Research on Line Loss of Low-Voltage Substation Based on Smart Electric Meter Clustering Algorithm

Yang Fuli
State Grid Chongqing Electric Power Research Institute
No. 80, Middle Section of Huangshan Avenue
Yubei District
Chongqing, China
E-mail: 18680887087@163.com

Hou Xingzhe
State Grid Chongqing Electric Power Research Institute
No. 80, Middle Section of Huangshan Avenue
Yubei District
Chongqing, China
E-mail: cqhhxz@163.com

***Abstract*—At present, electric power company judgment on abnormal line loss is that the line loss is abnormal when the line loss rate exceeds a certain threshold. Yet the judgement is one-sidedness and limited. Aiming at the large proportion of low-voltage distribution network loss in the total loss of power grid, this paper introduces the line loss analysis process, constructs the line loss analysis model based on smart electric meter the four dimensional eigenvalues of daily average line loss rate, line loss rate variation coefficient, three-phase imbalance, power factor abnormality, and puts forward a method of identifying line loss abnormality in low-voltage distribution network based on clustering algorithm. The experimental results show that the method has certain practical application effects, which can accurately analyze the line loss rate and improve the accuracy of line loss anomaly judgment.**

Keywords-Component; Line Loss Rate; Loss Abnormal; Low-Voltage Distribution Network; Clustering Algorithm

I. INTRODUCTION

In the process of power transmission and marketing, the power loss generated by power network enterprises from the outgoing line of power plant to the customer's power meter are called power loss of power grid, referred to as line loss[1] With the continuous growth of electricity sales in the current power grid, the line loss management in the low-voltage distribution network (the power supply scope of the transformer) has been paid more and more attention. The line loss rate is an important indicator reflecting the power grid planning and design, technical equipment level and economic operation level[2].

According to the calculation of theoretical line loss, the line loss rate of 380V power grid (low-voltage distribution distribution network) is the highest, about 6%~10%; according to the calculation of statistical line loss, the loss of low voltage substation area accounts for about 20% of the total loss of power grid. Due to the large number of low-voltage stations and the lack of effective technical means to support line loss management, line loss management is still an important problem to be solved by power companies[3]. The economic and social significance of line loss analysis is inestimable, which can improve the efficiency of energy utilization and improve the economic benefits of the whole power grid. Therefore, the study of line loss has a good theoretical significance and practical application value.

II. LINE LOSS ANALYSIS BASED ON CLUSTERING ALGORITHM

Relying on the power consumption information collection system, the massive data is deeply mined. Taking the line loss rate of low-voltage distribution network and the abnormal events of online monitoring of metering as the analysis object, a line loss analysis method of low-voltage distribution network based on clustering algorithm is proposed. The analysis flow of this method is shown in Figure 1.

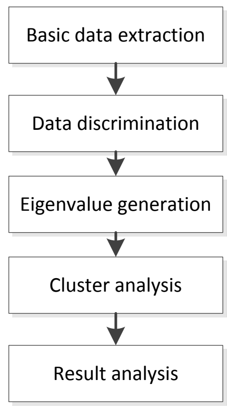

Figure 1. Low-voltage distribution network line loss analysis method flow

A. Basic data extraction

The purpose of line loss analysis in low-voltage substation area is to determine the rationality of distribution system structure operation, and to find out the weak links in distribution system operation, metering device operation and line loss management.

In order to realize the in-depth analysis of the line loss in low-voltage distribution network, relying on the power consumption information collection system, taking the low-voltage distribution network in a certain range (city and county power supply companies) as the research object, taking the daily line loss rate of the station area in the statistical cycle and the abnormal events of online monitoring of measurement as the analysis data, the original data matrix is constructed. Among them, the daily line loss rate of the station area is calculated automatically by the electricity supply and sale collected by the electricity consumption information collection system every day, and the three-phase imbalance and power factor abnormality which have significant influence on the line loss are selected as the measurement abnormal events[4].

B. Data discrimination

The daily line loss rate in low-voltage distribution network is greatly affected by the accuracy of basic files, the success rate of collection and the integrity rate of collected data, so it is necessary to screen the original data and remove the obviously abnormal line loss data. Considering that the size and fluctuation of line loss rate can better reflect the level of line loss in the station area, the variable daily average line loss rate and variation coefficient of line loss rate are defined. The daily average line loss rate in the station area can reflect the average level of line loss rate; the variation coefficient of line loss rate can reflect the dispersion of line loss rate in the station area, and eliminate the influence of the big difference in daily average line loss rate. The calculation formula is as follows:

$$E_x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - E_x\right)^2} \tag{2}$$

$$V_x = \frac{\sigma_x}{E_x} \tag{3}$$

C. Eigenvalue generation

The three-phase imbalance and power factor abnormality are the characteristic values that reflect whether the abnormal metering events occur in the statistical period of the low-voltage distribution network. They are defined as discrete variables, with occurrence recorded as 1 and non-occurrence recorded as 0. Although the daily average line loss rate can reflect the line loss level in the station area, if the k-means algorithm is directly used as the eigenvalue for clustering analysis, the grouping results can’t be fully consistent with the current line loss management indicators. The daily average line loss rate of a city company's low-voltage distribution network is taken as the eigenvalue and K-means[5] algorithm is used for clustering analysis. The grouping results are shown in Table 1 and Figure 2.

TABLE I. CLUSTERING RESULTS

GROUPING	QUANTITY	MINIMUM	MAXIMUM VALUE	PROPORTION
1	8726	2.46%	0.00%	35.86%
2	8690	6.80%	4.60%	35.73%
3	6012	12.02%	9.50%	24.68%
4	293	21.42%	17.01%	1.21%
5	173	34.02%	28.20%	0.72%
6	330	50.00%	41.20%	1.36%
7	50	61.23%	56.18%	0.20%
8	26	92.16%	88.23%	0.10%

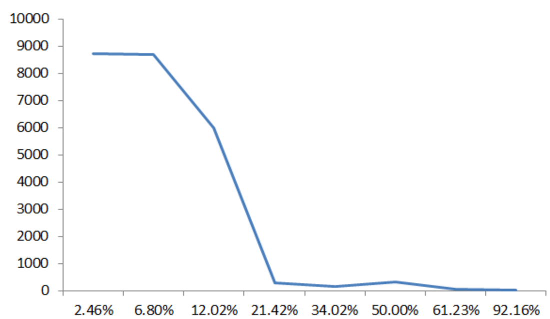

Figure 2. Clustering results

It can be seen from the clustering results that the line loss rate distribution is stepped, and the area above 20% accounts for a small proportion. Refer to the relevant regulations of the State Grid Corporation regarding the monthly linear loss of the station area, and define the daily average line loss rate as a discrete variable. A station with a line loss rate between 0 and 10% is defined as "good", a station with a line loss rate between 10% and 20% is defined as "medium", and a station with a line loss rate greater than 20% is defined as "poor". ". The coefficient of variation of the line loss rate is defined as a continuous variable, and the eigenvalue matrix is constructed together with the daily mean line loss rate, the three-phase unbalance, the power factor anomaly, and the station number.

III. DISCRIMINATION OF LINE LOSS ANOMALY BASED ON CLUSTERING

A. *Design ideas*

This method is improved on the application of traditional clustering algorithm[6-7]. According to the analysis of the actual line loss data, the general line loss rate is generally classified into three categories, high line loss rate, normal line loss rate and low line loss rate[8]. In some cases, the middle class will be closer to the rest of the categories, so it is determined that the number of primary clustering and secondary clustering categories are 3 and 2, respectively. At the same time, the mean value of each cluster is used as the cluster center of this cluster. By comparing the cluster center distance of one cluster result and the number of individuals of each cluster, we can judge whether to carry out secondary clustering. After clustering, the time dispersion of individuals in the categories with high line loss rate is analyzed to determine whether there is abnormal online loss rate. According to the above ideas, a k-means clustering based line loss anomaly detection model is designed, including data preprocessing, clustering analysis, time dispersion analysis, decision-making and so on. The flow is shown in Figure 3.

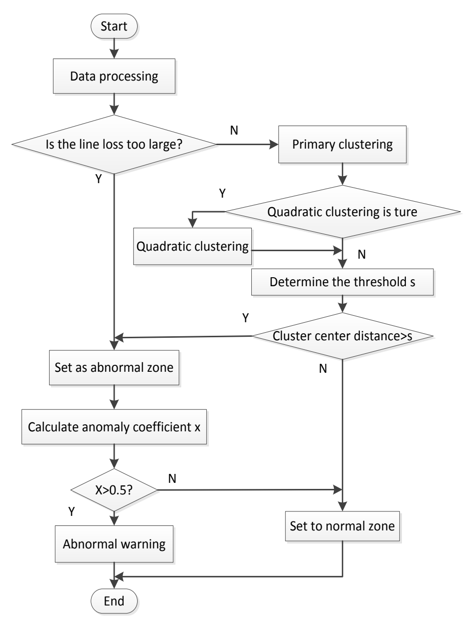

Figure 3. Low chart of line loss anomaly detection based on clustering algorithm

B. Data collection and preprocessing

The selected data is the low-voltage line loss rate data of 50 users in a certain area of a first tier city for four months. Firstly, the redundant data and blank data in the line loss rate data are removed, and then the averaged operation of the processed line loss data is performed. If the average line loss rate is greater than 10%, the line loss is too large, and the station area is determined to be an abnormal station area, and time dispersion analysis is performed; if the average line loss rate is less than 10%, further analysis is performed[8].

C. Cluster partial analysis

The key point of clustering part is to set the threshold value of distance judgment between clustering centers. Combined with the line loss rate data of some line loss abnormal stations that have been verified by the power company at present, on the basis of analyzing the data characteristics of the line loss rate itself, the formula for calculating the distance judgment threshold of clustering center is set as follows:

$$s=\begin{cases}0.5\times a_i, a_i<3\% \\ 3\%, a_i\geq 3\%\end{cases} \tag{4}$$

Where: s is the cluster center distance threshold; a_i is the average value of the station line loss rate.

Based on the case that the line loss rate is less than 10%, when the average line loss rate of the station area is 3% or more, the threshold value of the distance between the cluster centers is set as 3%, that is, the difference between the maximum cluster center and the minimum cluster center is 3%, the actual test effect is the best, and the maximum cluster center is still less than 10%, which is in line with the discussion scope of this method.

D. Decision making

In general, for a zone with a line loss rate of less than 10%, clustering is first performed to obtain three cluster categories. It is verified by analysis that if the quantity of data of a certain type is less than 10, the data in the class is a coarse error, and the class is eliminated, and the second cluster is performed to obtain two cluster categories. Finally, the distance between the cluster centers is compared. If the distance is less than the threshold, the station is determined to be a normal station; if the distance is greater than the threshold, the station is determined to be an abnormal station. In order to quantitatively determine the degree of line loss anomaly in the abnormal zone, the concept of time dispersion is introduced. The time dispersion is the time point corresponding to the line loss rate data of the class with the largest cluster center obtained by clustering in the abnormal line region loss rate data, and the average value of the interval time of these time points is calculated. As an indicator of the degree of line loss anomaly. The formula for time dispersion T_d is as follows:

$$T_d=\frac{1}{n-1}\sum_{k=1}^{n-1}\left(t_{k+1}-t_k\right), k=1,2,...,\mathrm{n}-1 \tag{5}$$

Where: n is the number of line loss rate data for the largest class in the cluster center; t_k is the time corresponding to each line loss rate data.

The smaller the time dispersion is, the more abnormal the online loss rate data is in the time zone of the station. The possibility of line loss anomaly is high in the station area; the time dispersion is larger, the line loss rate data is higher. The occurrence of anomalous occurrences is relatively scattered, indicating that the possibility of line loss anomalies in the station area is low. According to the corresponding data research and actual verification, after one cluster or quadratic clustering, the time dispersion degree is calculated for the line loss rate data of the largest class in the cluster center, and the anomaly coefficient x is obtained. In this paper, 4 months of line loss rate data is used for analysis. In equation (5), when n is arbitrary, the minimum value of T_d is 1, and when n = 10, the maximum value of T_d is about 13.55. The formula for calculating the line loss anomaly coefficient of the station area is:

$$x=1-\frac{T_d}{13.55} \tag{6}$$

Combined with the actual situation and data analysis, it is stipulated that when $x < 0.5$, the abnormal degree of line loss in the station area is low; when $x > 0.5$, the abnormal degree of line loss in the station area is high, and the abnormal alarm of line loss in the area is issued.

IV. RESULT VERIFICATION

This algorithm is applied to the analysis of 9 low-voltage stations in a certain area of a first-line city with abnormal line loss rate caused by power stealing. The line loss rate of 123 days from July 1 to October 31 in each low-voltage distribution network is extracted for analysis, and the following analysis results are obtained (see Table 2)

TABLE II. LINE LOSS ANALYSIS TABLE OF LOW VOLTAGE STATION AREA

Low-voltage distribution network	*Condition of station area*	*Abnormal coefficient of line loss*
A	Abnormal	0.54
B	Abnormal	0.83
C	Abnormal	0.74
D	Abnormal	0.72
E	Normal	
F	Abnormal	0.61
G	Abnormal	0.67
H	Abnormal	0.58
I	abnormal	0.87

The above analysis results show that, except for the zone E because the distance between the cluster centers does not exceed the threshold, the algorithm determines that the zone is normal, and the other eight zones are judged as the line loss abnormal zone, and The higher line loss anomaly coefficient, the discrimination rate is 90%, which proves that the method has certain practical significance.

V. CONCLUSION

Based on the current research results of line loss, combined with the clustering algorithm and the characteristics of low-voltage station line loss rate data, this paper proposes a line-loss anomaly discrimination method based on k-means clustering algorithm. After data acquisition and processing, clustering algorithm analysis and time dispersion analysis, etc., it is finally determined whether there is abnormality and abnormality in the line loss of the station area. This method largely solves the problem of the one-sidedness and limitation of the original line loss abnormality judgment, and has been recognized by the power grid company in the actual test. However, due to the diversity of the line loss problem in low-voltage distribution network, the phenomenon of false positives and false negatives is inevitable. It is necessary to further study in the future, improve the algorithm proposed in this paper, or fuse the algorithm in the paper with other data mining algorithms to further improve the accuracy of abnormal identification.

ACKNOWLEDGMENT

This work was supported by the State Grid Corporation Science and Technology Project "Research on key technologies of risk monitoring and life evaluation of smart meters".

REFERENCES

[1] WANG Tao, ZHANG Jianmin, LI Xiaoping. Calculation of scheduled loss ratio and its assessment [J]. Power System Tech-nology, 2003, 27(7) : 40－42, 55.

[2] TIAN Hongjie. Application of loss analyzing forecast in power supply management [J]. Power System and Control, 2010, 38(7) : 77-80

[3] LIU Jian, DUAN Jingjing. Line loss limitation analysis and optimal planning of loss reduction for distribution grids [J]. Power System Protection and Control, 2013, 41(12): 27-35

[4] K. Elissa, JIANG Huilan, AN Min, LIU Xiaojin. The calculation of energe losses in distribution systems based on RBF network with dynamic clutering algorithm [J]. Proceedings of the CSEE, 2005, 25(10): 35-39.

[5] LI Xueping, LIU Yiran, LU Zhigang. Phase theoretical line loss calculation and analysis based on clustering theory [J]. Transactions of China Electrotechnical Society, 2015, 30(12) : 367－376.

[6] ZHANG Yu. Analysis of line loss management in station area under electricity information acquisition system [J]. Global Market, 2017(9) : 301.

[7] ZHANG Yangrui, YANG Peng, SHEN Hongtao. Ｒesearch of low-voltage distribution network loss based on cluster analysis [J]. Hebei Electric Power, 2017, 36(2) : 8－10.

[8] LI Ya, LIU Lipin, LI Baiqing. Calculation of line loss rate in transformer district based on improved k- means clustering algorithm and BP neural network [J]. Proceedings of the CSEE, 2016, 36(17) : 4543－4551.