Research on Classification of Daily Load Curve of Distribution Network Based on Improved SSA-FCM

Haoran Shi¹, Rong Cao², Wenbo Hao², Mingyu Xu², Heng Hu², Peng Jiang², and Feng Zhou¹*

¹ College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong Province, 266590, P.R.China
² State Grid Heilongjiang Electric Power Company Limited, Heilongjiang Province, 150090, P.R.China
*Corresponding author’s e-mail: zhoufeng@sdust.edu.cn

Abstract. In the analysis of three-phase unbalance in distribution network, the accuracy of daily load curve classification results determines the size of three-phase unbalance. Aiming at the shortcomings of Fuzzy C-Means (FCM), a fuzzy C-Means clustering algorithm (SSA-FCM) optimized based on Sparrow Search Algorithm (SSA) is proposed. The cluster validity evaluation index is introduced to get the optimal quantity of clusters, and the SSA is used to search for the initial cluster center, which solves the problem that the FCM algorithm relies on the initial value and is easy to converge to local optimal solution. The simulation results show that, compared with the FCM algorithm, the load curves classified into the same category by SSA-FCM are closer together.

1. Introduction

With the progress of my country's technological level and economic development, the scale of the power system continues to expand and the structure becomes more complex. At present, in our country's low-voltage power supply system distribution network, the user's load types are complex, and the difference in daily load characteristics is becoming more and more huge[1]. It is important to study a fast and accurate classification method based on user's actual daily load curve[2].

At present, in the daily load characteristic curve classification, the main methods used are: K-means algorithm[3], FCM[4], DBSCAN algorithm[5], hierarchical clustering algorithm[6], Gaussian mixture model (GMM) algorithm[7], support vector machine (SVM) algorithm[8] and extreme learning machine (ELM) algorithm[9], etc.

Among these methods, the FCM algorithm is one of the most widely used power load characteristic classification algorithms. However, there are three shortcomings in applying FCM algorithm to classify users' daily load curves:

- The fuzzy weighted index m can only be selected based on experience, and there is still a lack of theoretical guidance.
- The number of clusters directly affects the final result, but the best number of clusters cannot be obtained in advance.
- FCM is easy to converge to the local optimal solution, but can not get the global optimal solution. The starts impact the effectiveness of FCM greatly.
In view of the current problems of the FCM, the sparrow search algorithm is used to improve FCM, and SSA is used to quickly seek the optimal initial cluster centroid for FCM, so as to reduce the possibility of FCM falling into local optimization and obtain an approximate global optimal cluster partition.

2. Fuzzy C-means clustering algorithm based on SSA

2.1. Sparrow search algorithm

The sparrow search algorithm is a swarm intelligence optimization algorithm proposed based on the behavior of sparrows foraging and evading predators[10]. In the sparrow search algorithm, individuals are divided into discoverers, followers, and guards, and each individual position corresponds to a solution. According to the algorithm setting, the population proportion of the guards is 10%~20%.

The following is the location update of the discoverer:

\[
X_{i,j}^{t+1} = \begin{cases} 
X_{i,j} \exp(-\frac{t}{\text{iter}_\text{max}}) & \text{if } R_t < ST \\
X_{i,j} + Q \Delta L & \text{if } R_t \geq ST 
\end{cases}
\]

Among them, \(t\) represents the current number of iterations; \(\text{iter}_\text{max}\) represents the maximum number of iterations; \(X_{i,j}\) represents the position information of the \(i\)-th sparrow in the \(j\)-th dimension. \(\alpha \in (0,1]\) is a random number. \(R_2(R_2 \in [0,1])\) and \(ST(ST \in [0.5,1])\) represent the warning value and safety value respectively; \(Q\) is a random number that obeys a normal distribution Number. \(L\) is a 1 by \(d\) matrix.

The location update description of the joiner is:

\[
X_{i,j}^{t+1} = \begin{cases} 
\frac{X_{\text{wor}e} - X_{i,j}^{t+1}}{\epsilon} & \text{if } f_i > f_G \\
X_{i,j}^{t+1} & \text{otherwise}
\end{cases}
\]

Among them, \(X_P\) is the optimal position of the discoverer in the population; \(X_{\text{wor}e}\) represents the worst position of the sparrow in the current population. \(A\) represents a 1×\(k\) matrix, where each element is randomly assigned a value of 1 or -1, and \(A^+ = A^T(AA^T)^{-1}\).

The initial position of the vigilant is randomly generated in the population. The expression is as follows:

\[
X_{i,j}^{t+1} = \begin{cases} 
X_{\text{best}} + R_{i,j} \frac{X_{i,j}^{t+1} - X_{\text{wor}e}}{\epsilon} & \text{if } f_i > f_G \\
X_{i,j}^{t+1} + K \frac{X_{i,j}^{t+1} - X_{\text{wor}e}}{(f_i - f_G) + \epsilon} & \text{if } f_i = f_G 
\end{cases}
\]

Among them, \(X_{\text{best}}\) is the current global optimal position; \(\beta\) is a random number subject to a normal distribution with a mean of 0 and a variance of 1; \(K \in [-1,1]\) is a random number, and \(f_i\) is the adaptation of the current sparrow individual Degree value; \(f_G\) and \(f_w\) are the current global best and worst fitness values respectively; \(\epsilon\) is a constant.

2.2. SSA-FCM algorithm

The SSA algorithm has good global search characteristics. Aiming at the FCM algorithm’s dependence on the initial value and the shortcomings of converging to local extremum easily, the sparrow search algorithm is used to optimize the original algorithm, and the global search of the sparrow algorithm is used to replace the gradient descent of the FCM algorithm. Process to avoid local extreme values and reduce dependence on initial values. In the search process, for the evaluation of each individual sparrow, the fitness function defined is consistent with the objective function of the FCM algorithm. The specific steps for SSA-FCM are:

- Determine the range of the number of clusters \(C=[2, C_{\text{max}}]\).
- Initialize the population. Determine the number of iteration terminations, the population size, initialize the discoverer-to-participant ratio, randomly select \(c\) cluster centers as sparrow
individuals from the sample, and repeatedly generate multiple sparrow individuals to generate the initial population.

- Execute FCM clustering algorithm. Calculate the membership degree of each individual sparrow, then update the cluster center, and traverse each particle to achieve the effect of compactness and separation between clusters.
- Calculate the fitness function value, calculate the fitness value of each individual, and record the individual extreme value and the global extreme value of the entire group.
- Check termination condition. If the maximum number of iterations is reached or if the fractional positive epsilon or the fuzzy matrix less than specified is not changed, the calculation ends. Otherwise, according to formula (1) and formula (2) And formula (3) update the location of the discoverer, joiner and guard. Form a new cluster center, go to step 2.
- Set the initial value of the FCM to the result obtained in step 5, the FCM algorithm is executed for classification, and output the result.

### 3. Daily load curve classification based on SSA-FCM algorithm

#### 3.1. Construction of the input data set

The user load data obtained by the smart collection terminal in the smart grid contains bad data and abnormal data, which cannot be cleaned and eliminated during the data collection process. Therefore, it is also necessary to clean the input samples. After removing the bad data and abnormal data in the daily load data, the curve data which is missing a small amount of data is processed by smoothing formula.

#### 3.2. Selection of fuzzy weighting index m

The previous article summarized the shortcomings of the FCM algorithm. Among them, the fuzzy weighting index m is determined based on experience. The value of m impacts the classification results greatly. The larger the value of m, the more fuzzy the clustering. Bezdek gave a range of experience; later, from the physical explanation, m=2 was the most meaningful; Chan et al. obtained from the application background of Chinese character recognition that the best value of m should be between 1.25 and 1.75; Pal et al. obtained from the experimental study on clustering effectiveness that the optimal selection interval of m was [1.5, 2.5], and generally take the median of the interval m=2 without special requirements.

#### 3.3. Cluster validity evaluation index

The traditional clustering effectiveness indicators are PC, MPC, PE, XB, UV, FM. However, these indicators can only exert their advantages on specific data sets, and cannot be applied to all data sets. This paper chooses a new fuzzy clustering effectiveness index $W(c,U)$, the definition is as follows:

$$W(c,U) = Var^c(c,U) + \frac{Cop^c(c,U)}{Sep^c(c,U)}$$  \hspace{1cm} (4)

Where, $Var^c(c,U)$ is intra-class compactness; $Cop^c(c,U)$ is the overall overlap degree; $Sep^c(c,U)$ is the overall degree of separation.

The $Var^c(c,U)$ smaller, the closer the distance within the class; the $Sep^c(c,U)$ larger, the more scattered the distance between classes; the $Cop^c(c,U)$ smaller, the smallest overlap. Obviously, the $W(c,U)$ smaller the clustering effect is the best.

#### 3.4. Algorithm flow design

The overall algorithm includes load data preprocessing and determining the optimal number of clusters in the SSA-FCM algorithm. The following is the overall process of user daily load curve classification using SSA-FCM algorithm:

- Carry out load data preprocessing on the original daily load curve data;
- Set and minimum cluster number \(c_{\text{min}}\) and maximum cluster number \(c_{\text{max}}\), SSA-FCM related parameters;
- Execute the SSA-FCM algorithm for cluster analysis;
- Use W index to evaluate the effectiveness of clustering;
- Determine whether the current \(c\) is greater than the maximum number of clusters \(c_{\text{max}}\), if yes, go to step 6. If not, \(c = c+1\), execute step 3;
- Output the optimal number of clusters and index values.

4. SSA-FCM algorithm daily load curve classification simulation analysis

4.1. Daily load curve classification

In June 2021, 55 users' daily load data were collected in a city of Heilongjiang. Each user collected 96 times a day for 15 minutes. After processing, it finally contains 50 daily load curves.

In order to remove the influence of the magnitude of the load on the classification results, the user's daily load data is normalized before the classification. Here, extreme value sequence normalization is used. After normalization, all the data are in \([0,1]\).

Use SSA-FCM algorithm and FCM to classify the processed data respectively, and compare the classification results. The parameters of the SSA-FCM algorithm are set as follows: minimum number of clusters \(c_{\text{min}}=2\), maximum number of clusters \(c_{\text{max}}=6\), fuzzy weighting index \(m=2\), population size 200, and maximum number of iterations \(\text{inter}_{\text{max}}=300\). The FCM algorithm parameters are: fuzzy weighting index \(m=2\); the maximum number of iterations is \(t_{\text{max}}=300\), and the objective function threshold is \(10^{-7}\).

Figure 1. Trend of W indicator.

Figure 3 shows that the clustering number of the two methods is 3 when the W index reaches the minimum value, so the optimal clustering number of the two algorithms is 3. In addition, when taking the same number of clusters, the W index values of SSA-FCM algorithm are all smaller than those of FCM algorithm, so SSA-FCM has better performance.

| Algorithm name | Type 1 | Type 2 | Type 3 |
|----------------|--------|--------|--------|
| SSA-FCM        | 41     | 4      | 5      |
| FCM            | 45     | 1      | 4      |

Table 1. Number of various types of curves.
Figure 2. SSA-FCM daily load curve classification results.

Figure 3. FCM daily load curve classification results.

Table 1 shows the number of daily load curves belonging to each type in the classification results of SSA-FCM and FCM, Figure 2 and Figure 3 compare the clustering effects of the two clustering algorithms, the bold curves in Figure 2 and Figure 3 are all kinds of clustering centers. The clustering centers obtained by SSA-FCM algorithm are different from those obtained by FCM algorithm. Combined with Table 1, Figure 2 and Figure 3, the clustering center obtained by SSA-FCM algorithm is different from that of FCM algorithm. Combined with Table 1, Figure 2 and Figure 3, the classification results of the two algorithms are different. The daily load curve classification results are obtained based on SSA-FCM algorithm, the daily load curves of the same category are relatively close, while in the results of FCM algorithm, the daily load curve deviates from the cluster center more seriously. Therefore, SSA-FCM algorithm has better clustering effect.
5. Conclusion
At present, the most widely used method in daily load curve classification is the FCM algorithm. Because it is a local search algorithm, it is very sensitive to initialization and converge easily to local minimums. At the same time, when using this algorithm, the best number of clusters cannot be obtained in advance. In order to solve the problem of FCM, a FCM algorithm based on SSA algorithm is proposed.

- Adopt a new fuzzy clustering effectiveness index. This indicator is composed of three parts: compactness, separation and overlap. It can accurately find the optimal number of clusters on noise data sets and overlapping data sets.
- Use the SSA algorithm to improve the FCM clustering algorithm, and replace the FCM with SSA instead of the iterative process based on gradient descent. Simulation experiments show that, compared with the FCM, the SSA-FCM is classified into the same category of daily load curves are closer, so the improved algorithm has better clustering accuracy.

The proposed algorithm can make the power consumption characteristics of the same class of users have high similarity and there are obvious differences between different classes of users. It provides an accurate and scientific user classification method for the power department, which is of great significance to solve the three-phase imbalance problem of distribution network and reduce the network loss of distribution network.

Acknowledgments
Funded by the 2020 R&D Science and Technology Project of State Grid Heilongjiang Electric Power Company Limited: Research on loss reduction technology of three-phase unbalance distribution load in Heilongjiang Province. (523437190010)

References
[1] Chaohong, B., Xu, W., Yuan, H. (2017) Research review and prospects of energy internet planning. J. Proceedings of the Chinese Society of Electrical Engineering, 37(22): 6445-6462+6757.
[2] Xiaoxin, Z., Shuyong, C., Zongxiang, L., et al. (2018) Technical characteristics of China's new generation power system in the energy transition. J. Proceedings of the Chinese Society of Electrical Engineering, 38(7): 1893-1904.
[3] Wenqing, Z., Yaqiang, G. (2016) Load curve clustering based on Kernel K-means. J. Electric Power Automation Equipment, 36(6): 203-207.
[4] Yongguang, L., Chao, L., Zhenzhao, N., et al. (2014) Research on power load characteristics classification technology of improved fuzzy C-means clustering algorithm. J. Electrical Measurement and Instrumentation, 51(18): 5-9.
[5] Guilin, W., Guoliang, Z., Hongshan, Z., et al. (2016) Fast clustering and anomaly detection technology for large-scale electricity data flow. J. Automation of Electric Power Systems, 40(24): 27-33.
[6] Xiaojian, G., Xiaoxia, L., Xiaoling, L. (2008) Improvement and analysis of hierarchical clustering algorithm. J. Computer Applications and Software, 25(6): 243-246.
[7] Jia, Y., Shitong, W. (2006) Research on EM Algorithm and Initialization in Gaussian Mixture Model Clustering. J. Microcomputer Information, 11(22):244-247.
[8] Xiaopu, F. (2010) Research on Power User Classification Technology Based on Actual Load Curve. J. Electric Power Science and Engineering, 26(09): 18-22.
[9] NAGI, J., YAPK, S., TIONG, S.K., et al. (2010) Nontechnical loss detection for metered customers in power utility using support vector machines. J. IEEE Transactions on Power Delivery, 25(2): 1162-1171.
[10] Jiankai, X. (2020) A novel swarm intelligence optimization approach: sparrow search algorithm. J. Systems Science & Control Engineering: An Open Access Journal, 8(1): 22-34.
[11] Jiayi, G., Xuezhong, Q., Shiping, Z. (2019) New fuzzy clustering validity index. J. Application Research of Computers, 36(4): 1001-1006.