

Orderly power utilization oriented twice clustering load data analysis

X. CHEN¹ & K. SHI² & G.J. XU² & D.Z.LI² & P. XU³ & H.YE³

¹Jiangsu Electric Power Company Research Institute. Nanjing 211103, China; State Grid Key Laboratory of Electric Power Metrology, Nanjing 211103, China

²Electric Power Research Institute. Beijing, China

³North China Electric Power University, Beijing, China

KEYWORD: Orderly power utilization; Twice clustering algorithm; Load data analysis; Power load data clustering

ABSTRACT: As the requirement of smart grid increasing, a twice clustering algorithm is proposed here to extract the characteristic of load data curve. In addition, the calculating algorithm of load performance is also provided to analyze the curves, including seasonal maximum load curve, seasonal safe load curve, economic production load curve, summer (winter) air conditioning load and production protected load curve, upon which the adjustable potential of user can be estimated, so that to set up the economic estimation for companies. Moreover, the above processing method is finally verified through the experiment design upon the data resources from an industry company in Jiangyin.

Introduction

As an important part of demand side management, orderly electricity utilization is to ensure the maintenance of the smooth power supply through administrative measures, economic means and technical methods, in case of insufficient power supply and emergency^[1]. In recent years, along with country economy sustainable development, people's living standard constantly improving, power consumption grows sharply and the power supply is relatively lagging. Most regions is in serious lack of electricity in summer and winter. Power companies everywhere was rationing by strengthening the demand side management to reduce conspicuous power brownouts. Therefore, the orderly management of power is an imminent task for the power companies, even if the growth of power supply satisfied the power consumption in future, orderly management of the electricity is still a very important task^[2].

The main content of this paper is to lay a foundation for the orderly electricity comprehensive evaluation, proposing user orderly power utilization value and power saving potential evaluation methods^[3]. On the basis of regional and industry classification, clustering analysis method is used for the user load data of different load characteristics to achieve accurate classification. Then the investigation and analysis on the typical user were extracted from each type of user to set up orderly power consumption value and electricity saving potential evaluation method. In addition, propose principles and strategy for user participation in power ordering on the basis of calculated potential value, which can be open to the society as a standard for the plan making and execution for order electricity, which does help to the optimal response plan making and reducing the blindness of user operations^[4].

General structure

To achieve considerable strategy for orderly power utilization, the paper mainly researches as the following general structure.

(1) Data preprocessing: to obtain user's 48 history data points through the marketing system and remove anomalies or invalid data, in preparation for the clustering analysis algorithm;

(2) Clustering analysis algorithm: the history of the user data is divided into four groups according to the season, do clustering algorithm process respectively for every quarter of the data, at the same time every quarter is clustered into 6 classes, copolymerization into 4 season 24 root curve;

(3) User load feature extraction: to extract the user load curve characteristic indexes as follows.

1) Seasonal maximum load curve: load curve with highest average power consumption of the quarter;

2) Seasonal safe load curve: load curve with lowest average power consumption of the quarter;

- 3) Economic production load curve: the most intensely occurred center curve of every quarter as a typical curve, and define the curve nearest to it as economic production load curve.
 - 4) Summer (winter) air conditioning load: the difference value between summer (winter) maximum load curve and spring season (autumn) maximum load curve, if the difference is a non-positive number, the take that point data of 0.
 - 5) Production protected load curve: the difference value between Seasonal maximum load curve and air conditioning load.
- (4) Evaluation of user electricity potential (interruptible) saving capacity:
- 1) User maximum electricity saving potential: the difference value between Seasonal maximum load curve and safe load curve.
 - 2) User economical electricity saving potential: the difference value between Seasonal maximum load curve and economic production load curve.
 - 3) User productive electricity saving potential: the difference value between Seasonal maximum load curve and production protected load curve.
- (5) Economic value index calculation: gathering the evaluation index of company economic value according to consumption on per unit tax, power consumption on per unit of GDP or energy consumption on per unit GDP.
- (6) Result output: output the user quarterly electricity saving potential and economic value index data as results forming the enterprise energy saving potential database, to provide a basis for orderly electricity schedule.

Twice clustering algorithm

Due to the huge amounts of sampling data and characteristic vectors, unique traditional clustering method cannot reach the ideal effect^[5]. Therefore, a new load clustering algorithm is necessary to suit the large sampling data and high dimensions, in order to achieve a more accurate and efficient classification^[6]. Observing the performance of clustering algorithms, it is obvious that the initial clustering center has great effect on the classification for fuzzy C clustering. While the hierarchical clustering method is an algorithm with simple process, quick clustering and no-initialization but also with much repeated step. Considering the two methods generally, a twice clustering method is proposed to study the load classification. That means firstly do a clustering with hierarchical clustering method and then a fuzzy C clustering with the classification results in first step as the clustering center of fuzzy C. In that way, the sensitivity of Fuzzy C to the initial center is largely reduced. The detail process flow of twice clustering algorithm is as following fig. 1.

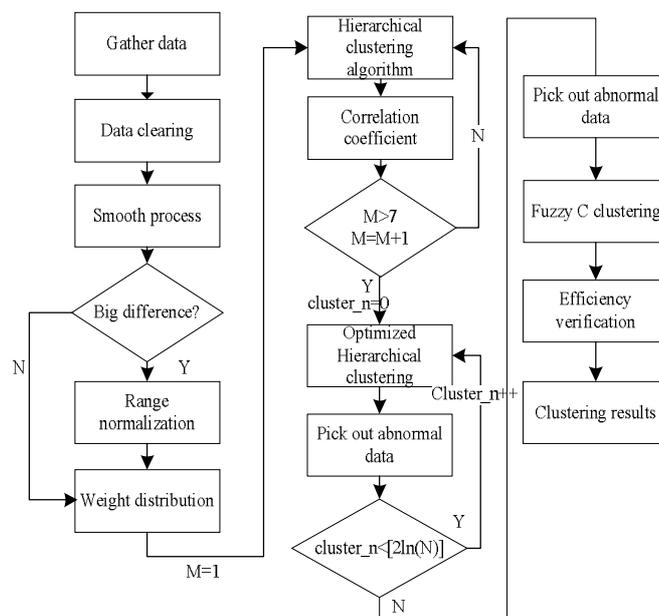


Fig. 1 Detail process flow of twice clustering algorithm

Preprocessing of load data

Abnormal data identification and process

Load data is often not comprehensive collected or distorted due to the signal interference, software failure, equipment performance and so on. Thus, the transverse recognition is used to identify and handle the abnormal data^[7].

A common agreement here is that the load data in a period is probably similar, namely the load curve of sampled day and near similar days are kindly the same. Thus, the abnormal data can be recognized by comparison of sampled data and anticipated result.

Step 1 calculate the mean and variance of data sequence.

$$\bar{x}_{n,i} = \frac{1}{N} \sum_{n=1}^N x_{n,i}, i=1 \sim 96 \quad (1)$$

$$S_i^2 = \frac{1}{N} \sum_{n=1}^N (x_{n,i} - \bar{x}_{n,i})^2 \quad (2)$$

Step 2 recognize the abnormal data with threshold (commonly 1-1.5).

$$\left| x_{n,i} - \bar{x}_{n,i} \right| > 3S_i e \quad (3)$$

Step 3 correct the data using weight distribution.

$$x_{n,i}^* = \frac{a_1}{2} \sum x_{n\pm 1,i} + \frac{b_1}{2} \sum x_{n,i}^{1,2} + g_1 \bar{x}_{n,i} \quad (4)$$

Here $x_{n,i}$ is the abnormal data. $a_1 + b_1 + g_1 = 1$, $x_{n,i}^*$ is the corrected data and $x_{n,i}^{1,2}$ is the nearest similar load data from sample.

Range normalization

To obtain a standardized data sequence, an effective method is the range normalization.

The $\min_{1 \leq k \leq p} x_{ik}$ here is minimum value of X_i , and X^R is the normalized matrix.

$$X^R = \begin{bmatrix} x_{11}^R & x_{12}^R & \mathbf{L} & x_{1p}^R \\ x_{21}^R & x_{22}^R & \mathbf{L} & x_{2p}^R \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ x_{n1}^R & x_{n2}^R & \mathbf{L} & x_{np}^R \end{bmatrix} \quad (5)$$

$$x_{ij}^R = \frac{x_{ij} - \min_{1 \leq k \leq p} x_{ik}}{\max_{1 \leq k \leq p} x_{ik} - \min_{1 \leq k \leq p} x_{ik}}, i=1, 2, \dots, n; j=1, 2, \dots, p \quad (6)$$

Weight distribution of load data

In the normal clustering algorithm, each performance character vector has the same effect weight, which doesn't suit the real situation that the different factor has different effects^[8]. Moreover, the effect weight also varies with the user type, conditions and so on. Thus weight distribution is needed to overcome the limitations of direct clustering analysis.

Considering the reference policies and real load characteristics, the TOU (time of use) price is a typical factor to which the load data has great relation. Power price with different weight may illustrate the relationship more accurately. Generally, set a high weight to the working period while a lower for the night rest period.

For instance, the weight of period with high load ratio (8:00-12:00, 17:00-21:00) here is 3, and that of rest period is as default value 1. In addition, this method is also suitable for other vectors to reflect their true effectiveness to the load data.

Hierarchical clustering algorithm

The principle of hierarchical clustering algorithm is: at the start each of n variables as a cluster, and the distance between the variables and the distance between the classes are ruled. Then nearest two merged into a new cluster, refreshing the distance between new cluster and other clusters. Finally, repeat the procedure until all the variables into a class. After finishing the process, a relationship graph is gathered which shows the clustering results clearly. However the distance calculation method is defined independently, different definition results in different algorithms. In this paper, Ward

clustering is used as its clear great clustering effect with clear concept and simple calculation. The implementation steps are as following^[9].

Step 1 Form the sample feature vector set;

Step 2 Standardize the original data, transforming the load characteristic composition data into a [0,1] interval data;

Step 3 Initial classification, set $K=0$, each sampled data as a cluster:

$$G_i^{(0)} = \{x_i\}, \quad i = 1, 2, \dots, \mathbf{L}, N \quad (7)$$

Step 4 Calculate the distance among clusters, generating a distance matrix $D^{(K)} = (D_{ij})_{n \times n}$:

Step 5 Select two clusters according to the square sum of deviations S as the clusters being merged.

$$S = \sum_{i=1}^k \sum_{p=1}^{n_p} (x_{ip} - \bar{x}_p) (x_{ip} - \bar{x}_p) \quad (8)$$

And the new distance between merged clusters and others can be represented as following:

$$d_{kr}^2 = \frac{n_k + n_p}{n_r + n_k} d_{kp}^2 + \frac{n_k + n_q}{n_r + n_k} d_{kq}^2 - \frac{n_k}{n_r + n_k} d_{pq}^2 \quad (9)$$

Here the n_p, n_q, n_r, n_k represent the sample numbers of cluster G_p, G_q, G_r, G_k .

Step 6 Repeat the step 4-5 until all sample clustered into a class.

Fuzzy C clustering

Use the equation editor of the selected word pro The main principle of FCM algorithm is to iterate and adjust (U,V) to get a minimum objective function J. The detailed steps are as following^[10]:

Step 1 Set expected cluster numbers C , fuzzy index m and initial clustering center v_0^L ;

Step 2 Calculate the distance matrix D ;

Step 3 Calculate the membership matrix U^L according to D ;

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}^2}{d_{kj}^2}\right)^{1/m-1}} \quad (10)$$

Step 4 Calculate the clustering center V^{L+1} according to U^L :

$$v_i^{L+1} = \frac{\sum_{j=1}^n (u_{ij}^L)^m x_j}{\sum_{j=1}^n u_{ij}^L (k)^m} \quad (11)$$

Step 5 Make a judgment of whether $\|U^{L+1} - U^L\| < e$, if it sets up, the process break up, otherwise back to step 2 and goes on.

About the expected cluster numbers C , its value haven't had an accurate setting. While Pal and Bezdek point out that the maximum limitation is that $c_{\max} \leq \sqrt{n}$, as \sqrt{n} increases much faster than $\ln n$, we set $c_{\max} \leq 2 \ln n$.

About the initial clustering center V , it is well known that the clustering result is much sensitive to initial center. Not only has the classification varied largely with different center initialization, an unexpected minimum point may also appears to lead to a slow constriction even an endless loop. Thus, the special factors are extracted to avoid that.

Experiment Analysis

To verify the efficiency of the proposed algorithm, an experiment is designed with the main principle with 6 steps in part 2, which is divided into four main parts as following:

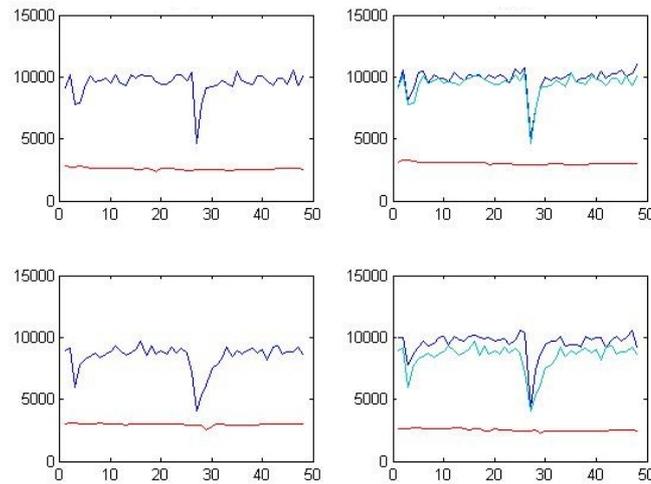
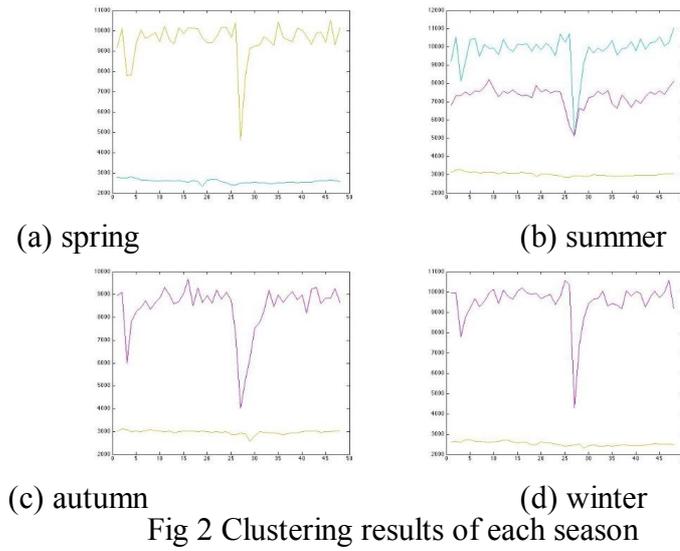
(1) Data processing: process the load data with twice clustering algorithm, getting the clustered load curves of each season.

(2) Calculate the seasonal maximum load curve, seasonal safe load curve, economic production load curve, summer (winter) air conditioning load and production protected load curve.

(3) Estimate the seasonal power saving potential, including user maximum electricity saving potential, user economical electricity saving potential and user productive electricity saving potential.

(4) Gather an economic evaluation along the power consumption on per unit tax.

The experiment results are shown in the Fig.2-3 and Tab.1.



— Maximum load curve of the season
— safe load curve and economic production load of the season
— production protected load curve of the season

Fig.3 load feature extraction curves of each season

Based on the above results, a refined analysis of the power load adjustment potential in each season is made to support the electric power distribution and dispatch. The detailed analysis results is as Tab.1, from which it can be seen that the adjustment potential goes to the peak mostly in summer or winter, since the air conditioner spends lots of power resources and its adjustable potential is considerable. However, also in the spring and autumn, if the production is properly controlled, the transferable load is also handsome, especially for the companies works at the peak point of power system load.

Tab.1 Load adjusting potential of each season

Season	SML	SSL	EPL	PPL	MESP	EESP	PESP
Spring	9524.559	2578.184	2578.184	/	6946.375	6946.375	/
Summer	9834.319	3030.155	3030.155	9450.562	6804.164	6804.164	383.7576
Autumn	8450.803	2973.843	2973.843	/	5476.96	5476.96	/
Winter	9531.716	2525.78	2525.78	8450.803	7005.936	7005.936	1080.913

Conclusion

In this paper, proposed a twice clustering algorithm to extract the characteristic of load data curve. In addition, provided the calculating method of load performance analyzing curves, including seasonal maximum load curve, seasonal safe load curve, economic production load curve, summer (winter) air conditioning load and production protected load curve, upon which the power adjustable potential can be estimated. Then the economic estimation can be set up for companies. Moreover, the above processing method is verified through the experiment design upon the general structure of this paper and data resources from an industry company in Jiangyin. Submission of material to the editor

Reference

- [1] ZHANG Zhi-sheng, SUN Ya-ming, ZHANG Shi-ying et al. Clustering Analysis of Electric Load Series Using Clustering Algorithm of Multi-Hierarchy and Detailed Decomposition Based on Data Mining[J]. Power System Technology, 2006, 30(2): 51-56.
- [2] HUANG Mei, HE Ren-mu, YANG Shao-bing. APPLication of Fuzzy Clustering in Measurement-Based Load Modeling[J]. Power System Technology, 2006, 30(14): 49-52.
- [3] ZHANG Chun-lei. Power system short-term load forecasting based on fuzzy clustering analysis and rough sets[J]. Journal of North China Electric Power University, 2008, 35(3): 38-43.
- [4] SHI Jian-zhong. Research on Nonlinear System Identification Using Fuzzy Clustering[D]. North China Electric Power University. 2012.
- [5] YU Long. Short-term Power Load Forecasting Based on Fuzzy Clustering similar days[D]. Shanghai Jiao Tong University.
- [6] LI Pei-qiang, LI Xin-ran, CHEN Hui-hua, et al. The characteristics classification and synthesis of power load based on fuzzy clustering[J]. Proceedings of the CSEE, 2005, 25(24): 73-78.
- [7] Li Cai-ling, Wang Jin, Li Xin-ran. An Optimized FCM Method for Electric Load Clustering[C]. DRPT2008, 6-9 April 2008, Nanjing China. 2008: 882-886.
- [8] SHI Guo-ping, LIANG Jun, LIU Xiang-sheng. Load Clustering and Synthetic Modeling Based On An Improved Fuzzy C Means Clustering Algorithm[C]. DRPT2008 6-9 July 2011:859-865.
- [9] Lin Yu-Hsiu, Men-Shen Tsai. Non-Intrusive Load Monitoring by Novel Neuro-Fuzzy Classification Considering Uncertainties[J]. IEEE Transactions on Smart Grid, 2014, 5(5): 2376-2384.
- [10] Kamwa, I. Pradhan, A.K. Joos, G. et al. Fuzzy Partitioning of a Real Power System for Dynamic Vulnerability Assessment[J].IEEE Transactions on power systems.24(3):1356-1365.