Refined Analysis of User Load Based on Weighted Fuzzy Clustering

J.FAN[1,2] & X.L.GAO[1,2] & X. CHEN[1,2] ,K. SHI[3] ,P. XU[4] & H.YE[4]

[1]Jiangsu Electric Power Company Research Institute. Nanjing 211103, China;

[2]State Grid Key Laboratory of Electric Power Metrology, Nanjing 211103, China

[3]Electric Power Research Institute. Beijing, China

[4]North China Electric Power University, Beijing,China

KEYWORD: Refined Analysis; Weighted Fuzzy Clustering; Load Classification Dividing; Clustering Efficiency Estimating

ABSTRACT: As the development of intelligent power system, much importance and requirement have been attached to the load data analysis. In view of the current rough classification of load data, propose a weighted fuzzy clustering algorithm to detail the load classification dividing, which adds a weight distribution process to balance the different influence of various factors. In addition, Two group of experiments are set to verify the efficiency of this method. The experiment results show that the algorithm is effective to accurately cluster the load data and supportive to the fine analysis of load data.

Introduction

Influenced by various factors, power system load changes constantly with time. Also, it has the characteristics of continuity and periodicity[1]. Power system load characteristic is refers to the characteristics and nature of power load, different types of users load show different load characteristics. To master the characteristics and variation law of power load can be beneficial to safe and steady operation of electric power system and get the best economic benefit for power supply departments. Besides, it can give full play to the benefit of every once electricity for the user[2].

Power system load has strong periodicity as well as randomness[3]. On one hand, development trend of power load changes obeying a certain regulation; on the other hand load, certain fluctuations may occur at any moment with society, economy, politics, weather, and many other complicated factors. Load analysis and load distribution management is the foundation of economic analysis of power distribution system. Under the control environment, load analysis mainly serves for load forecasting, system planning and demand side management providing information[4]. While in the retail market, which is customer-centric, service strategy, it is more vital to make a refined analysis on the user load characteristic, to help fine mining potential value of electricity characteristics[5].

Index of user load performance

At present, according to the historical development and current situation in our country a set of relatively comprehensive load characteristic index system has been put forward[6]. The index system contains several indexes involving different time ranges from daily to year. The representation mainly divides into numerical type and curve type. Some indexes reflect the overall situation of load characteristics, mainly applied on horizontal comparison between different regions at home and abroad. The others are specific load index, mainly applied on the analysis and calculation of power system planning and design[7]. Considering practicality this article uses the index system in table 1.

However, the utilization of power indexes is still not enough in our country. For instance, the user load usually is roughly divided into several types: load of industry, agriculture, commercial and city life, which just considered the difference between vocations but not accurate indexes. To achieve fine management of user-side, the rough classification could not meet the sharply developing load mining requirement. For instance, user load varies not only with the vocations. Much more issues should be taken into consideration, since even the users of the same profession. Their load performance also may be different from each other due to the weather and personal habits[8]. Therefore, an accurate

and reasonable analysis of power load data needs the detailed classification with more considered issues and more effective algorithms.

Tab.1 Indexes of load data

Indexes of description	Indexes of comparison	Indexes of curves
Maximum/minimum load daily	Minimum load ratio daily	Curve of daily load
Average load daily	Average load ratio daily	Curve of year load
Difference of daily load	Load ratio difference daily	
Maximum/minimum load monthly	Load ratio monthly	
Average load daily monthly	Average load ratio monthly	
Maximum/minimum load yearly	Minimum load ratio monthly	
	Average day load ratio yearly	
	Average month load ratio yearly	
Difference of year load	Load ratio seasonal	
	Difference ratio yearly	
	Load ratio yearly	
	Peak load points of year	

Weighted fuzzy clustering

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.

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.

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

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

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

$$\left| x_{n,,i} - \bar{x}_{n,i} \right| > 3s_i e \tag{3}$$

Step 3 correct the data using weight distribution.

$$x^*_{n,,i} = \frac{a_1}{2}\sum x_{n\pm1,,i} + \frac{b_1}{2}\sum x^{1,2}_{n,,i} + g_1 \bar{x}_{n,i} \tag{4}$$

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

Range normalization

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

$$X^R = \begin{bmatrix} x^R_{11} & x^R_{12} & \mathbf{L} & x^R_{1p} \\ x^R_{21} & x^R_{22} & \mathbf{L} & x^R_{2p} \\ \mathbf{M} & \mathbf{M} & & \mathbf{M} \\ x^R_{n1} & x^R_{n2} & \mathbf{L} & x^R_{np} \end{bmatrix} \tag{5}$$

$$x^R_{ij} = \frac{x_{ij} - \min\limits_{1\le k\le p} x_{ik}}{\max\limits_{1\le k\le p} x_{ik} - \min\limits_{1\le k\le p} x_{ik}}, i = 1,2,...,n; j = 1,2,...,p \tag{6}$$

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

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. 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.

Fuzzy clustering process

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[9]:

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.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}^2}{d_{kj}^2})^{1/m-1}} \tag{7}$$

Step 4 Calculate the clustering center V_{L+1} according to U_L.

$$v_i^{L+1} = \frac{\sum_{j=1}^{n} (u_{ij}^{\ L})^m x_j}{\sum_{j=1}^{n} u_{ij}^L (k)^m} \tag{8}$$

Step 5 Make a judgment of whether $\left\| U^{L+1} - U^L \right\| < e$, 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 $c_{\max} \le \sqrt{n}$, as $\sqrt{n}$ increases much faster than $\ln n$, we set $c_{\max} \le 2 \ln n$.

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.

Clustering efficiency estimating

A probability distribution based clustering efficiency function is applied here as the evidence of clustering result estimating[10]. The function mainly achieves the judgment via probability distribution function $F(U,c)$ and coefficient $P(U,c)$. For each sampled X_i, the probability distribution function $F(U,c)$ and coefficient $P(U,c)$ is as following[11].

$$F(U;c) = \frac{1}{n} \sum_{j=1}^{n} (\sum_{i=1}^{c} m_{ij}^2 / \sum_{i=1}^{c} m_{ij}) \tag{9}$$

$$P(U;c) = \frac{1}{c} \sum_{i=1}^{c} (\sum_{j=1}^{n} m_{ij}^2 / \sum_{j=1}^{n} m_{ij}) \tag{10}$$

For the given U and c, clustering efficiency function is as following (11), and the optimal result appears when it suits the condition of equation (12).

$$FP(U;c) = F(U;c) - P(U;c) \tag{11}$$

$$FP(U^*;c^*) = \min_{c}\{\min_{\Omega_c} FP(U;c)\} \tag{12}$$

Simulation and fine analysis

To study the load performance of a user with different situation, select an electronic component produce company as object, sampling its load data of April, June to September, 2010 with 24 points daily, totally 152 sets of data with 130 of them are available. The clustering result is as Fig.1.

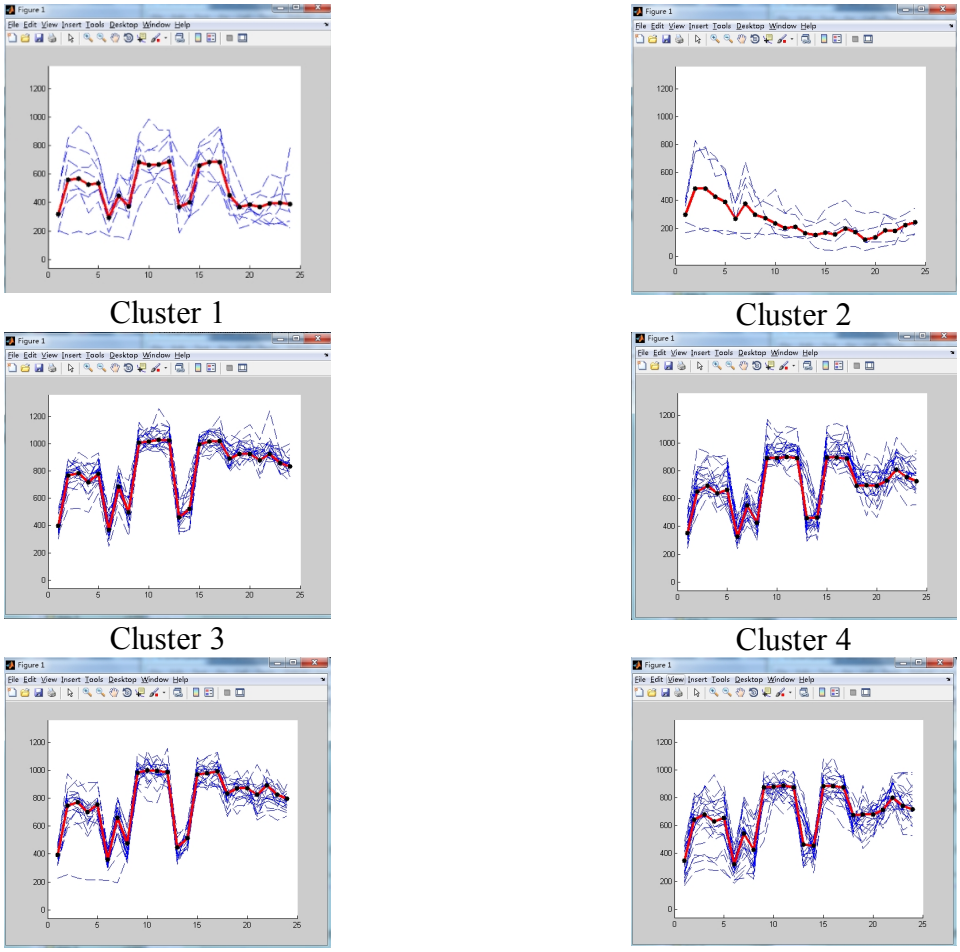

Figure 1. Clustering results of specific company

Tab.2 refined analysis of load utilization from one company

Industry type		Electronic manufacture
Load performance		Oscillation when folk shift
Peak point		2, 8-11, 14-16
Adjustable load for climate		Little related to climate
Adjustable load for production	Level 1	200-300 kW
	Level 2	500-600 kW
Suitable demand response strategy		TOU price, critical peak price , interruptible load, emergency demand response strategy

Integrating the above load clustering results, it can be known that the company works in three shift system. However, there exists a temporary oscillation when the folk hand over. That means, part of the production devices are not necessary to work constantly. So we can guess that when power grid is overloaded, these devices can be rest to lower the load. For each class, the first two classes

seem special to the others. According to their load values, class 1 and 2 seems to be in low production situation. Under normal production situation, the top load ranges from 800 kW to 1200 kW, not varying as the season shifts. There exist almost 200-300 kW reasonable load and 500-600 kW safe load available to adjust the load level. Therefore, we can predict that the company would be willing to transfer a portion of load into night period as the price motivation. In summary, the detailed analysis of user power consumption behaviors is as the following Tab.2.

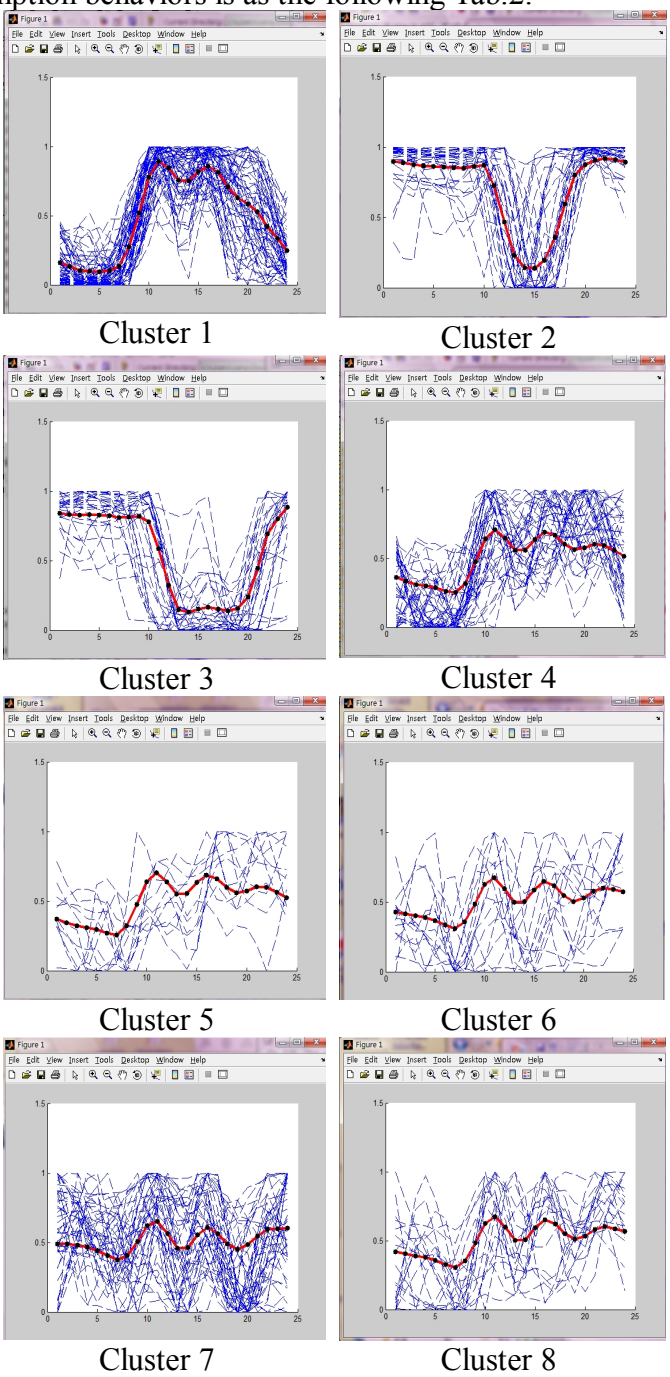

Fig.2 Clustering results of different industries

Moreover, a set of load data from different companies at the same time are also analyzed to study the load difference among different industries. Here selected 386 companies as the sample with 24 points a day each. The temperature is 15℃ and 264 data sets are available. The simulation result is shown in Fig.2.

From the above cluster results, it can be seen that the first class represents the typical companies which products in the day and rests at night, containing the 27% of all companies, such as comprehensive retail and tourism hotel industry. In view of their power transferring ability, the lighting and air conditioner are the main available devices to adjust power utilization. Therefore, critical peak

pricing and direct load control may be effective to them but with limited ability. While the second and third classes is reversed, which work at night and rest in day. It means that these companies have adapted some strategies to avoid the top price period. The companies of these classes, such as cement manufacturing, textile industry, non-metal manufacturing and so on, have little potential to be exploited. The other companies mainly work in triple shift schedule, which constantly works with 24 hours, however the production in the day is still a little higher. If their production could be changed, the adjustable potential are considerable. The refined analysis of load utilization is as the following tables.

Tab.3 indexes of load data performance from different classifications

Classification	Daily load ratio	Minimum load ratio daily	Difference of load daily	Peak hour	Valley hour
1	54.1%	10.6%	89.3%	10	4
2	76.2%	14.9%	85.2%	21	14
3	62.2%	14.6%	85.4%	23	13
4	70.8%	35.5%	64.4%	10	6
5	71.3%	36.6%	63.4%	10	6
6	75.3%	46.3%	53.7%	10	6
7	78.6%	57.7%	42.2%	10	6
8	74.8%	45%	54.9%	10	6

Tab.4 refined analysis of load data from different vocations

Vocation	Classification	Ratio	Load type	Suitable demand-response strategy
Business	1	61%	Peak forward	Critical Peak Pricing, direct control
	4	29%	Constant	TOU price/Critical Peak Pricing/direct control
	8	7%	Constant	TOU price/Critical Peak Pricing/direct control
Electronic production	1	35%	Peak forward	TOU price/interruptible load
	7	20%	Constant	Interruptible load
	4	14%	Constant	Interruptible load
Metal industry	1	41%	Peak forward	TOU price/Critical Peak Pricing/interruptible load
	3	18%	Peak avoiding	Emergency demand-response
	4	14%	Constant	Interruptible load/ emergency demand response
Textile industry	2	22%	Peak avoiding	Emergency demand-response
	3	19%	Constant	Critical Peak Pricing/ interruptible load/ emergency demand-response
	7	18%	Constant	Critical Peak Pricing/interruptible load/ emergency demand-response
Non-metal industry	7	54%	Constant	Interruptible load
	1	15%	Peak forward	TOU price/interruptible load
	2	15%	Peak avoiding	Interruptible load

From the above cluster results, it can be seen that the first class represents the typical companies which products in the day and rests at night, containing the 27% of all companies, such as comprehensive retail and tourism hotel industry. In view of their power transferring ability, the lighting and air conditioner are the main available devices to adjust power utilization. Therefore, critical peak pricing and direct load control may be effective to them but with limited ability. While the second and third classes is reversed, which work at night and rest in day. It means that these companies have adapted some strategies to avoid the top price period. The companies of these classes, such as ce-

ment manufacturing, textile industry, non-metal manufacturing and so on, have little potential to be exploited. The other companies mainly work in triple shift schedule, which constantly works with 24 hours, however the production in the day is still a little higher. If their production could be changed, the adjustable potential are considerable. The refined analysis of load utilization is as the following tables

Conclusion

Firstly, it was described in this paper that much importance and requirement have been attached to the load data analysis as well as the reference indexes. Then, in view of the rough classification of load data, proposed a weighted fuzzy clustering algorithm to detail the load classification dividing. Two group of experiments are set to verify the efficiency of this method, one of which orients to the difference of load data from one user but different moments, and the other to the difference among different companies but the same day. The experiment results show that the algorithm is effective to accurately cluster the load data and supportive to the fine analysis of load data, promoting the development of electric technology.

Reference

[1] DONG Rui, HUANG Min-xiang. An improved FCM algorithm based on subtractive clustering for power load classification[J]. East China Electric Power, 2014, 42(5): 917-921.

[2] ZHANG Shao-min, ZHAO Shuo, WANG Bao-yi. Research of power load curve clustering algorithm based on cloud computing and quantum particle swarm optimization[J]. Power System Protection and Control, 2014, 42(21): 93-98.

[3] LIU Yong-guang, Sun Chao-liang, NIU Zhen-zhen, et al. Research on the improved fuzzy C-means clustering algorithm based power load characteristic classification technology[J]. Electrical measurement and instrumentation, 2014, 51(18): 5-9.

[4] SONG Yi-yang, LI Cun-bin, QI Zhi-qiang. Extraction of Power Load Patterns Based on Cloud Model and Fuzzy Clustering[J]. Power System Technology, 2014, 38(12): 3378-3383.

[5] Wang Ke, Wang Tian-xiu. Identification of Abnormal Power Load Data Based on Two Times Clustering[J]. Electrical Engineering, 2014, 11:1-3+17.

[6] LI Can-bing, QU Fang, WANG Xiao-ning. Character analysis of power load based on fuzzy clustering[J]. Journal of Zhengzhou University (Engineering Science), 2010, 31(1): 107-110.

[7] JIANG De-long, LI Sheng, WANG Ke-wen, et al. Research and implementation of a fuzzy clustering analysis system[J]. Computer Engineering and Science, 2011, 33(12): 121-125.

[8] LIU Jian-hua, WANG Jin, MENG Ying, et al. Clustering analysis of power load by rough set and K-means based on simulation annealing method[J]. Modern Electric Power, 2012, 29(1): 10-14.

[9] LIU Jian-hua, WANG Jin, YANG Hong-chun et al. Power load clustering analysis based in ACO-PAM synthesis algorithm[J]. Journal of Electric Power Science and Technology, 2011, 26(4): 94-99.

[10] 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.

[11] 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.

Jie Fan received her M.S. degree in Dongnan University, China,2014. She is currently the manager of metering department of Jiangsu electric power company research institute. Her research interest focus on demand response, Electric Power Metrology and etc.