STUDY ON THE ANALYSIS OF FAULT DATA DURING INTELLIGENT OPERATION OF POWER EQUIPMENT WITH CLUSTER ANALYSIS

Meng Li[1], Jiajia Zhu[1], Zihan Liu[1]*
[1]Information and Telecommunication Branch, State Grid Jiangsu Electric Power Co, Ltd., Nanjing, Jiangsu 210000, China.

Abstract - Many fault-related data are generated in the intelligent operation of power equipment, and analyzing these data can help understand the operating condition of the equipment. This paper studied the clustering analysis methods and introduced K-means, kernel C-means (KCM), and fuzzy kernel C-means (FKCM) methods. The gray wolf optimization (GWO) algorithm was used to optimize the initial clustering center and kernel parameters of the FKCM algorithm to obtain the GWO-FKCM algorithm. Experiments were conducted taking transformer data as an example. It was found that the GWO-FKCM algorithm obtained the optimal clustering results at the 763rd iteration, and its accuracy rates for different fault types were all above 90%, with an average value of 92.92%, which was higher than the other clustering analysis methods. The results prove the reliability of the GWO-FKCM algorithm for equipment fault data analysis. This algorithm can be promoted and applied in actual power equipment.

Keywords: Clustering analysis, Power equipment, Fault data, Gray wolf optimization algorithm, Smart grid.

1. Introduction

Electric power equipment is a very basic and important part in the electric power system [1]. After a long period of operation, equipment inevitably fails to operate, causing huge losses to the power company and users [2]. However, due to the update of technology, the data involved in the intelligent operation of power equipment has further increased, bringing great difficulty to the analysis of fault data. With the advancement of data mining technology, many methods have been applied in the fault data analysis of power equipment [3]. Liu et al. [4] studied the fault analysis of distributed power grids based on the Bayesian algorithm combined with evidence theory and found through simulation experiments that the method had high fault tolerance and performed well in fault analysis under the situation of incomplete information. Sharan et al. [5] designed a method based on spectrum analysis to monitor open-circuit faults in grid-connection and verified the effectiveness of the method by simulation analysis. Stallon et al. [6] designed a method called kernel principal component analysis-enhanced spider monkey optimization to achieve the analysis of grid faults and found through experiments that the method obtained 98% accuracy. Badr et al. [7] proposed a support vector machine (SVM) based method for photovoltaic array fault monitoring and demonstrated the performance of the method through experiments in MATLAB/Simulink. This paper used the method of cluster analysis to study the fault data analysis, designed a gray wolf optimization-fuzzy kernel C-means (GWO-FKCM) algorithm, and conducted experiments with transformers. This work provides some referable ideas to further realize the reliable analysis of equipment faults and promote the smooth operation of power grids.

2. Clustering-based Analysis Method for Power Equipment Fault Data

2.1 Clustering Analysis Algorithm

The data generated in the intelligent operation of power equipment can be used to assess the status of the equipment [8], and these data have the characteristics of wide distribution, great varieties, and massive amount, resulting in a slow and inefficient fault data analysis, so a fast and accurate fault analysis method is particularly important [9]. Data mining techniques can effectively improve the effect of fault data analysis. Currently, the commonly used data mining techniques include neural networks, SVMs, and so on [10].

Cluster analysis is a method to understand the distribution of data [11], which can discover the data characteristics of all categories. It is based on the principle that data with high proximity are brought together into the same class. In the analysis of power

equipment data, there are large differences between fault data and normal data; therefore, the method of cluster analysis can be used to analyze power equipment fault data.

2.2 K-means Algorithm

The K-means algorithm is one of the most widely used clustering algorithms [12], and its basic steps are as follows.

(1) Suppose there are n samples, and the initial data center of k clusters is randomly selected.

(2) The distance between all data and the center is calculated, and it is classified into the nearest class. The calculation formula of the distance is:

$$d_{ij} = \sqrt{(x_{i1}-y_{j1})^2+(x_{i2}-y_{j2})^2+\cdots+(x_{in}-y_{jn})^2}, \quad (1)$$

where d_{ij} refers to the Euclidean distance between data points $\mathrm{x_i}$ and $\mathrm{y_j}$.

(3) New k clustering centers are calculated constantly. The clustering stops until the specified number of iterations is reached.

The K-means algorithm has a relatively simple principle and usually obtains stable clustering results, which has good applications in biological research [13], material analysis [14], etc. However, the method also has an obvious drawback: its initial center value is generated randomly, which cannot guarantee that the final result can converge to the globally optimal solution; therefore, in order to improve the effect of the K-means algorithm in power equipment fault data analysis, it needs to be improved.

2.3 Fuzzy C-means Clustering

The fuzzy C-mean clustering (FCM) algorithm is an algorithm for fuzzy processing based on K-means [15]. If there is a sample set $X=\{x_1,x_2,\cdots,x_n\}$ and it is divided into C classes. There exists affiliation matrix U whose value range is [0,1], $\sum_{i=1}^{c}u_{ij}=1,\forall j=1,2,\cdots,n$, then the objective function of the FCM algorithm is written as:

$$J(U,c_1,c_2,\cdots,c_n)=\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m d_{ij}^2, \quad (2)$$

where d_{ij} refers to the distance between the j-th sample x_j and the i-th center c_i. $m\in(1,\infty)$, which is a weight index; the larger its value is, the more fuzzy the clustering is.

Using the Lagrange multiplier method, $\mathrm{c_i}$ and $\mathrm{u_{ij}}$ are solved:

$$c_i=\frac{1}{\sum_{j=1}^{n}(u_{ij})^m}\sum_{j=1}^{n}(u_{ij})^m x_j, \quad (3)$$

$$u_{ij}=\left[\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}\right]^{-1}. \quad (4)$$

The steps of FCM calculation are as follows. First, the initial samples are input, the initial values are set, and U is initialized; secondly, c_i and u_{ij} are constantly updated following the formulas above until the cluster center no longer changes; finally, the clustering result is output.

2.4 Fuzzy Kernel Clustering

Compared with the FCM algorithm, the FKCM algorithm [16] introduces the idea of kernel methods in SVM to further improve the effectiveness of clustering.

Suppose there exists sample $x_k\in R^N$. Through nonlinear mapping $\Phi(\cdot):x_k\in R^N\to Y\in R^M, N<M$, the sample space is mapped to a high-dimensional feature space. The kernel function is $K(x_i,x_j)=\langle\Phi(x_i),\Phi(x_j)\rangle$, and $\langle,\rangle$ is the inner product operation. In the feature space Y, the objective function of FKCM is written as:

$$\mathrm{J_Y(X,U,C)=\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m d_{ij}^2=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left[K(x_j,x_j)-2K(x_j,c_i)+K(c_i,c_i)\right]}, \quad (5)$$

where $\mathrm{K(x_j,x_j)}$ is the the inner product of the auto-kernel function of the j-th sample $\mathrm{x_j}$, $\mathrm{K(c_i,c_i)}$ is the inner production of the auto-kernel function of the i-th center $\mathrm{c_i}$, and $\mathrm{K(x_j,c_i)}$ is the inner product of the kernel function of $\mathrm{x_j}$ and $\mathrm{c_i}$.

Using the Lagrange multiplier, the update equations of $\mathrm{u_{ij}}$ and $\mathrm{c_i}$ are:

$$u_{ij}=\left[\sum_{k=1}^{c}\left(\frac{K(x_j,x_j)-2K(x_j,c_i)+K(c_i,c_i)}{K(x_j,x_j)-2K(x_j,c_k)+K(c_k,c_k)}\right)^{\frac{1}{m-1}}\right]^{-1}, \quad (6)$$

$$c_i=\frac{1}{\sum_{j=1}^{n}(u_{ij})^m K(x_j,c_i)}\sum_{j=1}^{n}(u_{ij})^m K(x_j,c_i)x_j. \quad (7)$$

The calculation steps of the FKCM algorithm are as follows. Firstly, the initial samples are input, the initial values are set, and the kernel function and its parameters are selected. Secondly, the clustering center and the kernel function matrix are calculated. Thirdly, according to the formulas above, c_i and u_{ij} are constantly updated until the termination condition is satisfied.

2.5 Fuzzy Kernel C-means Algorithm Combined with Gray Wolf Optimization

The FKCM algorithm is greatly influenced by the initial clustering center and kernel parameters; therefore, in order to obtain good results for power equipment fault data analysis, this paper combines the GWO algorithm [17] to obtain the optimal initial clustering center and kernel parameters, i.e., the GWO-FKCM algorithm.

In the iterative process of the GWO algorithm, individuals with the top three best positions are denoted as α, β, and δ, and the other individuals are denoted as ω. Its optimization process is the hunting process of the wolf pack, which is written as:

$$D = \left|CX_p(t) - X(t)\right|, \tag{8}$$

$$X(t+1) = X_p(t) - AD, \tag{9}$$

$$C = 2r_1, \tag{10}$$

$$A = 2ar_2 - a, a = 2 - 2 \times \frac{t}{T}, \tag{11}$$

where t is the number of iterations, X_p is the position of the prey, $X(t)$ is the position of the gray wolf, A and C are random vectors, r_1 and r_2 are random numbers in [0,1], and a is a convergence vector, which decreases linearly from 2 to 0. A and C are adjusted to achieve the update of the gray wolf position:

$$X_1 = X_\alpha - A_1 D_\alpha, \tag{12}$$

$$X_2 = X_\beta - A_2 D_\beta, \tag{13}$$

$$X_3 = X_\delta - A_3 D_\delta, \tag{14}$$

$$X(t+1) = (X_1 + X_2 + X_3)/3. \tag{15}$$

In order to avoid falling into local optimum, the value of a is improved to obtain a better convergence effect, and the corresponding formula is:

$$|a| = (a_{max} - a_{min}) \times \left(1 + e^{-15\times\left(0.618 - \frac{t}{T_{max}}\right)}\right)^{-1}, \tag{16}$$

where a_{max} and a_{min} are the maximum and minimum values of a and T_{max} is the maximum number of iterations.

The steps of the GWO-FKCM algorithm are the same as those of the FKCM algorithm except that the optimal initial clustering centers and kernel parameters are obtained using the GWO algorithm.

3. Results and Analysis

3.1 Experimental Data

This paper focused on the study of transformers in power equipment. At present, transformers are usually serviced periodically, which is not conducive to meet the reliability of power supply. The transformer faults studied included:

① low-energy discharge: arcs and sparks formed due to poor connections;

② high-energy discharges: discharges over the surface due to short-circuiting, etc.;

③ partial discharge: partial discharge caused by incomplete impregnation, oil over-saturation, etc.;

④ medium-low temperature overheating (t < 700℃): winding oil flow blockage, etc.;

⑤ high-temperature overheating (t > 700℃): short circuit between core silicon steel sheets, etc.

Dissolved gases analysis (DGA) is currently the most effective method for analyzing transformer faults [18]. When a transformer fails, due to internal heating and discharge, different gases will appear in the insulating oil, and H_2, CH_4, C_2H_2, C_2H_4, and C_2H_6 have violent reactions.

387 samples of fault data and 314 samples of normal data were collected from the historical data of a power company in Jiangsu. One data was taken from every type, and the corresponding gas content is shown in Table 1.

Table 1. Content of different types of gases (μL/L)

	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6
Low-energy discharge	171	8	55	15	25
High-energy discharge	110	105	185	200	10
Partial discharge	650	50	0	20	35
Medium-low temperature overheating	120	120	0.5	85	30
High-temperature overheating	765	995	5	670	115
Normal	5	2	0	0	7

3.2 Clustering Results

According to the fault type, the number of clusters in the cluster analysis was 6, and the feature dimension was 5. The population size of the GWO algorithm was 100, and the maximum number of iterations was 1000. $a_{max} = 2$, $a_{min} = 0.1$, and weight index $m = 2$. The experimental samples were divided into a training set and a test set according to a ratio of 2:1. The transformer fault data were analyzed by the cluster analysis method.

The number of iterations required for different cluster analysis methods was compared, and the results are displayed in Figure 1.

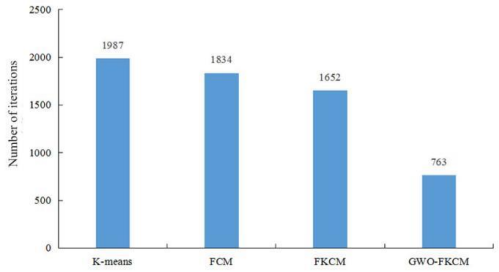
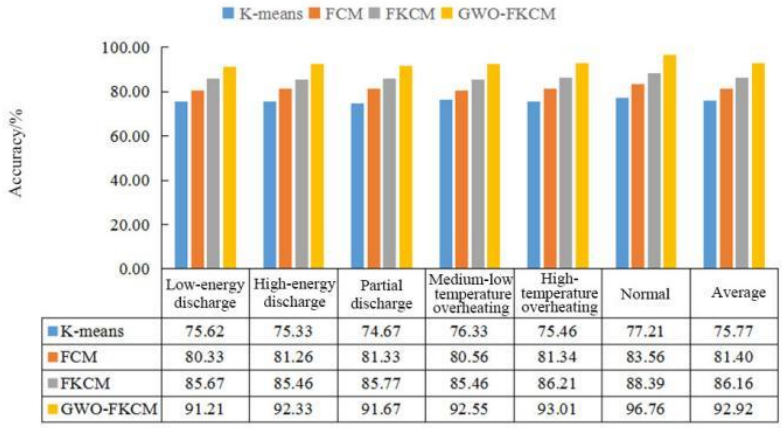


Figure 1: Comparison of the number of iterations between different clustering analysis methods

It was seen from Figure 1 that the number of iterations required for the K-means algorithm was as high as 1987; the number of iterations required for FCM and FKCM algorithms was 1834 and 1652, respectively; the GWO-FKCM method obtained the optimal results in the 763rd iteration, and the number of iterations it required was 61.6% less than the K-means algorithm and 53.81% less than the FKCM algorithm. This indicated that the optimization of the FKCM algorithm by the GWO algorithm effectively improved the efficiency of clustering analysis.

The accuracy of different clustering analysis methods was compared, and the results are presented in Figure 2.



	Low-energy discharge	High-energy discharge	Partial discharge	Medium-low temperature overheating	High-temperature overheating	Normal	Average
K-means	75.62	75.33	74.67	76.33	75.46	77.21	75.77
FCM	80.33	81.26	81.33	80.56	81.34	83.56	81.40
FKCM	85.67	85.46	85.77	85.46	86.21	88.39	86.16
GWO-FKCM	91.21	92.33	91.67	92.55	93.01	96.76	92.92

Figure 2: Comparison of the accuracy between different clustering analysis methods

It was observed in Figure 2 that these cluster analysis methods had the highest accuracy when analyzing the normal data, which might be due to the fact that the largest amount of data made the algorithm fully trained. Then, the comparison of different cluster analysis methods demonstrated that the traditional K-means algorithm had an accuracy of around 75%, with an average accuracy of 75.77%, while the average accuracy of the FCM algorithm was 81.40%, which was 5.63% higher than that of the traditional K-means algorithm.

After kernel function optimization, the accuracy of the FKCM algorithm was about 85% when analyzing different types of data, with an average value of 86.16%, which was 10.39% higher than the traditional K-means algorithm and 4.76% higher than the FCM algorithm. After the parameters of the FKCM algorithm were further optimized using the GWO algorithm, the GWO-FKCM algorithm achieved an average accuracy of 92.92% when analyzing fault data, which was 17.15% higher than the K-means algorithm, 11.52% higher than the FCM algorithm, and 6.76% higher than the FKCM algorithm.

Ten of these samples were used as an example to compare the fault analysis results of the traditional three-ratio method [19] with the GWO-FKCM algorithm, as shown in Table 2.

Table 2. Comparison of the three-ratio method with the GWO-FKCM algorithm

Sample number		1	2	3	4	5
Gas content (μL/L)	H_2	35	112	655	95	770
	CH_4	25	107	51	70	996
	C_2H_2	22	186	0	14	4
	C_2H_4	22	205	23	14	672
	C_2H_6	0	11	33	2	113
Actual faults		Low-energy discharge	High-energy discharge	Partial discharge	High-temperature overheating	High-temperature overheating

Three-ratio method		**High -energy discharge**	High-energy discharge	Partial discharge	**Low-temperature overheating**	High-temperature overheating
GWO-FKCM algorithm		Low-energy discharge	High-energy discharge	Partial discharge	High-temperature overheating	High-temperature overheating
Sample number		6	7	8	9	10
Gas content (μL/L)	H_2	6	72	175	125	126
	CH_4	3	50	9	105	125
	C_2H_2	0	2	57	220	0.4
	C_2H_4	0	60	17	155	87
	C_2H_6	8	70	24	10	32
Actual faults		Normal	High-temperature overheating	Low-energy discharge	High-energy discharge	Medium-low temperature overheating
Three-ratio method		Normal	High-temperature overheating	Low-energy discharge	**Partial discharge**	Medium-low temperature overheating
GWO-FKCM algorithm		Normal	High-temperature overheating	Low-energy discharge	High-energy discharge	Medium-low temperature overheating

It was observed in Table 2 that when using the traditional three-ratio method for fault analysis, the low-energy discharge of sample 1 was wrongly judged as high-energy discharge, the high-temperature overheating of sample 4 was wrongly judged as low-temperature overheating, and the high-energy discharge of sample 9 was wrongly judged as high-energy discharge, but the results obtained by the GWO-FKCM algorithm were all accurate, further proving the effectiveness of the cluster analysis method designed in this paper for equipment fault data analysis.

4. Conclusions

In this paper, an FKCM method optimized by the GWO algorithm was designed for analyzing fault data generated during the intelligent operation of power equipment. The performance of the method was analyzed with transformer fault data as an example. It was found through experiments that compared with traditional K-means, FCM and FKCM algorithms, the GWO-FKCM algorithm required fewer iterations to reach the optimal clustering result, presented higher accuracy when analyzing fault data, and achieved an average accuracy of 92.92%. Compared with the traditional three-ratio method, the GWO-FKCM algorithm was more accurate in judging transformer faults. The GWO-FKCM algorithm can be further applied in the actual fault data analysis of power equipment. At the same time, the research in this paper proved the superiority of the FKCM method and the reliability of the improved FKCM method, which provides a new idea to further improve the performance of cluster analysis methods. In the research of cluster analysis methods, the FKCM method can be a focus of in-depth research, and more optimization algorithms can also be combined with cluster analysis methods to obtain better results in data analysis in order to improve the reliability and applicability of cluster analysis methods.

References

[1] Yang Q. “Application and Research of Machine Learning Simulation Technology in Equipment Fault Monitoring under Smart Grid,” IOP Conference Series: Materials Science and Engineering, 2020, 782:1-7.

[2] Wang B, Yang K, Wang D, Chen SZ, Shen HJ. “The applications of XGBoost in Fault Diagnosis of Power Networks,” 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), 2019:3496-3500.

[3] Benkercha R, Moulahoum S. “Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system,” Solar Energy, 2018, 173(OCT.):610-634.

[4] Liu X, Yang X, Li C, Liu J, Zhang F. “Distributed Power Grid Fault Diagnosis Based on Naive Bayesian Network and D-S Evidence Theory,” Journal of Physics: Conference Series, 2020, 1549(5):1-12.

[5] Sharan B, Jain T. “Spectral analysis-based fault diagnosis algorithm for 3-phase passive rectifiers in renewable energy systems,” IET Power Electronics, 2020, 13(16):3818-3829.

[6] Stallon S, Rajkumar M N. “Improving the performance of grid-connected doubly fed

induction generator by fault identification and diagnosis: A kernel PCA-ESMO technique," International Transactions on Electrical Energy Systems, 2021, 31(4):e12844.1-e12844.23.

[7] Badr MM, Hamad MS, Abdel-Khalik AS, Hamdy RA. "Fault Detection and Diagnosis for Photovoltaic Array Under Grid Connected Using Support Vector Machine," 2019 IEEE Conference on Power Electronics and Renewable Energy (CPERE), 2019: 546-553.

[8] Ntalampiras S. "Fault Diagnosis for Smart Grids in Pragmatic Conditions," IEEE Transactions on Smart Grid, 2018, 9(99):1964-1971.

[9] Zhang X, Guo Z, Zheng Y, Liu J, Yan P, Zheng L. "Power grid fault diagnosis using polar PMU data plots," International Journal of Electrical Power & Energy Systems, 2022, 141:1-13.

[10] Sahri Z B, Yusof R B. "Support Vector Machine-Based Fault Diagnosis of Power Transformer Using k Nearest-Neighbor Imputed DGA Dataset," Journal of Computer & Communications, 2018, 02(9):22-31.

[11] Atsa'am D D, Wario R. "Hierarchical cluster analysis of the morbidity and mortality of COVID-19 across 206 countries, territories and areas," International Journal of Medical Engineering and Informatics, 2022, 14(2):125-133.

[12] Bigdeli A, Maghsoudi A, Ghezelbash R. "Application of self-organizing map (SOM) and K-means clustering algorithms for portraying geochemical anomaly patterns in Moalleman district, NE Iran," Journal of Geochemical Exploration: Journal of the Association of Exploration Geochemists, 2022, 233:1-13.

[13] Majdina N S, Soeleman M A, Supriyanto C. "Application of Particle Swarm Optimization (PSO) to Improve K-means Accuracy in Clustering Eligible Province to Receive Fish Seed Assistance in Java," IOSR Journal of Computer Engineering, 2022, 24(1):43-49.

[14] Kurita H, Suganuma M, Wang Y, Narita F. "k-Means Clustering for Prediction of Tensile Properties in Carbon Fiber-Reinforced Polymer Composites," Advanced Engineering Materials, 2022, 24(5):1-6.

[15] Pimentel B A, Silva R, Costa J. "Fuzzy C-Means Clustering Algorithms with Weighted Membership and Distance," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2022, 30(04):567-594.

[16] Rustam Z, Faradina R. "Application of Fuzzy Kernel C-Means in face recognition to identify look-alike faces," Journal of Physics: Conference Series, 2019, 1218:1-8.

[17] Balraj R, Stonier A A. "A novel PV array interconnection scheme to extract maximum power based on global shade dispersion using grey wolf optimization algorithm under partial shading conditions," Circuit World, 2022, 48(1):28-38.

[18] Rao UM, Fofana I, Rajesh KNVPS, Picher P. "Identification and Application of Machine Learning Algorithms for Transformer Dissolved Gas Analysis," IEEE Transactions on Dielectrics and Electrical Insulation, 2021, 28(5):1828-1835.

[19] Kong X, Zhou D, Gu Z, Ma H. "Dissolved Gas Analysis of Insulating Oil for Power Transformer State Evaluation Based on Entropy Weight and TOPSIS Method," 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2020:1-5.