Research on Fault Monitoring Technology of Distribution Network Based on Fuzzy Association Rules Mining

Wei Zhang1,*, Yumin Liu1, Zhu Liu1, Shilin Wang1, Zijing Zhen1 and Yixin Li2
1State Grid Information and Telecommunication Group, Changping District, Beijing, China
2North China Power Engineering Co., Ltd. of China Power Engineering Consulting Group, Xicheng District, Beijing, China

*Corresponding author email: zhangwei1@sgitg.sgcc.com.cn

Abstract. In order to monitor the operation status of distribution network in real time and ensure the security and stability of distribution internet of things, a fault monitoring method based on fuzzy association rule mining was proposed. Fuzzy association rules mining was used to build the normal state parameter model of distribution network. The degree of interest index replaced the traditional index, aiming to improve the association degree between the rule parameters and form an effective expert knowledge base. According to the rules in the expert knowledge base, the normal state data of distribution network were obtained by fuzzy reasoning. Further, the state similarity function between the monitoring states and the normal states was used as the state judgement index. When the similarity is lower than the threshold, the fault alert should be triggered. Taking the single-phase ground fault data of distribution network simulated by MATLAB Simulink as an example, it was verified that the proposed method can effectively monitor the single-phase ground fault.

Keywords: Distribution network; Intelligent distribution terminal; Fuzzy association rule mining; Fault monitoring.

1. Introduction

The distribution network is located at the end of the power transmission system, which directly distributes power to the consumers. Thus, its reliability directly affects the satisfaction with electricity consumption. According to the statistical analysis of the causes of the power shortage and interruption, more than 90% letting off of customers is leaded by distribution system fault. Therefore, improving the safety and reliability of distribution network is an important indicator of power grid capacity.

Based on the current demand of distribution network development, building the distribution Internet of things is an essential and necessary improvement of distribution network. Many regions in China take the intelligent distribution terminal as the starting point to explore and practice the distribution Internet of things technology. The intelligent distribution terminal and its general software platform have been built [3]. Further, the intelligent distribution terminal has strong local computing processing and service capabilities, so in theory, the intelligent distribution terminal has unlimited expandable functions [4]. The State Grid takes the intelligent distribution terminal as the application platform, with the help of the edge computing technology, the real-time business and intelligent application of distribution network are deployed in the form of functional software app [1]. Through the real-time monitoring of the intelligent distribution terminal, it can provide scientific basis for the state monitoring and power grid planning, and improve the reliability and operation level of the distribution network [5].
At present, there are some limitations in the research of fault monitoring for the safety and reliability of distribution network: First, the monitoring efficiency is low. Most of the existing technologies need to send the collected data information to the cloud server by the data acquisition terminal, and then through the server calculation and judgment [6,7]. This will reduce the efficiency of fault handling in distribution network, and a large number of fault information in distribution network area will increase the workload of cloud server and occupy communication resources. This limitation highlights the significance of the edge computing technology in the power Internet of things. The distribution terminal unit can realize the functions of monitoring, storage and calculation near the data source by using the edge computing technology, which will effectively improve the response speed of distribution network monitoring and the accuracy of fault diagnosis, reduce the workload of upper master station and the excessive use of communication resources. Second, the fault judgment standard is extensive. The existing traditional fault judgment standards mainly include: the distribution network current and voltage data threshold method [8,9], the frequency transformation analysis judgment method [10,11], the matrix theory statistical index method [12,13], the multivariate state analysis judgment method [1]. Compared with the fault threshold obtained by real-time data mining, the fixed reference value may reduce the sensitivity of fault monitoring in different fault conditions, to some extent, it will weaken the effect of monitoring and alarm.

This research mined the fuzzy association rules of the normal operation state data of the distribution network, and established the association rule expert knowledge base and the normal state parameter inference model. Furthermore, the terminal real-time monitoring parameters of distribution network and normal state parameters were used to calculate the similar function value function. When the value of similar function is lower than the mining standard value, it means that the state parameters of the distribution network area deviate from the normal state, and an alarm should be triggered. Generally, in off-line part, training and learning were carried out according to the actual state parameters of the distribution network area; in on-line part, the software app deployed in the intelligent distribution terminal, which greatly improved the rapidity, adaptability and accuracy of fault monitoring.

2. System Structure

The proposed distribution network fault monitoring system includes the following modules: Module 1, the distribution terminal platform and database; Module 2, the data sampling, selection and preprocessing; Module 3, the fuzzy association rule mining and expert knowledge base; Module 4, the state parameter similarity function value judgment. The structure flow chart is shown in Figure 1.

![Figure 1. Structure flow chart.](image-url)

The fault monitoring system of the distribution network was divided into the offline part and the online part: the offline part was based on the operation monitoring data of the distribution network, the mining algorithm of fuzzy association rules was used to establish the inference model of the normal state parameter, and the above inferred normal state parameter was the basis for the study and judgment of the distribution network state; in the online part of the fault monitoring system, the similarity function value between the real-time monitoring state and the inferred normal state parameter was calculated. When the similarity was lower than the fault alert threshold value, it indicated that the current monitoring state deviated from the normal state, and the alarm should be triggered, so as to realize the monitoring and maintenance of distribution network.
3. Method Introduction

The normal historical data from distribution terminal platform and database should be sampled and preprocessed offline. According to the sampling time of 50 μs, taking the local time of the terminal as the benchmark, the relevant data group of variables in distribution network was sampled, including: the three-phase voltage, the current, the active power, the reactive power, the frequency, the voltage and current imbalance, the frequency deviation, the load rate and other parameters at the low-voltage side of the distribution transformer. In order to ensure the accuracy and representativeness of data, a data preprocessing method based on K-means algorithm [14] was used to clean the data samples and eliminate outliers caused by measurement errors, sensor failures and environmental changes.

3.1. Fuzzy Association Rule Mining

After off-line preprocessing, the normal state parameters of distribution network were input into the fuzzy association rule reasoning model. The core of this model is to use the historical data of normal state parameters to mine the relationship between the various state parameters of distribution network, that is, fuzzy association rules. The rule table with the fuzzy levels corresponded to the relationship between the normal parameters of distribution network in different states. In this study, this relationship was expressed by conditional statements, thus the following $X_1$-$X_n$ is the parameters of different distribution network states at the same time. If $X_1$ is a range level value in the domain, the distribution network state parameters, except $X_1$, correspond to the range level of the specific domain according to the association rules.

The traditional fuzzy association mining methods usually take the minimum support $s_{\text{min}}$ and the minimum confidence $c_{\text{min}}$ as the selection criteria, and their expressions are as follows:

$$s(X_1 \Rightarrow X_{2\sim n}) = s(X_1 \cup X_{2\sim n}) = \sum_{i=1}^{n} \mu(X_1 \cup X_{2\sim n}) / |D|$$  \hspace{1cm} (1)

$$c(X_1 \Rightarrow X_{2\sim n}) = s(X_1 \cup X_{2\sim n}) / s(X_1) = \frac{\sum_{i=1}^{n} \mu(X_1 \cup X_{2\sim n}) / \sum_{i=1}^{n} \mu(X_1)}$$  \hspace{1cm} (2)

Where $\mu(X)$ is the membership function of $X$; $D$ is the total number of transactions in the data set. The purpose of data mining is to find out the credible and representative rules. The minimum support $s_{\text{min}}$ and the minimum confidence $c_{\text{min}}$ specify the thresholds of support and confidence. They respectively specify the minimum support and confidence that must be achieved when the association rules are established, and their expressions are as follows:

$$s(X_1 \cup X_{2\sim n}) \geq s_{\text{min}}, c(X_1 \cup X_{2\sim n}) \geq c_{\text{min}}$$  \hspace{1cm} (3)

In practice, under the framework of support and confidence rule mining, a large number of redundant rules contradict each other, resulting in the failure of reasoning and decision-making. Therefore, the traditional screening criteria of fuzzy association rules can not ensure that all association rules are valuable, and even the rules are deceptive [15].

A large number of parameters in the distribution network were distributed and regular. The concept of degree of interest [15] was introduced to screen the most valuable rules when mining fuzzy association rules. Under its guidance, the correlation of parameters was directly proportional to the value of degree of interest. The expression is as follows:

$$I(X_1 \cup X_{2\sim n}) = \frac{1 - s(X_2\sim n)}{[1 - s(X_1)] \times [1 - s(X_1 \cup X_{2\sim n})]}$$  \hspace{1cm} (4)

Based on this, the fuzzy association rules with the maximum degree of interest were selected to form the expert knowledge base, which provided the rule knowledge and experience for fuzzy reasoning, thus reasoning and judging. The detailed construction steps are as follows:
Step 1: Determine the distribution network parameters $X_1-X_n$, and fuzzy the historical data of $X_1-X_n$ after sampling and preprocessing. The fuzzy domain was set to [-9,9], and the number of fuzzy sets was set to 9. The corresponding language variables were NB (negative big), NM (negative middle), NS (negative small), NZ (negative zero), Z (zero), PZ (positive zero), PS (positive small), PM (positive middle) and PB (positive big), which were expressed by trigonometric membership function.

Step 2: According to the fuzzy data obtained from a large number of historical data in step 1, the initial fuzzy association rule base $A_1$ of distribution network state parameter $X_1-X_n$ was established.

Step 3: The same fuzzy association rules between $X_1-X_n$ in $A_1$ were merged, and the fuzzy association rule base $A_2$ was obtained. According to formula (1) and (2), the minimum support $s_{\text{min}}$ and the minimum confidence $c_{\text{min}}$ corresponding to each fuzzy association rule in $A_2$ were calculated.

Step 4: According to the traditional rule selection criteria, the minimum support $s_{\text{min}}$ and confidence $c_{\text{min}}$ are set. According to formula (3), the fuzzy association rules in $A_2$ which are lower than $s_{\text{min}}$ and $c_{\text{min}}$ were deleted, and the fuzzy association rule base $A_3$ was sorted out.

Step 5: According to equation (4), the degree of interest of different fuzzy association rules were calculated, the fuzzy association rules with the largest degree of interest in $A_3$ were selected, and a fuzzy association rule base $A_4$ was formed. The final rule base $A_4$ was the expert knowledge base in Module 3.

Step 6: The fuzzy association rules in the expert knowledge base were built into the fuzzy controller to get the fuzzy value output of the state parameters in the normal parameter domain. The basic flow of fuzzy logic controller for fuzzy reasoning is shown in Figure 2.

Step 7: The output of the fuzzy controller was processed by the defuzzification algorithm. The inferred data of different normal state parameters was obtained and imported into the XML file, which was used to configure the software in the intelligent distribution terminal platform.

3.2. Calculation of State Similarity Function

According to the algorithm flow, the judgment of state similarity function value Module (Module 4), was used to compare the difference between the real-time monitoring state and the inferred normal state, so it is necessary to build the state similarity function. The state similarity function read the normal state of distribution network in the XML configuration file in Module 3 as the judgment index, received the real-time monitoring state data of distribution network in Module 2 as the input, and output the state judgment results to the distribution terminal platform and data base in Module 1 in real time after judgment. The Euclidean distance, which is the most commonly used in vector space distance, is used to construct the state similarity function [16]. The Euclidean distance formula is:

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2}$$  

(5)

Where, $d_{ij}$ represents the Euclidean distance between the real-time monitoring state group $x_i$ and the corresponding normal state group $x_j$ in the XML file; $n$ is the state group dimension; $k$ represents the parameter number in the state group; $x_{ik}$ and $x_{jk}$ respectively represent the $k$ parameter element corresponding to the normal state group $x_i$ and the monitoring state group $x_j$ of the distribution network. In engineering applications, $S_{ij} = 1/(1+d_{ij})$ is often used to standardize the Euclidean distance range to [0,1] [16]. Therefore, the state similarity function used to evaluate the normal state group $x_i$ and the monitoring state group $x_j$ was set as follows:
\[ S(x_i, x_j) = \frac{1}{1 + \sqrt{\sum_{l=1}^{n} (x_{il} - x_{jl})^2}} \]  

(6)

The historical data of distribution network under massive normal state were sampled, the minimum state similarity function between the inferred fuzzy normal state groups and the historical normal state groups was defined as \( S_m \), and the product of minimum state similarity function value \( S_m \) and threshold coefficient \( k_t \) was set as alert threshold \( S_a \), the formula of \( S_a \) is as follows:

\[ S_a = k_t \cdot S_m \]  

(7)

The threshold coefficient \( k_t \) was set according to the monitoring sensitivity of the distribution network. When the state similarity function value of the monitoring states is lower than the threshold \( S_a \), the monitoring system should send out an alarm signal.

According to the above algorithm, C++ programming process was used for embedded programming development. After compiling the program on Linux platform, the functional app was obtained and used to cooperate with the XML configuration file in Section 3.2. With the terminal debugging tool, the function app was transmitted to the distribution terminal equipment. When the fault occurs, the function app should report the alarm signal to the terminal platform through the interactive interface, and finally realize online fault monitoring.

4. Simulation Verification

4.1. Simulation Case Introduction

The single-phase ground fault was selected as a simulation example to verify the performance of fault monitoring system. For comparison, the simulation fault parameters were set to be the same as those in Reference [1]. Using the Simulink tool of MATLAB software to build the simulation system model, the model is composed of the three-phase power supply, the main transformers, the transmission lines, the distribution transformers and the user load. The parameters of the main transformer and the distribution transformer are shown in Table 1. The parameters and length of the six transmission lines are shown in Table 2 and Table 3 respectively, and the load is three-phase symmetrical load.

| The main transformers | The distribution transformers |
|-----------------------|-------------------------------|
| Wiring mode           | YY0                           |
| Rated capacity        | 31.5 MV·A                     |
| Primary voltage       | 110 kV                        |
| Secondary voltage     | 10 kV                         |

| The main transformers | The distribution transformers |
|-----------------------|-------------------------------|
| Rated capacity        | 1 MV·A                        |
| Primary voltage       | 10 kV                         |
| Secondary voltage     | 400 V                         |

Table 1. Parameters of main transformer and distribution transformer.

| Overhead line | Cable line |
|---------------|------------|
| Positive sequence resistance/(Ω/km) | 0.17 | 0.078 |
| Positive sequence to ground inductance/(mH/km) | 1.21 | 0.270 |
| Positive sequence earth mobility/(μF/km) | 9.70 | 695.000 |
| Zero sequence resistance/(Ω/km) | 0.23 | 0.106 |
| Zero sequence to ground inductance /(mH/km) | 5.48 | 1.223 |
| Zero sequence earth mobility/(μF/km) | 6.00 | 358.000 |

Table 2. Parameters of transmission line.

| Overhead line/km | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------|---|---|---|---|---|---|
| Cable line/km    | 0 | 0.5 | 1 | 0 | 0 | 0.5 |

Table 3. Length of transmission line.
In the simulation case, 58 process variables were selected for monitoring state parameters, including the three-phase voltage of the secondary side of the main transformer, the three-phase current of each transmission line, the three-phase voltage and current of the secondary side of each distribution transformer, and the zero sequence current of the fault line. The simulation time was set to 1 s, the sampling time was set to 50 μs, and the single-phase ground fault occurred at 0.8 s. Thus, the simulation data volume was 20000 groups of 58-dimension data samples. Besides, 14000 groups of samples in the first 0.7 seconds were intercepted as training data, and 6000 groups of training samples in the last 0.3 seconds were used as test data, so the 2001 point in the test data started to contain fault information. The training data and test data were input into the distribution network fault monitoring system for monitoring.

4.2. Reasoning Model Construction
Before the establishment of the model, the historical data of the distribution network was cleaned and fuzzed. In the data domain, the triangular membership function was selected to improve the control sensitivity. As shown in Figure 3, \(c_1-c_9\) were obtained as the cluster center through K-means clustering algorithm, and they were set as the data fuzzy set center.

![Membership function diagram.](Figure 3)

After the fuzzy set of each variable was determined, according to the triangular fuzzy membership function, the membership of different monitoring variables in fuzzy sets was calculated and finally transformed into fuzzy data. The fuzzy vector, membership degree and initial expert knowledge rule base can be obtained by fuzzy level division and membership degree calculation. According to the mining steps of fuzzy association rules in Section 2.2, the minimum support \(s_{\text{min}}\) and the minimum confidence \(c_{\text{min}}\) were set to be 0.45 and 0.75 respectively. After calculation and selection, the most representative fuzzy association rules were selected, and the expert knowledge base of the distribution network condition monitoring system was established. Thus, the fuzzy association rules among 58 variables in the single-phase ground fault simulation example were finally obtained according to the measurement standard of the maximum degree of interest. Due to space limitation, only the fuzzy association rule table of 9-dimension variables is displayed, as shown in Table 4.

| Variable | NB | PS | PS | NB | PS | PS | NB | PS | PS |
|----------|----|----|----|----|----|----|----|----|----|
| NB       |    |    |    |    |    |    |    |    |    |
| NM       |    | PB | NB | NS | PB | NS | NB | PB | NB |
| NS       | PB | NM | NM | PB | NM | NM | PB | NB | NB |
| NZ       | PB | NS | NZ | PB | NS | NZ | PB | NS | NS |
| Z        | PM | NS | Z  | PM | NB | Z  | PM | NS | NS |
| PZ       | PM | NB | PZ | PM | NB | PZ | PM | NB | NB |
| PS       | PS | NB | PS | PS | NB | PS | PS | NB | NB |
| PM       | PZ | NB | PZ | NB | PM | PZ | NB | PZ | NB |
| PB       | NM | NM | PB | NM | NM | PB | NM | NM | NM |
4.3. Judgment of State Similarity Function Value

The state similarity function between the actual monitoring state and the inferred normal state was mainly based on the Euclidean distance between the state vectors. The real-time difference between the states represented the fault information of the equipment parameters. However, in the actual operation of the distribution network, it is difficult to ensure 100% accuracy of the sampling data due to sensor measurement errors and environmental changes. Thus, the distribution network state fluctuation caused by such non-trend short-term deviation should not be identified as fault state. Therefore, the elimination of accidental fluctuation error should be considered in the study of state judgement. In this paper, the sliding window method was used to process the data, and the trend of the state similarity function value was found by calculating the moving average of data by updating data.

The width of sliding window was set to 100, the minimum state similarity function value \( S_m \) was 0.726, and the alert threshold coefficient \( k_t \) was set to 0.85, so the final alert threshold \( S_t \) was 0.597. The fault monitoring result of distribution network is shown in Figure 4.

In the test data, the real fault occurred at point 2001, and the proposed state monitoring system triggered an alarm after point 2032 according to the monitoring results in Figure 4, thus the alarm time was only 0.0016 s away from the fault occurrence time. In the case of the same simulation parameters, the proposed monitoring method was compared with the method in Reference [1], and the compared results are shown in Figure 5.

In Figure 5, the state "1" represents the normal state and the state "0" represents the monitoring fault state. It can be seen from Figure 5 that the method in Reference [1] repeatedly presents unstable alarm, but the effective fault alarm time should be defined as the last state change time before the monitoring state is stable. Therefore, compared with the fault detection time, the effective alarm time of the results of Reference [1] is 0.0025s, and the effective alarm time of this study is 0.0016s, so the proposed method is faster. In terms of accuracy, the monitoring results in Reference [1] have different unstable alarm, but the monitoring results in this study are more stable and accurate. In general, the fault monitoring system proposed in this study has excellent performance in rapidity and accuracy.
5. Conclusion
In this study, a fault monitoring system based on the mining of fuzzy association rules and the index of similar function values was proposed. The system used the degree of interest index instead of the traditional indexes as rule selection basis, the expert knowledge of fuzzy association rules representing the normal state of distribution network was selected. Besides, the fuzzy controller was guided to get the inferred normal state parameters of distribution network. By constructing the state similarity function, the system can effectively evaluate the fault state of distribution network. Compared with the existing methods, the proposed fault monitoring system has better performance in monitoring alarm time and fault judgment accuracy. The simulation results showed that the monitoring system based on intelligent distribution terminal platform can effectively use edge computing technology to improve the monitoring response speed and the judgment accuracy of distribution network.

Acknowledgments
This project is supported by the Science and Technology Project of State Grid Headquarters (Project No. 5700-202041163A-0-0-00). The authors would like to thank the company and colleagues for the support. In addition, the authors expect positive criticism and suggestions.

References
[1] Shilin Wang, Zhu Liu, Zhe Zhang, et al. Research on single phase grounding fault detection technology of distribution network based on intelligent variable terminal. Power System Technology, 2019, 43(12): 4291-4298.
[2] Jun Lv, Wenpeng Luan, Riliang Liu, et al. Architecture of distribution internet of things based on widespread sensing & software defined technology. Power System Technology, 2018, 42(10): 3108-3115.
[3] Jichuan Zhang, Lei Chen, Mingyu Zhang, et al. Conception and application of smart terminal for distribution internet of things. High Voltage Engineering, 2019, 45(6): 1729-1736.
[4] Xinhua Xiao, Mingye Huang, Zhiqiang Song. Research on construction of intelligent transportation inspection system for low-voltage distribution network based on intelligent distribution terminal. Distribution & Utilization, 2019, 36(6): 55-61.
[5] Sifang Liu, Chunyu Deng, Guobin Zhang, et al. Research on collaborative architecture for edge computing of residential intelligent usage of electricity. Electric Power Construction, 2018, 39(11): 60-68.
[6] Hong Xiao, Di Zhou, Jing Tang, et al. Application and discussion of the measurement terminal in the monitoring of low voltage line in power grid. Smart Grid(Hans), 2018, 8(2): 105-110.
[7] Chuqiao Lin, Hui Tong, Wenfang Yu, et al. Distribution grid intelligent state monitoring and fault handling platform based on ubiquitous power internet of things. Journal of Northeast Dianli University, 2019(4): 1-4.
[8] S Hanninen, M Lehtonen. Method for detection and location of very high resistive earth faults. European Transactions on Electrical Power, 1999, 9(5): 285-291.
[9] Yongrui Xue, Bingyin Xu, Zuren Feng. Fault current detection method using measurement current transformer for distribution automation. Automation of Electric Power Systems, 2001(21): 51-54.
[10] Hongsen Liu, Yinghua Li, Heng Wang, et al. Fast detection technology of short-circuit fault in distribution network based on the real-time wavelet transform. High Voltage Apparatus, 2018, 54(3): 225-230.
[11] Jin Qiu, Xin Cui, Ye Tian, et al. Analysis of the arc high impedance grounding faults voltage characteristics in non-effective grounding feeders. Power System Protection and Control, 2019, 47(16): 115-121.
[12] Xianshun Zheng, Yi Han, Yi Zhuo, et al. Application of random matrix theory in early fault detection of distribution network. Electrotechnical Application, 2019(3): 58-63.
[13] Jianye Huang, Minwei Chen, Gonglin Zhang, et al. Non-Solidly Earthed Fault Judgement Method Based on Mutation Logic Matrix. Smart Power, 2017, 45(8): 45-50.
[14] Chao Cui, Tingting Yang, Jizhen Liu, et al. Denitration cost modeling for boilers based on BP variable selection and LSSVM. Proceedings of the CSEE, 2015, 35(24): 6413-6420.

[15] Chao Cui, Jizhen Liu, Tingting Yang. Economical optimization of a boiler denitration system based on GA and fuzzy association rules. Journal Of Chinese Society Of Power Engineering, 2016, 36(4): 300-306.

[16] Tao Liu, Jizhen Liu, You Lv, et al. Early fault warning of power plant fans based on MSET and the deviation degree. Journal Of Chinese Society Of Power Engineering, 2016, 36(6): 454-460.