Province are used to verify the effectiveness of the model. The results show that the park portrait obtained by the model and algorithm can quantitatively analyze the relevant attributes of the park.

1. Introduction

The development of the park economy and the creation of leading industrial clusters can promote the adjustment of industrial structure and change the mode of economic growth. Provinces and cities in China have launched park construction. Considering that the power consumption level of the park is relatively high, and there are large differences in the power consumption levels of different parks, which undoubtedly increases the difficulty of power management in the park [1]. In order to ensure the power supply and service quality of the park, the utility company need to provide differentiated and personalized power supply services for the park.

At present, with the development of Big Data, Internet of Thing and Artificial Intelligence technology, user portrait technology has a wide range of applications in many fields, i.e. the analysis of the occupational adaptability [2] and APP download preferences of mobile phone users [3], the built of recommendation system [4-6], and precise marketing [7]. In the field of electricity, in [8], based on the data provided by the utility company, the user label information is extracted, the user label library and user portrait system are established. With the help of fuzzy c-means (FCM) clustering algorithm, a user power consumption behavior portrait is created from three perspectives, i.e. temperature sensitivity, electricity price sensitivity and power consumption stationarity [9]. In [10], the K-means algorithm is used to realize the clustering of power user labels and analyze the typical characteristics of their power behav
or. However, the above research is based on a single user as the research unit, without analyzing the characteristics of the user group.

An algorithm for quantitative analysis of park data characteristics and establishment of park portraits are presented in this paper. Firstly, according to electricity behavior, energy demand, and demand response, a user profile for the park users is established. Based on SpectralBiclustering algorithm, the electricity consumption behavior is analyzed [11]. The combined forecasting model, combined with Logistic model and improved gray Verhulst model is applied to predict power demand growth. Then the evaluation index system of the power demand is established. The demand response evaluation algorithm of electric mode and demand response load reduction rate realizes a comprehensive analysis of the demand response ability of the park users. After analyzing the user's three-dimensional characteristics through the above algorithm, the results are clustered to obtain the park portrait.

2. Establishment of Multi-Dimensional Park Portraits

The park portrait is based on the analysis of the user portraits of the users in the park and the clustering of the user portrait results.

In this paper, the three dimensions of electricity behavior, power configuration, and demand response related to the power supply strategy are selected as the components of the user portrait. The electricity behavior analysis will show the user's habitual electricity behavior pattern. The power configuration analysis will reflect the quantitative analysis result of the user's intention to expand the power configuration in the future. And the demand response analysis will show the quantitative evaluation of the user's ability to cooperate with the power grid to change the electricity consumption habit.

Afterwards, K-means clustering is performed on the three dimensions of user portraits, and the proportion of users with different clustering labels to the total number of users in the park is counted. The sequence of proportions is used as the result of the park portrait. However, the result of the user's electricity consumption behavior is the label, and the analysis result has no numerical correlation, so the proportion of different labels is directly counted. These sequences are the results of quantitative analysis of the park.

Figure 1. Park portrait construction process.
3. Establishment of User Portraits in Multi-Dimensional Parks

3.1. Analysis of User's Electricity Behavior
The purpose of electricity behavior analysis is to categorize and summarize the user's electricity habits, and help utility companies cooperate with users' electricity habits to provide personalized services.

The SpectralBiclustering algorithm is chosen to analyze in this paper. The biclustering algorithm can analyze the data from both rows and columns to obtain the local clustering results of the data. The flow chart of SpectralBiclustering algorithm is as Fig.2.

The preprocessing part converts user data of different power voltage level to the same energy unit.

3.1.1. Bistochastic Normalization
The Bistochastic Normalization method can normalize rows and columns at the same time, thus ensuring that the impact of useless information and noise on local feature mining is reduced without destroying the connection between data. The process is shown in Table 1.

Table 1. Bistochastic Normalize Normalization Process.

| Step | Description |
|------|-------------|
| ①   | Suppose the number of iterations is k, and k=1; |
| ②   | Suppose the data matrix \( \mathbf{A} \) is the data of the \( i \)th row, the \( j \)th column in matrix \( \mathbf{A} \). The diagonal matrix \( R_i \) and \( C_j \) represent the sum of the rows and columns of the data matrix respectively. \( R_i \) is the sum of the values of the \( i \)th row, \( C_j \) is the sum of the values of the \( j \)th column. The auxiliary computing matrix is \( \mathbf{R} = R^{-1} \mathbf{A} C^{-1} \). |
| ③   | Calculate the difference matrix before and after normalization \( D^k = \mathbf{A}^k - \mathbf{R} \mathbf{A}^k \). Calculate the Euclidean norm of the matrix by rows, i.e. \( D^k = \| D^k \| = \sqrt{\sum_{i=1}^{k} (D^k_i)^2} \). |
| ④   | Let \( \mathbf{A}^k = \mathbf{R} \mathbf{A}^k \), that is, the result obtained by iteration is \( \mathbf{A}^k \). |
| ⑤   | Let \( D^k < \epsilon \) or \( k = K \) is satisfied, stop the iteration and get the normalized result \( \mathbf{A}^k \). Otherwise, repeat process \( ②-④ \). \( \epsilon \) is the preset threshold, and \( K \) is the preset maximum number of iterations. |
3.1.2. **Singular value decomposition algorithm (SVD decomposition).** The obtained normalized data is decomposed using SVD decomposition algorithm. The reason is that if the data is regarded as a matrix, the local feature is the sub-matrix. The SVD decomposition algorithm is used to split the data matrix into the form of feature vector groups and feature value multiplication, which is convenient to find the feature vectors forming the local feature sub-matrix from the feature vector groups.

The SVD decomposition algorithm can decompose the data into two eigenvectors \( U \) and \( V \) to form a matrix and an eigenvalue matrix. Among them, the matrix \( U \) contains the information in the row direction of the data matrix \( A \), and the matrix \( V \) contains the information in the column direction of the data matrix \( A \).

Note that the matrix normalized by the Bistochastic Normalization method still contains global information, which is information that affects the analysis from both rows and columns. Global information can interfere with local information mining and needs to be removed. Therefore, after SVD decomposition, the feature vector corresponding to the largest feature value needs to be discarded. Because the eigenvector corresponding to the largest eigenvalue is the most important vector for the reconstructed data matrix, which contains global information.

3.1.3. **Use K-means clustering algorithm to filter vector groups for data transformation.** After obtaining the feature vectors of the data matrix, it is necessary to filter out the vectors that constitute the sub-matrix with local feature. Since the dimension of the sub-matrix with local features must be smaller than the dimension of the data matrix. Therefore, the rank of the eigenvector of the searched sub-matrix must be smaller than the dimension of the data matrix. This means that the vectors in \( U \) and \( V \) that contain local feature information can be transformed to zero or near zero at a partial position in the vector by row-column transformation. Therefore, it is possible to measure the possibility that the vector contains local features by clustering the vector itself. For clustering vectors, the smaller the sum of the differences among the clustering results, the more likely the clustering vector is the sub-matrix feature vector.

Perform clustering on each vector in the vector group \( U \), calculate the distance from the value in the vector to the cluster center, and obtain the euclidean norm of the distance. Sort the vectors according to norm calculation results from small to large. The vector group formed by the first \( N_{\text{best}} \) vectors is used for the subsequent classification calculation.

(3) **Classification with the help of K-means clustering algorithm**

After executing the above algorithm, the vector group can be obtained and multiplied by the normalized data matrix. K-means clustering is performed on the multiplied result \( Pr \). Because of filtering vectors from matrix \( U \), the clustering result is the row label of the data. And that is the user behavior label.

\[
Pr = A \cdot V_{tr}
\]  

Statistically count the occurrence of user behavior labels of users' daily electricity data. The user behavior labels that appear the most times are the analysis results of the user's behavior data, and reflect the user's electricity habits. Since the electricity consumption data has both active and reactive power, the analysis result is a combination of active and reactive power's electricity habits, which can reflect the user's power consumption habits in both active and reactive.

3.2. **Analysis of User Power Configuration**

Power configuration analysis is to quantitatively analyze the growth trend of users' energy demand, so as to provide a reference for the power supply company to add capacity and corresponding equipment.

In order to analyze the demand for power configuration, medium and long-term load forecasting is required. For medium and long-term load forecasting, the prediction results obtained by using a certain forecasting method alone have large errors. In order to make full use of the useful information in different load forecasting models, a combination forecasting model based on equal weight recursion theory is proposed. Combined with the gray Verhulst model [12-13], the forecast method is used to predict the electricity consumption and annual maximum load of regional power users in the next 3 years.
The predicted results are compared with the current electricity load scale, so as to obtain two main indicators to evaluate the user's power configuration demand, i.e. annual average power consumption growth rate and annual average maximum load growth rate. In addition, comparing the current maximum load scale and the maximum load forecast scale in the next 3 years with the current contract operating capacity, two other evaluation indicators can be obtained, i.e. the current load factor of the transformer and the expected load factor of the transformer. The calculation of the above four indicators is shown in (2):

\[
\begin{aligned}
G_{rate,1} &= \frac{1}{3} \left( \frac{Q_1}{Q} + \frac{Q_2}{Q_1} + \frac{Q_3}{Q_2} \right) - 1 \\
G_{rate,2} &= \frac{1}{3} \left( \frac{P_1}{P} + \frac{P_2}{P_1} + \frac{P_3}{P_2} \right) - 1 \\
L_{rate,1} &= \frac{p}{P_{max}} \\
L_{rate,2} &= \frac{P_1 + P_2 + P_3}{3P_{max}}
\end{aligned}
\]

In the formula, \(G_{rate,1}, G_{rate,2}\) respectively represent the user's average annual power consumption growth rate and annual average maximum load growth rate. \(L_{rate,1}, L_{rate,2}\) respectively represent the current load rate of the transformer and the expected load rate of the transformer in the next 3 years. \(Q, Q_1, Q_2, Q_3\) respectively represents the user's current annual power consumption in the next one two and three years. \(P_{max}\) represents the user's current contract operating capacity.

The index system for evaluating the maturity of energy allocation needs of power users in the park is shown in Table 2.

| Target                                      | Index                                                                 |
|---------------------------------------------|-----------------------------------------------------------------------|
| Maturity of power configuration needs of power users in the park | J1 Annual average power consumption growth rate (%)                   |
|                                             | J2 Annual average maximum load growth rate (%)                        |
|                                             | J3 Current load factor of transformer (%)                             |
|                                             | J4 expected load factor of the transformer in the next 3 years (%)     |

After calculating the four indicators, the weights are confirmed according to actual experience, and the analysis results of the indicators are weighted and summed. The result is used as the evaluation result of power configuration. The analysis of power configuration here should include electricity, heat and gas. But the analysis methods of the three are the same, so this will only introduce electricity as an example.

3.3. Analysis of User Demand Response

Demand Response (DR) reflects the user's ability to participate in the adjustment and interaction process of peak shaving and valley filling in the system [14-17]. To evaluate the DR capacity, select the load mode, load reduction rate, power consumption coefficient of variation and total power outage time to establish an evaluation index system. These four indicators are the reflection of user DR capabilities on different sides. After completing the calculation of the four indicators, the four indicators are weighted and summed to obtain the final DR capability evaluation result.
The user DR process based on the minimum load electricity consumption mode is as follows. The user's minimum load power consumption mode refers to the load mode with the smallest total load among all typical daily load modes of the user.

\[ P_{\text{min}}(t) = C_g(t) \text{ if } \sum C_g(t) = \min \sum C_k(t) \quad k = 1, 2, \ldots, m \]  

(3)

In the formula, \( P_{\text{min}}(t) \) is the user's minimum load power mode. \( C_k(t) \) is the user's k-th load mode. \( m \) is the total number of users' typical daily load modes.

The user's DR potential is as (4)

\[ \text{DRP}_1 = \sum_{k=1}^{M} (\sum C_k(t) - P_{\text{min}}(t)) \cdot N_k \]  

(4)

In the formula, \( \text{DRP}_1 \) is the user's DR potential. \( N_k \) is the total number of similar days under the k-th load pattern.

Then consider the coefficient of variation of electricity consumption, which reflects the fluctuation level of electricity consumption of users during peak, valley and even periods.

The calculation formula of the coefficient of variation of electricity consumption during peak, valley, and equalization of users are shown in (5).

\[ CV_1 = \frac{SD_1}{MN_1} \]  

(5)

In the formula, \( CV_1 \) represents the coefficient of variation of electricity consumption at peak, valley, and halves of users. \( SD_1 \) represents the standard deviation of electricity consumption at peak, valley, and halves of users. \( MN_1 \) represents the average value of electricity consumption at peak, valley, and halves of users.

In addition, the total power outage duration of power users in the park can also reflect the user's DR ability to a certain extent. The longer the user's total outage duration, the easier it is for users to accept load-cutting orders issued by regional grid operators.

In summary, the indicator system for DR of power users in the park is shown in Table 3.

| Target | Index |
|--------|-------|
| Maturity of DR of power users in the park | \( K_1 \) DR potential \( \text{DRP}_1 \) |
| \( K_2 \) Variation coefficient of electricity consumption | |
| \( K_3 \) Total power outage (h) | |

The four indicators in the table are given weights for summation, and the result is named DR maturity \( K \) as the user DR evaluation result.

4. Case Study

4.1. Electricity Behavior Characteristics

Through experiments, when \( N_{\text{best}} = 40 \) in the algorithm and the number of clusters is 8, the electricity consumption data can be clearly classified. The clustering results of active electricity consumption data are shown in Fig.3. The user behavior analysis algorithm is applied to obtain all users' (active, reactive) electricity consumption behavior labels.
Figure 3. Active behavior data clustering results.

The ratio of the number of different types of users in the park to the total number of users in the park is shown in Table IV. The sequence composed of these proportional results contains the structural characteristics of the users in the park, and is used as the characteristic sequence of the electricity consumption behavior of the park.

Table 4. Characteristic sequence of electricity consumption in the park.

| Label | (0,0) | (0,4) | (0,6) | (2,5) | (3,4) | (4,7) | (6,1) | (6,6) |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Percentage | 0.869 | 0.036 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.036 |

It can be seen from Table IV that most park users are the (0,0) type. According to the analysis in Figure 3, it can be seen that the fluctuation of this type of electricity consumption is relatively small, and it belongs to the typical electricity consumption enterprise in the park.

4.2. Power Configuration Characteristics

The power configuration maturity of each user can be obtained by using the algorithm of power configuration feature analysis. After calculating the power configuration maturity of all users, K-means clustering algorithm is used for analysis. Experiments show that when the number of clusters is 3, the error between clusters can keep a low level and the number of clusters is small. When the cluster number is 3 as the basis for cluster analysis, the effect is shown in Fig.4.
Figure 4. Clustering results of power configuration demand maturity.

It shows that most users in the park are still at a relatively low level of power configuration maturity and are great potential. Considering that the Henan park selected for calculation is a new park that has just opened for two years, this result is consistent with the actual situation in which a large number of users in the park are planning to expand production.

On this basis, the experimental statistics on the proportion of users of different types of power configuration levels in the park accounted for the total number of users in the park. The sequence formed by these three proportional results contains the characteristics of the energy distribution level distribution structure of the park users.

Table 5. The Power Configuration Sequence.

| Label | Percentage |
|-------|------------|
| 0     | 0.375      |
| 1     | 0.488      |
| 2     | 0.138      |

4.3. DR Characteristics
The analysis ideas in the characteristics of the same energy allocation. After calculating the DR maturity of all users in the park, K-means clustering is performed on the results. When the number of clusters is 3, the clustering effect is better. The clustering results of DR maturity are shown in Fig.5.
Finally, the proportion of users with different DR levels in the park to the total number of users in the park is counted. The sequence composed of the three proportional results contains the distribution structure characteristics of the DR level of the park users. As the park DR sequence, it clearly that most enterprises in the park have low DR maturity from the Figure 5.

**Table 6. DR Sequence.**

| Label | 0    | 1    | 2    |
|-------|------|------|------|
| Percentage | 0.8  | 0.188| 0.012|

4.4. Park Portrait Results

Due to the limited number of parks included in the example, in order to show the effectiveness of the algorithm more intuitively, the calculation results of the three characteristics of all users are drawn in this paper, and the characteristics of different parks are shown by user distribution, as shown in Fig.6.
Figure 6. Distribution of users in three characteristic dimensions of three parks.

As can be seen from Figure 6, most users in different parks have stable power consumption, but there is a certain difference in the distribution of users in the park. The users of Zhengzhou International Logistics Park have the characteristics of stable electricity consumption behavior, low maturity of power configuration demand, and poor DR capacity. The reason is that the logistics park has no complicated production lines, but has high energy-consuming equipment that requires long-term stable operation and needs protection. Compared with other parks, Zhengzhou High-tech Development Park has better DR capabilities. This is because park users are mostly software program development industries that is the service industry, they can bear more energy regulation and therefore have greater potential for DR regulation. The power configuration needs of users in the Jiaozuo Industrial Cluster are more mature. This is in line with the fact that Jiaozuo is an industrial park that is becoming mature. These analyses all prove that the park portrait here can be explained in combination with the actual situation, and a quantitative analysis of the park is completed.

5. Conclusion
In this paper, a multi-dimensional park portrait model based on clustering is proposed, and based on the calculation examples of the parks in Zhengzhou and Jiaozuo in Henan Province, the following conclusions are obtained: ① Different types of parks have different user distribution characteristics. But the energy consumption of users in the park is basically stable. ② Logistics parks have more stable energy consumption, lower DR maturity and lower power configuration maturity compared with other parks. ③ High-tech parks have higher DR capabilities.

The above conclusions reflect that the cluster-based multi-dimensional park portrait model can quantify the characteristics of multiple dimensions of the park, and the analysis results obtained have certain practical interpretation value. Therefore, power supply companies can implement differentiated and personalized power supply services based on the park portrait technology. In future, follow-up work will explore the correlation between the characteristics of the park and the power supply policy of the park, and make efforts to complete the integrated recommendation system.
Acknowledgments
This work was supported by Science and Technology Foundation of SGCC under Grant 5217L019000 6.

Reference
[1] AoD , Xing-Xing P , Liang J , et al. Energy Management Optimization Based on Energy Efficie
ncy Monitoring in Shanxi Industrial Park[J]. energy conservation technology, 2013.
[2] H. Si, H. Wu, L. Zhou, J. Wan, N. Xiong and J. Zhang, "An Industrial Analysis Technology Ab
out Occupational Adaptability and Association Rules in Social Networks," in IEEE Transacti
ons on Industrial Informatics, vol. 16, no. 3, pp. 1698-1707, March 2020.
[3] H. Jiang, Z. Hu, X. Zhao, L. Yang and Z. Yang, "Exploring the Users’ Preference Pattern of Ap
lication Services Between Different Mobile Phone Brands," in IEEE Transactions on Comp
utational Social Systems, vol. 5, no. 4, pp. 1163-1173, Dec. 2018.
[4] PAZZANIM J, BILLSUS D. Content-based recommendation system[M].Springer, 2007:325-34
1.
[5] LIUQ, NIUK. Microblog user interest modeling based on feature prop agation[C].Sixth Internati
onal Symposiumon Computational Intelligence Design, 2013:383-386.
[6] BLEI D M, NG A Y, JORDAN M I.Latentdirichletallocation[J].Journ al of Machine Learning R
esearch, 2003 (3) :993-1022.
[7] Liu Hai, Lu Hui and RuanJinhual, "Research on Precision Marketing Segmentation Model Base
d on Mining ‘Persona’", Silk, vol. 52, pp. 37-42, December 2015.
[8] T. Wang, J. Hu, C. Li and Z. Zhang, "Research on tariff recovery risks assessment method based
on electrical user portrait technology," 2017 6th International Conference on Computer Scie
nce and Network Technology (ICCSNT), Dalian, 2017, pp. 218-221.
[9] S. Lu, H. Jiang, G. Lin, X. Feng and Y. Li, "Research on Creating Multi-Attribute Power Consu
mption Behavior Portraits for Massive Users," 2018 8th InternationalConference on Power a
nd Energy Systems (ICPES), Colombo, Sri Lanka, 2018, pp. 54-59.
[10] C. Zhong, J. Shao, F. Zheng, K. Zhang, H. Lv and K. Li, "Research on Electricity Consumption
Behavior of Electric Power Users Based on Tag Technology and Clustering Algorithm," 201
8 5th International Conference on Information Science and Control Engineering (ICISCE), Z
hengzhou, 2018, pp. 459-462.
[11] KLUGER Y, BASRI R, ChANG J T, et al. Spectral biclustering of microarray data: coclustering
genres and conditions[J]. Genome Res, 2003,13(4):703-16.
[12] Q. Li, "Application of grey Verhulst model to commercial flights at the Macau International Air
port," Proceedings of 2013 IEEE International Conference on Grey systems and Intelligent S
ervices (GSIS), Macao, 2013, pp. 161-163.
[13] P. J. G. Lisboa et al., "Partial Logistic Artificial Neural Network for Competing Risks Regulariz
ed With Automatic Relevance Determination," in IEEE Transactions on Neural Networks, vo
l. 20, no. 9, pp. 1403-1416, Sept. 2009.
[14] C. Li, Z. Xu, Y. Zhou, W. Chen and X. Meng, "A Frequency Control Strategy Using Electric Ve
hicles Considering Travelling Demand," 2019 IEEE Innovative Smart Grid Technologies - A
sia (ISGT Asia), Chengdu, China, 2019, pp. 2709-2713.
[15] J. Wang, Y. Shi and Y. Zhou, "Intelligent Demand Response for Industrial Energy Management
Considering Thermostatically Controlled Loads and EVs," in IEEE Transactions on Industria
l Informatics, vol. 15, no. 6, pp. 3432-3442, June 2019.
[16] Y. Wang, L. Wu and S. Wang, "A Fully-Decentralized Consensus-Based ADMM Approach for
DC-OPF With Demand Response," in IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 26
37-2647, Nov. 2017.
[17] S. H. Tindemans, V. Trovato and G. Strbac, "Decentralized Control of Thermostatic Loads for F
lexible Demand Response," in IEEE Transactions on Control Systems Technology, vol. 23, n
o. 5, pp. 1685-1700, Sept. 2015.