Research on Abnormal Signal Identification Algorithm of Distribution Network Operation Based on Big Data Mining

Yang Li[1*], Sheng Liu[2], Ziyi Zhao[3], Lin Yan[4], Zhaohua Hu[5]

1828937@qq.com[1*], 309503096@qq.com[2], 850583975@qq.com[3], yanlinll@163.com[4], 412301613@qq.com[5]

Shenzhen Power Supply Co., Ltd. Shenzhen, Guangdong, 518000, China

Abstract. Conventional algorithm for identifying abnormal signals of distribution network mainly uses ESD (extreme studied deviate test) to obtain signal time series, which is easily influenced by proximity clustering relationship, resulting in poor identification performance. Therefore, it is necessary to design a new algorithm for identifying abnormal signals of distribution network based on big data mining. That is, the abnormal signal of distribution network operation is collected, and the characteristics of abnormal signal of distribution network operation are extracted by big data mining, and the optimization algorithm of abnormal signal identification of distribution network operation is generated, thus realizing the identification of abnormal signal of distribution network operation. The experimental results show that the designed algorithm for identifying abnormal signals of distribution network operation has good recognition performance, high accuracy, recall and F1 score, and it is reliable and has certain application value, which has made certain contributions to reducing the operation risk of distribution network.

Keywords: Big data mining; Power distribution; Power grid; Operation; Abnormal signal; Identification; algorithm.

1 Introduction

With the acceleration of urbanization and the development of economy and society, the scale and complexity of distribution network are increasing, and its development and importance are increasingly prominent. With the growth of power demand and the adjustment of energy structure, the load structure and distribution of distribution network are also changing, so it is necessary to upgrade and transform the distribution network to meet the new demand [1-3] . In addition, how to improve the intelligence and automation level of distribution network, reduce loss and improve efficiency is also a problem that needs attention in the development of distribution network. Under the background of rapid energy structure transformation, the distribution network structure in China is becoming more and more complex [4-6] , which is often affected by abnormal signals, resulting in serious operation accidents. Therefore, effective identification of abnormal signals in distribution network operation is needed.

Identification of abnormal operation signals of distribution network is one of the important tasks in power system. Due to the influence of various factors, such as load fluctuation,

BDEDM 2024, January 12-14, Ningbo, People's Republic of China

DOI 10.4108/eai.12-1-2024.2347156

equipment failure, external interference, etc., the power grid operation may be abnormal, such as voltage fluctuation, current imbalance, frequency deviation, etc. [7-10] . If these abnormal signals are not found and handled in time, it may have a serious impact on the stable operation of power grid and the safety of power equipment. Therefore, it is necessary to design an effective algorithm for identifying abnormal signals in distribution network operation to support the identification of abnormal signals in distribution network. The identification of abnormal signals in distribution network operation mainly includes the following steps. First, it is necessary to collect real-time data of distribution network, including parameters such as current, voltage and power factor [11-13]. These data can be obtained through on-site monitoring equipment or remote monitoring system of power system. Secondly, the collected data need to be preprocessed such as denoising and filtering to remove interference and outliers and ensure the accuracy and reliability of the data. Extract features from the preprocessed data, which can be time domain features, frequency domain features, waveform features, etc. These characteristics can reflect the operation state and abnormal situation of distribution network. Finally, anomaly classification and recognition [14-17] is carried out, and the extracted features are input into classifiers, such as support vector machine, neural network and decision tree, for classification and recognition. By comparing the training samples with the test samples, the accuracy and recall rate of the algorithm are evaluated. According to the classification results and the characteristics of fault types, fault diagnosis and location are carried out. In order to solve the current identification problem, this paper designs a new algorithm for identifying abnormal signals in distribution network operation based on big data mining.

2 Design of abnormal signal identification algorithm for distribution network operation based on big data mining

2.1 Collect abnormal signals of distribution network operation

In order to improve the accuracy of abnormal signal identification, it is necessary to ensure the acquisition quality of the original abnormal signal of distribution network. Therefore, this paper first collects the abnormal signal of distribution network operation [18-20] . Firstly, an abnormal signal set T can be generated according to the coordinate relationship of abnormal signals, as shown in the following (1).

$$T = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\} \tag{1}$$

In the formula (1), $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ are the abnormal signal representative of the distribution network operation collected for the first time, respectively. According to the above-mentioned abnormal signal sets, the fault signals can be linearly converted, and the distribution network abnormal signal collection kernel function x_n can be generated by using the relationship between the horizontal and vertical coordinate sets, as shown in the following (2).

$$x_n = \frac{1}{w}(\frac{1-\theta}{y_n} - D) \quad (2)$$

In the formula (2), w represents an angular variable between elements, θ represents the relaxation factor, y_n represents the transformation coefficient of kernel function, D represents the penalty parameter, according to the above-mentioned kernel function transformation coefficient [21-22] , an effective abnormal signal recognition filter factor $s(k)$ can be obtained, as shown in the following (3).

$$s(k) = ky_{\max} \quad (3)$$

In the formula (3), k represents the number of abnormal signals, $y_{\max}$ represents the maximum abnormal signal value. Based on this, the generated abnormal signal acquisition flow chart is shown in Figure 1 below.

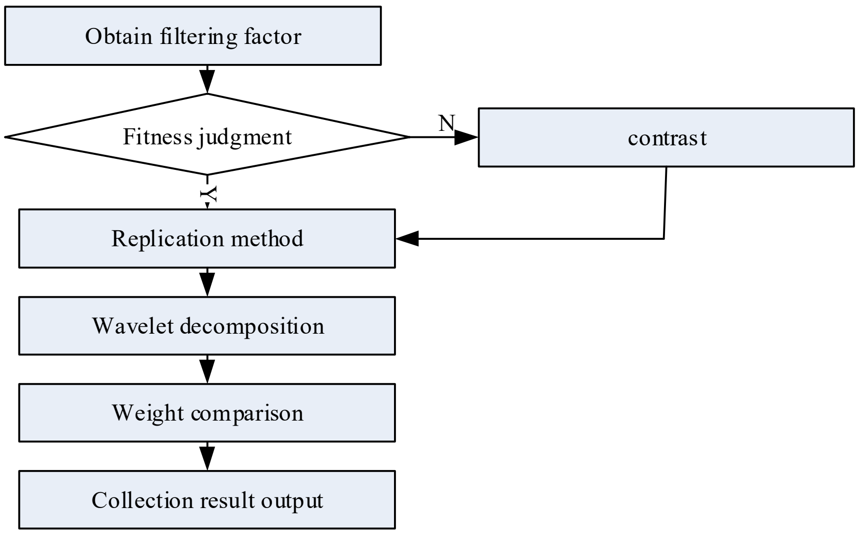

Fig. 1. Flow chart of abnormal signal acquisition in distribution network.

As can be seen from Figure 1, according to the above-mentioned abnormal signal collection flow of distribution network, abnormal signal collection nodes can be effectively arranged, and the quality of collected abnormal signals can be improved to the greatest extent, which can be used as a reference for subsequent abnormal signal feature extraction.

2.2 Based on big data mining, extract abnormal signal characteristics of power grid operation

Big data mining technology is a high-performance data analysis and processing technology, which can obtain the required abnormal features from massive data, and combine the correlation between features for comprehensive mining analysis. Therefore, this paper extracts the abnormal signal features of power grid operation based on big data mining technology. Firstly, we need to denoise the signal and calculate the local mean value of the signal, as shown in the following (4).

$$m_i = \frac{n_i + n_{i+1}}{2} \tag{4}$$

In the formula (4), n_i 、 n_{i+1} respectively represent the time-frequency distribution weights of the signals before and after the estimation, and according to the above-mentioned local mean value of the signals, abnormal signals can be separated, and the original signals are subdivided. At this time, the separated abnormal signals $h(t)$ is shown in the following (5).

$$h(t) = x(t) - m(u) \tag{5}$$

In the formula (5), $x(t)$ represents a status signal, $m(u)$ represents the locally separated signal, after the above steps are completed, the feature extraction feature of abnormal signal in distribution network can be obtained, that is, the collected high-quality signal is decomposed by LMD method, the components containing abnormal features are selected, and then the energy values of different components E are calculated, as shown in the following (6).

$$E = \sum_{p=1}^{n} E_P(t) \tag{6}$$

In the formula (6), $E_P(t)$ represents the energy component representing the main fault information can be used to process the calculated energy values of different signal components. At this time, the abnormal signal characteristics identification processing formula $f(x)$ based on hidden layer neural network can be used, as shown in the following (7).

$$f(x) = \sum B_I G(X) \tag{7}$$

In the formula (7), B_I represents the hidden-layer output matrix, $G(X)$ represents the sample output matrix. According to the above processing procedure, the representative sample output matrix can be used to mine big data, output the extreme learning function, and describe it in a matrix way. By combining the output connection relationship between the hidden layer and the sample layer, the abnormal signal feature recognition result can be obtained, which improves the comprehensive performance of the abnormal signal recognition algorithm.

2.3 Generation of distribution network operation abnormal signal identification optimization algorithm

The identification of abnormal signals in actual distribution network operation can be roughly divided into several steps, that is, generating a similarity matrix according to the sample set, extracting the characteristic vectors of abnormal signals according to the similarity matrix, and standardizing them to obtain the final identification results of abnormal signals. However, the conventional abnormal signal identification algorithm ignores the multidimensional relationship of feature vectors, which is prone to identification distortion and has poor overall performance. Therefore, based on big data mining and the principle of adaptive spectral clustering, this paper uses Euclidean distance to determine the identification relationship of abnormal signals. Based on this, the generated distribution network operation abnormal signal identification optimization algorithm is shown in Figure 2 below.

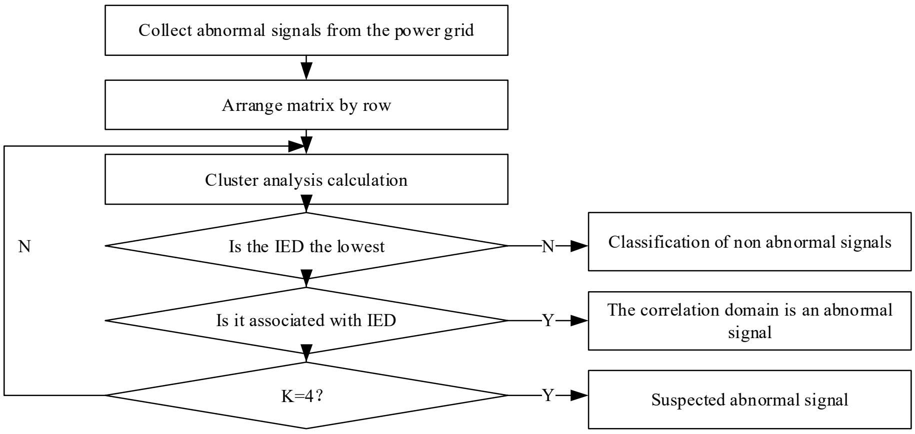

Fig. 2. Optimization algorithm of abnormal signal identification in distribution network operation.

As can be seen from Figure 2, using the above-mentioned optimization algorithm for identifying abnormal signals in distribution network operation can quickly obtain feature vectors, classify them according to the permutation matrix, accurately identify the threshold of abnormal signals, and reduce the possible interference in identifying abnormal signals.

3 Experiment

In order to verify the recognition effect of the designed abnormal signal recognition algorithm for distribution network operation based on big data mining, this paper configures a basic experimental platform, and compares it with the conventional abnormal signal recognition algorithm for distribution network operation based on random matrix and the abnormal signal recognition algorithm for distribution network operation considering strategy fusion, and carries out experiments as follows.

3.1 Experimental preparation

According to the performance verification and identification requirements of the abnormal signal identification algorithm for distribution network operation, this paper selects Lab Windows/CVI as the experimental platform to carry out the performance verification experiment. First, it is necessary to collect the abnormal data of distribution network operation with a data acquisition card, which is relatively large and has many types. Therefore, this paper selects SK2010 data acquisition card to perform the collection task and carry out effective interface matching. At this time, an effective experimental data collection process can be generated, as shown in Figure 3 below.

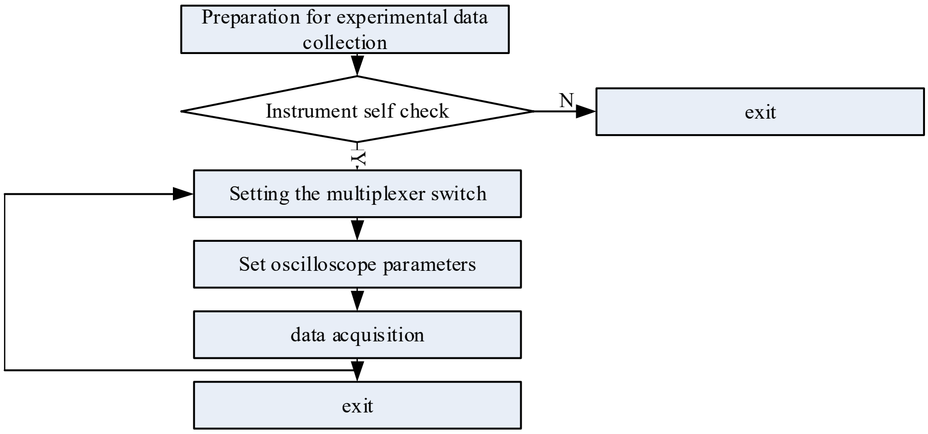

Fig. 3. Experimental data acquisition process.

As can be seen from Figure 3, this experiment mainly uses multi-channel signal acquisition method to collect experimental data signals. In order to meet the signal measurement and receiving requirements of the experimental platform, this paper uses PXI software to preprocess the signals, converts the basic receiving frequency of the signals, and installs brand-new Active experimental components. At this time, the experimental self-checking database can be generated. The attributes and data type parameters of different types of abnormal signals are shown in Table 1 below.

Table 1. Database Parameters of Experimental Self-inspection.

Attribute	Data type	Data examples	Source	Structured data
Serial Number	INT	5	automatic	Y
area	VARCHAR	XX market	OMS	Y
Factory Station	VARCHAR	XX stand	D5000	Y
protection equipment	VARCHAR	busbar protection	D5000	Y
Alarm element	VARCHAR	intelligent terminal	D5000	Y
Alarm content	VARCHAR	Device abnormality	D5000	Y
Alarm time	TIME	2020.1.6/22：33	D5000	Y
Faulty component	VARCHAR	terminal	OMS	N
Action situation	VARCHAR	maloperation	OMS	Y

According to the data parameters in Table 1, the abnormal signals are stored, and the structured data attributes extracted from the experimental database are Region, Factory and Alarm Time. At this time, the recognition accuracy , recall rate can be selected, and F values are used as experimental preparations, and the calculation formulas are shown in the following (8)~(10).

$$P=\frac{\sum\left|R(u)\cap T(u)\right|}{\sum\left|R(u)\right|} \tag{8}$$

$$R=\frac{\sum\left|R(u)\cap T(u)\right|}{\sum\left|T(u)\right|} \tag{9}$$

$$F=\frac{2PR}{P+R} \tag{10}$$

In formulae (8) to (10), $R(u)$ represents the actual accurate identification signal, $T(u)$ represents a predicted abnormality recognition signal. The higher the recognition accuracy, recall and F value represents the predicted abnormal signal recognition, it proves that the recognition performance of the abnormal signal recognition algorithm is better, on the contrary, it proves that the recognition performance of the abnormal signal recognition algorithm is relatively poor.

3.2 Experimental results and discussion

According to the above experimental preparation, the performance verification of distribution network abnormal signal identification algorithm can be carried out, that is, the distribution network abnormal signal identification algorithm based on big data mining, the conventional distribution network abnormal signal identification algorithm based on random matrix and the distribution network abnormal signal identification algorithm considering strategy fusion are used to identify different operation abnormal signals respectively, and the identification performance of the three algorithms is calculated by formulas (8)~(10). The experimental results are shown in the following Table 2.

Table 2. Experimental results.

Identification signal category number	The recognition accuracy of the distribution network abnormal signal recognition algorithm based on big data mining designed in this article (%)		Identification accuracy of distribution network abnormal signal recognition algorithm based on random matrix (%)		Recognition accuracy of abnormal signal recognition algorithm for distribution network operation considering strategy fusion (%)	
	data	Add abnormal samples	data	Add abnormal samples	data	Add abnormal samples
the distribution network abnormal signal identification algorithm based on big data mining						
FA01	98.747	98.747	65.157	68.185	65.418	65.847
FA02	97.742	94.841	61.251	64.148	64.275	54.412
FA03	99.384	95.285	64.368	65.355	62.417	62.328
FA04	98.245	95.475	76.545	62.276	66.286	53.546
FA05	95.685	96.369	68.269	51.985	59.385	56.284
FA06	96.173	94.854	75.875	66.175	65.475	59.854
FA07	99.248	94.246	76.478	79.124	64.147	58.226
FA08	98.256	95.278	69.145	78.286	65.269	65.841
FA09	97.584	98.478	67.352	67.284	52.278	64.233
FA10	95.254	97.254	78.986	78.278	63.465	65.578

the conventional distribution network abnormal signal identification algorithm based on random matrix						
FA01	95.485	95.478	65.841	65.586	55.845	65.175
FA02	98.175	96.243	66.253	67.142	58.698	68.475
FA03	94.328	99.286	75.965	58.398	64.785	64.398
FA04	95.246	98.278	79.369	54.655	56.243	56.646
FA05	96.963	94.174	78.574	56.727	52.286	52.752
FA06	99.542	95.283	74.257	62.245	63.896	53.475
FA07	98.175	92.238	78.284	52.688	69.246	69.398
FA08	97.246	94.145	66.276	63.576	65.278	55.585
FA09	95.386	97.278	72.298	55.273	66.698	67.139
FA10	99.954	98.289	76.765	59.284	57.527	76.525
the distribution network abnormal signal identification algorithm considering strategy fusion						
FA01	95.125	97.418	65.185	66.185	79.418	72.486
FA02	98.741	95.243	68.136	72.475	76.248	72.246
FA03	94.253	98.968	64.274	73.368	74.778	63.486
FA04	96.845	95.571	65.285	75.246	65.323	76.574
FA05	99.456	94.287	72.332	64.452	63.418	79.274
FA06	98.274	93.653	73.964	74.486	72.414	75.381
FA07	97.283	92.296	76.248	62.268	65.143	77.687
FA08	95.957	96.475	79.256	78.346	74.986	75.386
FA09	96.178	99.143	68.246	74.262	65.848	64.456
FA10	95.268	98.286	67.285	75.546	78.176	61.274

As can be seen from Table 2, the recognition accuracy, recognition recall and recognition F value of the distribution network operation abnormal signal recognition algorithm designed in this paper are high under different types of abnormal signals, while the recognition accuracy, recognition recall and recognition F value of the conventional distribution network operation abnormal signal recognition algorithm based on random matrix and the distribution network operation abnormal signal recognition algorithm considering strategy fusion are relatively low under different types of abnormal signals. The above experimental results prove that the abnormal signal identification algorithm of distribution network based on big data mining designed in this paper has good performance, reliability and certain application value.

4 Conclusion

Distribution network is an important part of power system, which undertakes the important task of transmitting power from power generation side to user side. The identification of abnormal signals in distribution network operation is of great significance in ensuring the stable operation of power system, improving the operation efficiency of power equipment, optimizing the planning and construction of power system, ensuring social economy and people's life, preventing and reducing accidents and improving the service level of power supply. This paper introduces the methods and steps of identifying abnormal signals in distribution network operation, including data acquisition, data preprocessing, feature extraction, abnormal classification identification and fault diagnosis. At the same time, this paper also introduces the importance and significance of abnormal signal identification of distribution network operation, as well as the existing problems and future development trends, and designs a brand-new algorithm for abnormal signal identification of distribution network operation based on big data mining. The experimental results show that the designed algorithm based on big data mining has good performance, reliability and certain application value, and has made certain contributions to improving the operation reliability of distribution network.

References

[1] Yang, J. J., Cheng, R. F., & Xu, L. B. (2022). Multi-objective optimization simulation of active distribution network considering new energy access. Comput. Simul, 39, 108-113.

[2] Yang, K., & Wang, X. (2021). Abnormal identification of lubricating oil parameters and evaluation of physical and chemical properties based on machine learning. In IOP Conference Series: Materials Science and Engineering (Vol. 1043, No. 5, p. 052053). IOP Publishing.

[3] Zhou, L., Li, Q., Zhang, Y., Chen, J., Yi, Y., & Liu, S. (2021). Consumer phase identification under incomplete data condition with dimensional calibration. International Journal of Electrical Power & Energy Systems, 129, 106851.

[4] Jeong, G., Lim, G., & Kang, D. (2021). Identification of unintended isolation segments in water distribution networks using a link-by-link adjacency matrix. Journal of Water Resources Planning and Management, 147(2), 06020013.

[5] Lin Y , Yan B , Gu D. (2021). Research on Identification Algorithm of Fragile Nodes: Based on Static and Dynamic Combination of Uncertainty in New Energy.Electronics Science Technology and Application, 8(1):14.

[6] Lim, X. Y., Gan, K. B., & Abd Aziz, N. A. (2021). Deep ConvLSTM network with dataset resampling for upper body activity recognition using minimal number of IMU sensors. Applied Sciences, 11(8), 3543.

[7] Tsvetkov, E. I., & Suloeva, E. S. (2021, May). Metric Identification for Random Value Density Distribution. In Journal of Physics: Conference Series (Vol. 1864, No. 1, p. 012092). IOP Publishing.

[8] Jiaren Y U , Tian Y , Lin H.(2022). Design of miner type identification mechanism based on reputation management model.Chinese Journal of Network and Information Security, 8(1):128-138.

[9] Xie Y , Zhang Y. (2021). Electricity meter area identification technology based on channel coding theory.Chinese Journal of Internet of Things, 5(4):137-144.

[10] Khan, A. I., Quadri, S. M. K., & Banday, S. (2021). Deep learning for apple diseases: classification and identification. International Journal of Computational Intelligence Studies, 10(1), 1-12.

[11] YixuanGUO,XiangdongJIA,ShengnanCAO,et al. (2022).Outage Performance Evaluation of Drone HetNet Based on Spatial Poisson Point Process.Journal of Signal Processing, 38(3):536-542.

[12] Lyu, Z., Pei, X., Yang, Y., & Li, J. (2021). General small-signal modeling and parameter optimization of the soft normally-open point connected to the distribution network. IEEE Access, 9, 112573-112584.

[13] Duarte, S. N., Ghetti, F. T., de Almeida, P. M., & Barbosa, P. G. (2021). Experimental evaluation of negative-sequence voltage compensation in distribution networks by a modular multilevel static synchronous compensator. Electric Power Systems Research, 194, 107020.

[14] Gupta, S., Agrawal, M., & Deepak, D. (2021). Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning. Biomedical Signal Processing and Control, 70, 102947.
[15] Song, K., Park, T. J., & Chang, J. H. (2021). Novel data augmentation employing multivariate gaussian distribution for neural network-based blood pressure estimation. Applied Sciences, 11(9), 3923.
[16] Mieński, R., Urbanek, P., & Wasiak, I. (2021). Using Energy Storage Inverters of Prosumer Installations for Voltage Control in Low-Voltage Distribution Networks. Energies, 14(4), 1121.
[17] Zhao, R., Xia, Y., & Zhang, Y. (2021). Unsupervised sleep staging system based on domain adaptation. Biomedical Signal Processing and Control, 69, 102937.
[18] Ilgaz, M. A., & Batagelj, B. (2021). Opto-electronic oscillators for micro-and millimeter wave signal generation. Electronics, 10(7), 857.
[19] Lee, S. H. (2021). Classification of epileptic seizure using feature selection based on fuzzy membership from EEG signal. Technology and Health Care, 29(S1), 519-529.
[20] Maričić, S., Milošević, N., Drajić, D., Milić, D., & Anastasov, J. (2021). Physical Layer Intercept Probability in Wireless Sensor Networks over Fisher–Snedecor $\mathcal{F}$ Fading Channels. Electronics, 10(12), 1368.
[21] Yan, L., Zheng, W., Wang, F. Y., & Gou, C. (2021). Joint image-to-image translation with denoising using enhanced generative adversarial networks. Signal Processing: Image Communication, 91, 116072.
[22] Yue, B., Chen, L., & Shi, H. (2022). Ship Detection in SAR Images Based on Improved RetinaNet. J. Signal Process, 38, 128-136.