

A method of electricity consumption behaviour clustering and pricing packages based on data mining

Baoyi Wang^{a,*}, Yang Fan^b and Shaomin Zhang^c

School of Control and Computer Engineering, North China Electric Power University, China
a. wangbaoyiqj@126.com, b. fywithnv@163.com, c. zhangshaomin@126.com

Keywords: smart grid, distributed clustering ,pricing packages recommendation

Abstract: In this paper, under the background of the reform of electricity sales side, a method of electricity consumption behaviour clustering and pricing packages based on data mining is proposed. Firstly, a distributed clustering framework combining DTW k-medoids algorithm and CFSFDP algorithm is proposed. Secondly, typical load curves of local data are extracted under the framework to construct local model. Then, quadratic clustering analysis is carried out for the local model results, and the global typical load curve is obtained to construct the global model. Finally, recommend the most suitable electricity sales plan to the target users. The experimental result shows that the subdivision of electricity consumption behavior can realize the effective personalized electricity package recommendation service for users and improve the power supply service quality for power companies and provide technical support for improving the operation efficiency.

1. Introduction

More than 500 million smart meters have been installed in China, and data collected every 15 minutes can add up to a terabyte of data a day. The surge of electricity consumption data of massive residential users has posed challenges to data storage, communication and analysis. On the one hand, electricity consumption data is collected and distributed in different locations, and users' electricity consumption data is collected and stored in different substations to which users belong. It is very expensive and time-consuming to transfer the whole data from each distributed site to a central site. On the other hand, the analysis and clustering of large data sets collected from each distributed site requires significant time and memory overhead. At present, there have been a lot of parallel clustering studies on the application of big data [1-2]. For these algorithms, the entire data set should be located in the same data center and then assigned to different clients, but it does not match the reality of power consumption data collection and storage. In addition, some completely distributed clustering algorithms are proposed to solve these problems by aggregating the information of local data and sending it to the central site for central analysis.

At present, there are many researches on power load clustering for users, but few of them are the application scenarios of providing value-added services to power companies. At home and abroad, the research on the electricity side mainly focuses on the reform of the electricity market, the construction of the electricity retail and wholesale market, and the bidding system. Literature [3-4] constructed the model of users' choice of electricity selling companies through psychological methods, and studied how to guide users to actively adjust their choice behavior through electricity price, so as to achieve a win-win situation between users and electricity selling companies. Power companies mine the massive data of users [5-7] and understand users' electricity consumption behavior, which is a necessary way to realize package push [8].

In summary, this paper proposes a resident user clustering and package recommendation method based on data mining for the selling power companies under the application background of the reform of the selling power side. Different electricity price packages are pushed to different types of users to save the electricity cost, so as to truly realize differentiated and refined targeted service of electricity selling companies.

2. DDKC

The algorithm is mainly divided into the following three steps: first, the local data scattered in different regions are locally clustered, and the local model is constructed; Secondly, the results of local clustering are further clustered and combined to obtain the results of global clustering. Finally, according to the result of global clustering the attribution of local data is adjusted.

2.1. Local clustering based on DTW k-medoids

DTW is a technique for finding the best alignment between two time-dependent sequences. In order to compare two sequences X and Y of length N , the total cost $c_p(X, Y)$ of a curved path p between X and Y is defined as follows, and the distance is calculated by $d(x_n, y_m) = (x_n - y_m)^2$:

$$c_p(X, Y) = \sum_{l=1}^L d(x_{nl}, y_{ml}) \quad (1)$$

The best warping path p^* is the path with the lowest cost of all possible paths. The DTW distance between X and Y is defined as the total cost of DTW(X, Y):

$$DTW(X, Y) = c_p^*(X, Y) \quad (2)$$

$p^* = \arg \min c_p(X, Y)$, Given a set of family load curve L and the number of clusters K , the goal of the DTW based clustering algorithm is to find an allocation C_k for each cluster K , and minimize WC in the cluster and cluster prototype matrix μ_K , as defined below:

$$WC = \sum_{k=1}^K \sum_{X \in C_k} DTW(X, \mu_k) \quad (3)$$

In the allocation step, we assign each load curve to the nearest cluster based on DTW(X, Y). After the find all load curve of the cluster members, we update the cluster prototypes, μ_k^* is when $X \in C_k$ the minimum DTW(X, μ_k), this algorithm is equivalent to k-medoids clustering using DTW as distance measurement.

2.2. Global clustering based on modified CFSFDP

CFSFDP clustering algorithm is a kind of clustering algorithm based on density, due to its simple parameters setting, beautiful form and is widely used in a short period of time, its core idea is to use the relative distance δ_i and local density ρ_i to determine the density of the center, and then to other clustering objects on a distribution is completed. Each clustering object of the original CFSFDP algorithm is equal. In order to fully consider the typicality of each load curve in the calculation of density, the ω_j is introduced as follows:

$$\rho_i = \sum_{j=1}^N \omega_j \chi(D_{ij} - d_c) \quad (4)$$

ω_j is the weight of each generation of clusters in M_i , is the weight of load curve j , which is equal to the load of the class in the local data center.

3. Results of lustering and package recommendation

3.1. Experiment design

3.1.1. Data set

The data set used for example analysis in this paper comes from the actual measurement data of Irish smart electricity meters published by SEAI. The experiment was conducted with the electricity consumption data of 1,443 household users for 22 days, and 48 data points were collected for each household user every day.

3.1.2. Package selection

Considering that there are many kinds of electricity price packages for domestic and foreign electricity selling companies, this paper selects two types of electricity price packages at home and abroad for experiments, and designs four kinds of electricity price packages.

Table 1: Electricity charge package based on monthly accumulated electricity consumption.

Electricity packages	Accumulated monthly electricity consumption (kW·h)	Electricity prices (prices /kW·h)
Package 1	[0,208]	0.5469
Package 2	[208-375]	0.5969
Package 3	>375	0.8469

Table 2: Electricity package 4.

Period of time		Electricity prices (prices /kW·h)
The peak time	8:00am-10:00pm	0.5963
The ordinary times	other times	0.3153

Table 3: Variable rate electricity package 5.

Period of time		Electricity prices (prices /kW·h)
The peak time	6:00am-10:00am 6:00pm-10:00pm	0.646
The ordinary times	other times	0.169

Table 4: Variable rate electricity package 6.

Period of time		Electricity prices (prices /kW·h)
The peak time	12:00am-6:00pm	1.56
The ordinary times	At other times	0.44

3.2.Data processing

In the process of data cleaning, the specific steps of verifying and normalizing each original daily load curve of each local data include: deleting the original daily load curve containing the blank value load collection point. Through data processing, 31202 load curves were obtained.

3.3.Clustering results and analysis

The power load curve after processing is evenly divided into 10 local data, each local data contains about 3000 load curves. Based on the existing DTW clustering algorithm of MATLAB, the example is invoked and rewritten in the local clustering stage. Figure 1 shows the number of clustering centers obtained by constantly adjusting the size of ∂ . With the increase of ∂ , the number of local clustering centers decreases. $\partial=0.4$ and the number of clustering centers contained in each local data is: 882, 757, 735, 701, 765, 589, 573, 775, 816, 782. Then, in the global analysis, the total number of load curves analyzed is 7375, which is equivalent to reducing the clustering workload of load curves by about 80%.

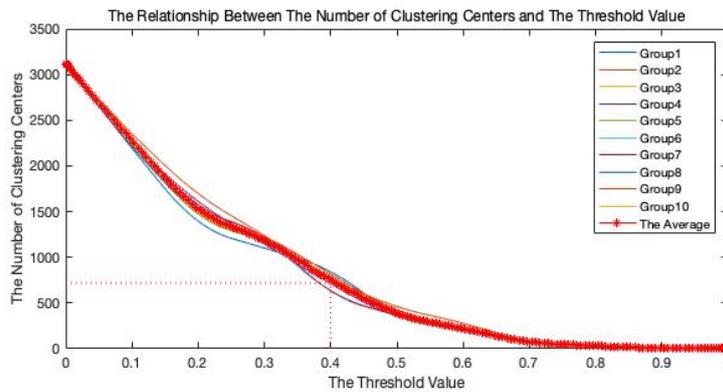


Figure 1: The relationship between the number of clustering centers and the threshold value.

According to the local clustering results, the weighted CFSFDP was used for global clustering, as shown in Figure 2, the final global clustering number $M=5$.

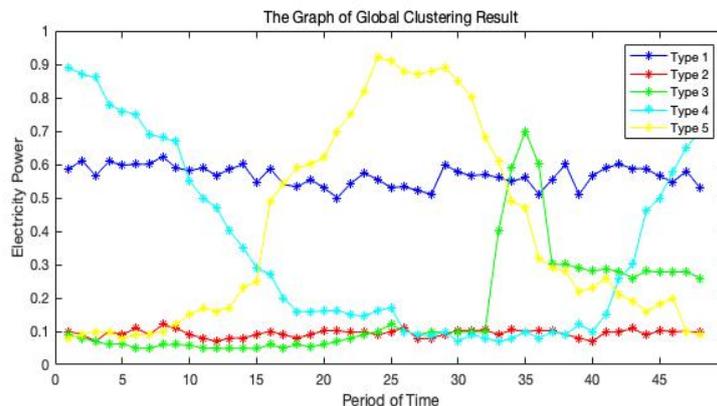


Figure 2: The graph of global clustering result.

Figure 2 shows the curve of all eventually find the typical electricity clustering model, it can be seen that residents users of electricity mode are diverse, the next will be based on the clustering results and the analysis of the power package together for each subclass users choose pricing packages.

3.4. Package recommendation results and analysis

3.4.1. Preliminary analysis of user patterns

This paper USES the concept of information entropy to measure the stability of user behavior. For a user n , assume that there are 5 possible values of load data L in a certain period of time, namely C_1, C_2, C_3, C_4, C_5 and the probability of each value is P_1, P_2, P_3, P_4 and P_5 , then the entropy value S_n of load data L in that period can be defined as:

$$S_n = -\sum_{i=1}^k P_i \log P_i \quad (5)$$

The entropy value reflects the fluctuation degree of the curve, the higher the entropy value, the greater the curve fluctuation degree, the entropy histogram of all resident users is shown in figure 3.

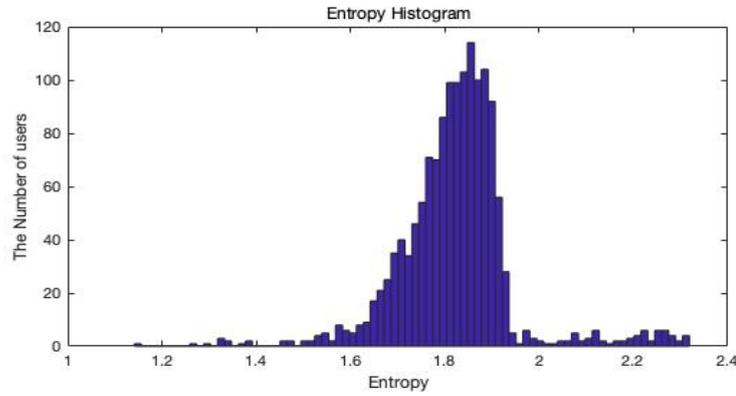


Figure 3: The graph of entropy histogram.

It can be seen that the entropy value of different types of users is different. Therefore, this paper will analyze the change degree of power consumption behavior of users according to the entropy value of users. If a family's entropy quantile over 65%, it can be classified as a variable family, otherwise is classified as a stable family.

3.4.2. Recommendation of package

According to the monthly cumulative consumption of various users, the monthly cumulative consumption of users will be respectively corresponding to the electricity price packages of each grade in the packages 1-3. Therefore, this paper will extract the monthly cumulative consumption Q_{month} of users as the characteristic attribute of the second clustering of users in the variable electricity consumption type.

Since the number of classified users has been greatly reduced compared with the initial data set, and there is only one characteristic quantity Q_{month} . For electricity user patterns, considering the type 1 and type 2 of consumer behavior had no obvious electricity peaks and troughs, therefore the package is recommended according to the accumulated monthly electricity consumption Q_{month} ; Other types will be recommended as the most suitable package for users of this subclass according to the rate type with the least overlap between the typical load curve and the peak of electricity consumption curve and the peak of electricity price package, the final recommendation results are shown in the table:

Table 5: The recommendation of package by accumulated monthly electricity consumption.

The type of users	Accumulated monthly electricity consumption (kW·h)	The number of users	Recommendation
Type 1	[0,208]	131	Package 1
Type 2	[208-375]	20	Package 2
Type 3	>375	16	Package 3

Table 6: The recommendation of package by the type of electricity behaviour.

The type of electricity behavior	The number of users	Recommendation
Type 3	428	Package 6
Type 4	369	Package 4
Type 5	479	Package 5

As can be seen from the table that users of all subclasses are recommended the time-of-use price packages that are out of position between the high electricity price period and the peak period of their own electricity consumption, we can according to the corresponding relationship between the total monthly electricity consumption of users and the package to choose the package which price is lowest when the electricity price is calculated.

4. Conclusion

In this paper, under the background of the reform of electricity sales, through the massive amounts of electricity, data mining is constructed under the electricity market reform residential user behavior analysis model, this paper proposes a distributed clustering algorithm DDKC, and on this basis, the behavior of electricity to users in the subdivision, according to the power mode, the classification of a residential user power package recommended method is given. It is helpful to optimize the energy structure of residents and ensure the stability of power system load by analyzing and studying the power consumption behavior of users to recommend personalized packages.

References

- [1] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data[J]. *IEEE Transactions on Smart Grid*, VOL. 7, NO. 1, Pages: 136-144, 2016.
- [2] Ofuji K, Tatsumi N. Wholesale and retail electricity markets in Japan: Results of market revitalization measures and prospects for the current reform[J]. *Economics of Energy & Environmental Policy*, VOL. 5, NO. 1, Pages: 31-49 2016
- [3] Sun Yi, Gu Wei, et al. Electricity-Retail-Market-Reform Oriented Recommendation Method of User's Hierarchical Clustering and Pricing Packages [J]. *Grid technology*, Volume:2, NO. 2, Pages: 447-454, 2018.
- [4] Hu Chen, et al. Preliminary Research of Trading Approach and Management Modes of Electricity Retail Companies Under New Electricity Market Reform [J]. *Grid technology*, Volume:40, NO. 11, Pages: 3293-3299, 2016.
- [5] Luo Shuhan, Jiang Chuanwen, et al. Power Trading Strategy and Risk Assessment of Electricity Retailing Company Under Power System Reform [J]. *Grid technology*, Volume:43, NO. 4, Pages: 944-951, 2019.
- [6] F. Luo et al. Cloud-based information infrastructure for next-generation power grid: Conception, architecture, and applications [J]. *IEEE Transactions on Smart Grid*, VOL. 7, NO.1, Pages: 1896-1912, 2015.
- [7] Luo F, Ranzi G, Wang X, et al. Social Information Filtering Based Electricity Retail Plan Recommender System for Smart Grid End Users [J]. *IEEE Transactions on Smart Grid*, VOL. 10, NO. 1, Pages: 95-104, 2019.
- [8] Yi Wang, Qixin Chen, Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications [J]. *IEEE Transactions on Smart Grid*, 2016, 7(5):2437-2447.