

Application of Improved Fuzzy C-means Clustering Analysis Method of Load Characteristics Stats

Li Lin^{1,a}, Liu Dong^{1,b}, Du Ying^{1,c}, Liu Junli^{2,d}

¹State Grid Shandong Electric Power Research Institute, Ji'nan 250002, China

²State Grid Liaocheng Electric Power Company, Liaocheng 252000, China

^alynn963@163.com, ^brzdludong@163.com, ^cduyliaocheng@126.com, ^dlililovejt@163.com

Keywords: fuzzy C-means clustering; load characteristic; power load; cluster analysis

Abstract. Power load is the active part of safe and stable operation of the entire power system. Establishing realistic dynamic load model is important meaning for power system planning and operation. In this paper practical load modeling idea is adopted for clustering analysis of the data of load characteristics with fuzzy C-means method to provide a basis for load management decisions.

Introduction

Cluster analysis is a new multivariate statistical method, is a combination of modern taxonomy and multivariate analysis. Cluster analysis is to classify objects in a multidimensional space, according to the degree of closeness of their [1][2][3]spatial relationship to categorize. In other words, different attributes things were identified each other, things which have similar properties were clustered together.

Currently, [4][5][6]applications of cluster analysis in power system have diagnosis of power system, power load forecasting and power load modeling. There are many literature about power system short-term load forecasting in domestic and foreign, adopted forecasting methods and forecasting accuracy are also different. [7][8][9]Various prediction methods have certain limitations because many factors which affect the load and the uncertainty of load. Traditional short-term load forecasting methods have regression model, time series, [10][11][12]etc., with the rise and development of artificial intelligence, fuzzy clustering method is also introduced into the short-term load forecasting, and achieved good results.

Overview of Fuzzy C-means Clustering Algorithm

Fuzzy C-means clustering is called for FCM, is used to determine the membership of each data point belongs to a clustering algorithm clustering degree.

FCM put the n vectors $x_i(1, 2, \dots, n)$ into c fuzzy groups, and calculate cluster centers of each group, so that the value function of dissimilarity index is minimized. FCM can calculate the degree belonging to each group of each data point with the value given in membership between 0 and 1, values of sparse matrix U were allowed between elements 0 and 1, Corresponded on introducing fuzzy division. However, the sum of membership of a data set is equal to 1 with normalized provisions.

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n \quad (1)$$

Then the value function of FCM is a generalization,

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2, \quad (2)$$

Here value of u_{ij} is between 0 and 1; c_i is fuzzy cluster center of group i ; $d_{ij} = \|c_i - x_j\|$ is Euclidean distance between i -th clustering center and j -th data point; and $m \in [1, \infty)$ is a weighted index.

A necessary condition which make equation (2) reach a minimum can be obtained through constructing the followed new objective function:

$$\begin{aligned}\bar{J}(U, c_1, \dots, c_c, \lambda_1, \dots, \lambda_n) &= J(U, c_1, \dots, c_c) + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{ij} - 1) \\ &= \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{ij} - 1)\end{aligned}\quad (3)$$

Here, the value of j is 1 to n in λ_j which are Lagrange multipliers of constraint formula of the equation (1). The necessary conditions which make the formula (2) reach the minimum are:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (4)$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (5)$$

Fuzzy C-means clustering algorithm is a simple iterative process under the above two necessary conditions. FCM cluster centers c_i and membership matrix U is determined by the following procedures when run on batch mode.

Step 1: a random number between 0 and 1 is used to initialize the membership matrix, which can satisfy the constraints in formula (1).

Step 2: formula (4) is used to calculate a cluster center c_i , $i = 1, \dots, c$.

Step 3: the value function is calculated according to the formula (2). If it is less than a determined threshold or its amount of change respected to the last change less than a certain threshold, the algorithm stop.

Step 4: the new matrix U is calculated with formula (5), and then return to step 2.

Above algorithm can first initialize the cluster center, then perform the iterative process. Because it is not sure that FCM converge to an optimal solution, performance of the algorithm depends on the initial cluster centers. So, either determine the initial cluster centers with additional fast algorithm, or adopt the different initial cluster centers each time to start algorithm.

Fuzzy C-means Clustering Algorithm Flowchart

Program flow shown in Fig. 1:

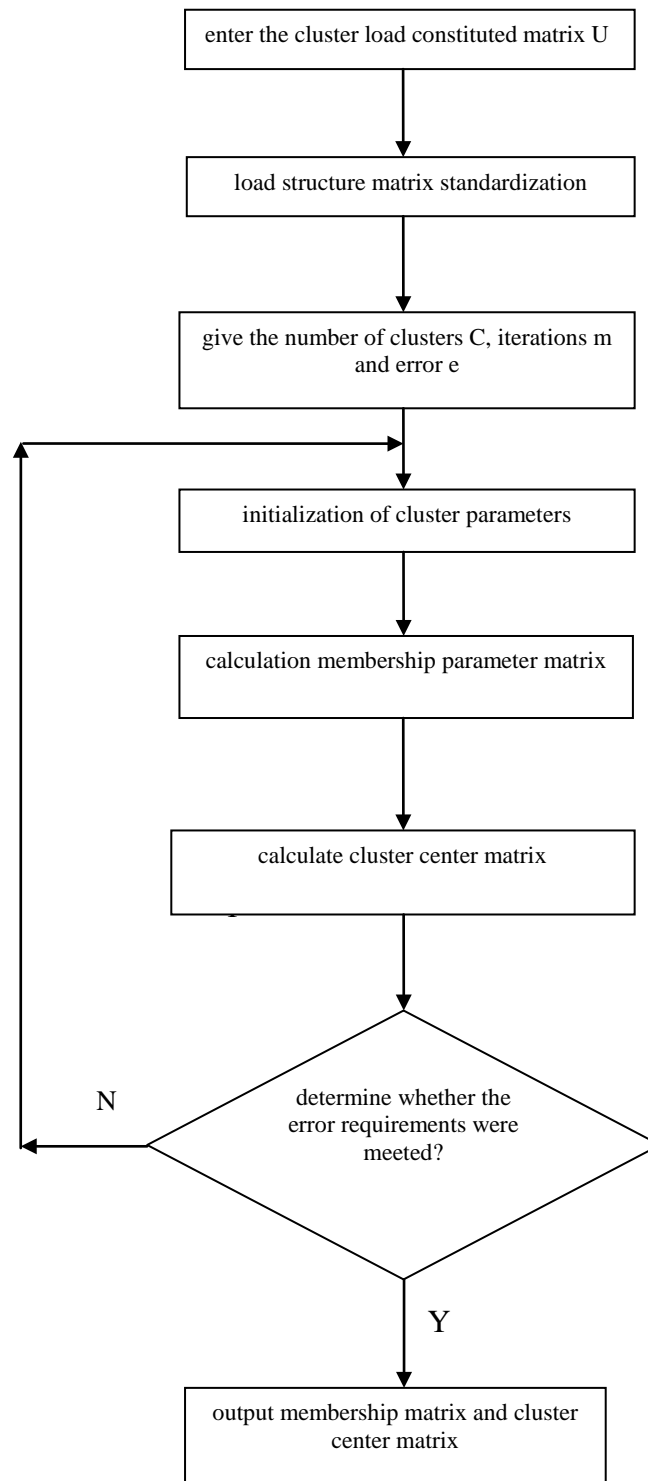


Fig.1. flow chart based on fuzzy C-means clustering algorithm

Improvement of Fuzzy C-means Method

FCM algorithm is simple and can achieve a good clustering result, is one of the most popular algorithms of the cluster analysis, but it has a fatal weakness: first fuzzy clustering objective function is a non-convex function, there are a lot of local extreme point, which will not get the optimal fuzzy partition of datasets because of improper initialization; the second is its practical application was restricted when the algorithm calculate large amounts of data.

In this paper, a user-defined selection method is used to select the initial cluster centers. First the system clustering method is used for data of initial clustering, and then the average distance of each class is calculated as the initial cluster centers Fuzzy C-means, so that you can successfully avoid

reducing the clustering effect because of data outliers or noise, etc. So the dependence on the initial cluster centers of fuzzy C-means greatly reduced, thereby increase the effect of fuzzy C-means clustering method.

Application of Fuzzy C-means Clustering Algorithm

5021 large industrial customers of run capacity of more than 2000 KVA and total reported installed capacity more than 49,961,600 KVA in the Shandong Province is analyzed in this paper. There are 4565 useful data and reported installed capacity of 43,959,000 KVA in above customers, wherein there are 3259 continuous production customers, reported installed capacity of 36,318,100 KVA and 1306 non-continuous production customers, reported installed capacity of 7,640,900 KVA. The time of a typical daily morning peak is 8: 30-11: 00, the load can be staggered is 1.1175 million kilowatts and the load can be off-peak is 8.0739 million kilowatts. The time of waist peak is 13: 30-17: 30, the load can be staggered is 1.2106 million kilowatts and the load can be off-peak is 9.0971 million kilowatts.

This design is programmed by MATLAB which has the software toolbox with tools function on fuzzy C-means method and the tool function is changed in this design. In this design part of automatically generating membership is removed; first the initial membership matrix by the system clustering method is entered, then fuzzy C-means clustering is implemented by the tool function. The membership degree matrix generated by the cluster center is inputted into improved fuzzy C-means function to obtain the fuzzy C-means clustering results.

The results will be in the following Table 1.

Table 1 ability of off-peak and staggering of industrial key customers above 2000kVA

| No. | Company | Continuity manufacturer | | Discontinuous manufacturer | | State of off-peak | | State of staggering | |
|--------|------------|-------------------------|----------|----------------------------|----------|---|--|---------------------------------------|--------------------------------------|
| | | User number | capacity | User number | capacity | Load of staggered in time of 8: 30-11: 00 | Load of staggered in time of 13:30-17:30 | Off-peak load in time of 8: 30-11: 00 | Off-peak load in time of 13:30-17:30 |
| 1 | jinan | 157 | 269.91 | 71 | 48.76 | 59.72 | 64.9 | 13.49 | 12.43 |
| 2 | qingdao | 284 | 304.5 | 133 | 57.6 | 44.1 | 50.11 | 15.84 | 17.14 |
| 3 | zibo | 416 | 408.16 | 60 | 26.27 | 92.12 | 107.21 | 7.62 | 7.32 |
| 4 | yantai | 370 | 292.85 | 174 | 90.56 | 73.01 | 81.01 | 14.39 | 16.83 |
| 5 | weifang | 341 | 240.81 | 252 | 133.41 | 56.61 | 68.31 | 12.07 | 14.6 |
| 6 | weihai | 70 | 68.37 | 81 | 34.47 | 16 | 17.7 | 4.85 | 5.23 |
| 7 | dongying | 206 | 316.03 | 37 | 31.51 | 34.25 | 39.32 | 3.65 | 3.23 |
| 8 | taian | 154 | 164.57 | 47 | 25.06 | 23.55 | 29.28 | 2.84 | 3.2 |
| 9 | rizhao | 96 | 144.72 | 45 | 18.05 | 27.75 | 27.91 | 2.25 | 2.39 |
| 10 | linyi | 293 | 195.46 | 69 | 28.94 | 54.78 | 65.08 | 7.92 | 8.4 |
| 11 | liaocheng | 143 | 154.81 | 77 | 41.03 | 28.9 | 29.51 | 3.51 | 3.42 |
| 12 | laiwu | 55 | 188.67 | 17 | 18.62 | 39.84 | 43.73 | 3.19 | 1.95 |
| 13 | jining | 186 | 266.25 | 56 | 37.66 | 160.03 | 168.78 | 5.19 | 7.66 |
| 14 | heze | 123 | 148.76 | 61 | 47.75 | 26.9 | 28.48 | 2.76 | 2.79 |
| 15 | dezhou | 143 | 160.04 | 62 | 66.25 | 27.09 | 37.55 | 6.9 | 7.86 |
| 16 | binzhou | 127 | 168.52 | 26 | 22.55 | 26.06 | 30.11 | 3.27 | 3.94 |
| 17 | zaozhuan g | 95 | 139.38 | 38 | 35.59 | 16.7 | 20.7 | 2.03 | 2.68 |
| toltal | | 3259 | 3631.81 | 1306 | 764.08 | 807.41 | 909.69 | 111.77 | 121.07 |

Conclusions

The data is clustered by improved fuzzy C-means method in this paper. The traditional fuzzy C-means method is improved to overcome the sensitivity of initial values and effect of isolated points. Clustering results validate the feasibility and validity that the improved FCM clustering method is applied to the load characteristics. On the other hand, the precision and the computing speed of FCM clustering method can be improved by using a composite clustering algorithm, the number of clusters can be determined by comparing the relevant statistics to increase the accuracy of clustering and overcome randomness and subjectivity from empirically.

This paper has a certain value for the orderly conduct electricity regulation work.

Acknowledgements

This work was financially supported by the National Natural Science Foundation (51277112).

References

- [1] ZHANG Hao,XIONG Jieqing,CHENQian.Component analysis and power model of traction load[J]. Proceedings of the CSU-EPSC,2015,27(6):37-42.
- [2] Li Long, Wei Jing, Li Canbing.Prediction of Load Model Based on Artificial Neural Network[J]. Transactions of China Electrotechnical Society,2015,30(8):225-230.
- [3] ZHENG Xiao-yu, ZHENG Jing-yuan.Real-Time Load Modeling for Smart Grid[J]. Power and Energy,2015,36(1):42-45.
- [4] Fahmy F H. Scheduling and resource allocation of non-conventional power systems via multi-level approach[J].Solar Energy Materials and Solar Cells,1994, 35(11): 401-407.
- [5] LI Juan,XUE Yongrui,XU Bingyin,GAO Houlei,SUN Bo.Power system transient model based on analytic hierarchy evaluation method[J]. Power System Technology,2013.8, 37(8):2207-2211
- [6] LIU Jianfen,ZHAO Kai. The electric power network load early warning model based on improved genetic algorithm[J]. Electrotechnical Application, 2015,34(6):27-31.
- [7] LIANG Hong-hao, DING Guo-mao, JIANG Da-chuan. Research and Application in Smart Grid Suited User Interaction Management Platform[J]. Electrical Measurement & Instrumentation, 2012, 49(1): 49-51.
- [8] JIANG Li-min, HUO Da-wei, YAN Hua-guang, et al. The Numerical Analysis in the Technical Loss of Distribution Network[J]. Electrical Measurement & Instrumentation, 2012,49(9): 47-51.
- [9] Les Pereira,Dmitry Kosterev,Peter Makin,et al.An interim dynamic induction motor model for stability studies in the WSCC. IEEE Transactions on Power Systems. 2002.
- [10] Krishnaswamy Srinivasan,Claude Lafond. Statistical Analysis of Load Behavior Parameters at four Major Loads. IEEE Transactions on Power Systems. 1994
- [11] IEEE Task Force on Load Representation for Dynamic Performance.Standard load models for power flow and dynamic performance simulation. IEEE Transactions on Power Systems. 1995
- [12] WANGHongfu,WANGYi,GAOChong. Iterative algorithm of DC power flow based on network loss equivalent load model[J]. Automation of Electric Power Systems,2015,39(1):99-103.