Comprehensive assessment and early warning of the risk of repeated power outages in distribution networks based on multi-source information fusion

Fulai Fan\textsuperscript{1}, Xuesong Tang\textsuperscript{1}, Jianming Wang\textsuperscript{1}, Yunkai Xiao\textsuperscript{1}, Zhutao Yin\textsuperscript{1}, Kai Ye\textsuperscript{1}, Jianchao Ma\textsuperscript{2,}\textsuperscript{*}
\textsuperscript{1} State Grid Chongqing Tongliang Electric Power Supply Branch, Tongliang, Chongqing, China
\textsuperscript{2} College of Electrical Engineering, Guangxi University, Nanning, Guangxi, China
*Corresponding author’s e-mail: majianchao@st.gxu.edu.cn

Abstract: In order to reduce the occurrence of power outages with repeated occurrences in the distribution network, this paper first defines the quantitative index of the concept of repeated occurrences. The FP-Growth data mining algorithm based on the optimal frequent itemset is proposed to obtain the correlation between the power outage factors. Combined with the fuzzy analytic hierarchy process, the main reasons for repeated power outages are obtained. The analytic hierarchy process combining subjective and objective weights is used to comprehensively evaluate the risk of repeated power outages in the distribution network. This paper realizes the early warning of power outage risk based on the comprehensive probability value of power outage. Finally, an actual distribution network in Chongqing is taken as an example to verify the practicability and effectiveness of the method.

1. Introduction
The reasons for the power outage of the distribution network include damage to the natural environment, aging or damage of equipment, unscientific operation and management. Unreasonable grid structure and relay protection configuration are also factors that cause power outages in the distribution network [1]. Research on power outages in distribution networks usually focuses on isolated power outages, and there is a lack of correlation analysis between the main factors and temporal spatial evolution of repeated power outages.

The industry does not have a clear definition of repetitive power outages, and related researches are based on experience. There is a lack of systematic research on this feature. For example, reference [2] shows that the low level of comprehensive power outage management in the distribution network will cause repeated power outages. Reference [3] mentioned that with the rapid increase of the distribution network load and the slow equipment update, the incidence of repeated outages has gradually increased. An effective way to reduce power outages in the distribution network is to conduct risk assessment and early warning. Reference [4] established a distribution network outage risk evaluation index system, and conducted research on power outage risk assessment methods. Reference [5] summarized the risk factors of distribution network operation, and established a calculation formula for the risk loss value. Relevant research either focuses on the general evaluation of failure risks, or focuses on the analysis around a specific power outage inducing factor. Therefore, this paper focuses on the main factors of repeated power outages, and proposes comprehensive evaluation and early warning strategy.
2. Data mining based on multi-source information

2.1 Quantification of the concept of repeated power outage

This paper is oriented to reduce user complaints caused by power outages. We refer to the acceptance standards for customer complaints about repeated power outages in the "Explanation on Business Acceptance Standards for Low Voltage and Frequent Power Outage Complaints" by the marketing department of the grid company. Repeated power outages are defined as 3 or more power outages in a 10kV line within 2 months. If the above criteria are met, it is considered that the line has repeated power outages.

2.2 FP-Growth algorithm based on OFI

Our distribution network blackout data comes from information systems such as 95598 blackout complaint records, distribution network emergency repair management platform, distribution network operation monitoring platform and so on. This paper aims to obtain the multi-dimensional correlation between repeated power outages on this basis. The classic FP-Growth algorithm only depends on the degree of support, which may lead to the negligence of serious failures and small probability events due to improper branch reduction, resulting in deviations in the final results. Therefore, this paper first divides the distribution network fault data samples into "season S, day and night T, area R, terrain D, fault point micro-topography M, meteorological conditions W, faulty equipment E, fault phenomenon C, fault location P, power failure cause L" to group to form a new data table. Secondly, each dimension in the multi-dimensional attribute is decomposed into several subsets according to the combination, and then the FP-Growth algorithm is used to generate frequent item sets for each dimension subset. This avoids the inefficiency of the algorithm after a certain critical point, and the union of all generated frequent patterns is still the frequent pattern set of the entire database. Finally, the appropriate support and confidence levels are set through actual experience to select the optimal frequent items set (OFI) for

![Fig 1. FP-Growth algorithm flow based on OFI](image-url)
the generated frequent items, so as to minimize the low-value frequent sets and redundant rules. The efficiency and accuracy are improved. The FP-Growth algorithm flow based on OFI is shown in Figure 1.

2.3 Analysis of main factors
According to the fault data and algorithm, strong association rules related to repeated power outages can be obtained. On this basis, a corresponding power outage scenario library is formed. This paper uses the fuzzy analytic hierarchy process to determine the degree of influence of each factor in the scene library on repeated power outages. The analytic hierarchy process is based on the scale method of 1-9 and its reciprocal, and the importance of each element in the indicator layer and the factor layer of each unit is compared by the way of expert scoring. The indicator scores are formed into a judgment matrix \( R \). Then the relative weight of each factor of the judgment matrix is calculated. We normalize the feature vector to get the relative importance weight value of the index of this layer relative to the index of the previous layer. Because the constructed judgment matrix has more orders, it needs to meet the consistency test \( (C.R < 0.1) \) according to formula (1) (2).

\[
\begin{align*}
\text{C.R} &= \frac{C.I}{R.I} \quad (1) \\
C.I &= \frac{\lambda_{\text{max}} - n}{n - 1} \quad (2)
\end{align*}
\]

Where \( R.I \) is a random consistency index. \( C.R \) is the consistency ratio. \( C.I \) is the consistency index.

According to the actual situation, this paper divides the factors of repeated power outages in the distribution network into 5 points according to the cascade theory: \{very serious, serious, normal, slight, very slight\}. The fuzzy set of the weight of the criterion level index relative to the target level and the actual score result of each index can be calculated, and the main factors are finally obtained.

3. Comprehensive risk assessment of repeated power outages

3.1 Benchmark probability index for repeated power outages
This paper establishes a set of indicators to obtain the benchmark probability of repeated power outages, and combines the improved analytic hierarchy process to determine the probability of repeated power outages. This paper selects equipment level, operation and maintenance level, and network architecture level to represent internal factors. The basic indicators of equipment level are lightning protection level, insulation level, number of old equipment and protection equipment configuration. Operation and maintenance level includes uninterrupted operation level, emergency repair ability and online detection ability. The grid structure level includes n-1 passing rate, power supply radius and line section. Meteorological factors represent external factors. Among them, user factors are beyond the control of the power grid and are not considered for the time being.

3.2 Weight determination based on a combination of subjective and objective
The statistical data for repetitive power outages can only reflect the difference in the first-level indicators, and it is difficult to refine the second-level indicators. Therefore, this paper uses the objective weight method for the primary indicators, and the comprehensive weight evaluation method of the subjective weight method for the secondary indicators. Subjective aspects are the same as in section 2.3. On the objective side, scores are based on the difference between the number of repeated power outages caused by various primary index factors obtained by statistics, and an objective judgment matrix is constructed according to formula (3)(4)(5).

\[
\Delta C = \begin{bmatrix}
\Delta c_{1,1} & \Delta c_{1,2} & \cdots & \Delta c_{1,m} \\
\Delta c_{2,1} & \Delta c_{2,2} & \cdots & \Delta c_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta c_{n,1} & \Delta c_{n,2} & \cdots & \Delta c_{n,m}
\end{bmatrix}
\] (3)
Where $\Delta C$ is the difference matrix of the number of repeated power outages caused by various secondary index factors. $\Delta c_{ij}=c_i-c_j$. $c_i$, $c_j$ represent the number of repeated power outage lines caused by the i-th and j-th factors.

$$B = \begin{pmatrix}
  b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\
  b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n,1} & b_{n,2} & \cdots & b_{n,n}
\end{pmatrix}$$

(4)

$$b_j = \begin{cases}
  \frac{9\Delta c_{ij}}{\max(\Delta c_{ij})}, & \Delta c_{ij} > 0 \text{ and } \frac{9\Delta c_{ij}}{\max(\Delta c_{ij})} > 1 \\
  1, & \Delta c_{ij} > 0 \text{ and } \frac{9\Delta c_{ij}}{\max(\Delta c_{ij})} \leq 1 \\
  \frac{1}{9}, & \Delta c_{ij} < 0 \text{ and } \frac{\max(\Delta c_{ij})}{9\Delta c_{ij}} \geq 9 \\
  \frac{\max(\Delta c_{ij})}{9\Delta c_{ij}}, & \Delta c_{ij} < 0 \text{ and } \frac{\max(\Delta c_{ij})}{9\Delta c_{ij}} < 9
\end{cases}$$

(5)

Where $B$ is the objective difference matrix.

The analytic hierarchy process is used to calculate the degree of importance based on objective aspects. Finally, the subjective and objective weights corresponding to each indicator are multiplied together to obtain the final weight based on the subjective and objective aspects.

3.3 Quantitative assessment of the risk probability of repeated power outages

According to the concept of distribution network risk assessment, the formula for calculating the comprehensive probability of repeated power outages is:

Comprehensive probability value of repeated power outage = reference probability value × meteorological factor × repeated power outage factor value

The specific step is to obtain the scores and weights of the indicators of repeated power outages in the distribution network, and then calculate them layer by layer according to the calculation method of the analytic hierarchy process to obtain the benchmark probability scores. The value range is 1-100. The meteorological factor reflects the impact of weather on the randomness and dynamics of power outages in the distribution network. The specific values are shown in Table Ⅰ. At the same time, the number of power outages that occurred in the past month will affect the possibility of repeated power outages this month. Therefore, the repeated power outage factor value represents the magnification effect of the historical number of power outages on the probability of repeated power outages. The specific values are shown in Table Ⅱ. The comprehensive probability value range of repeated power outages is also 0-100. If the calculated value exceeds 100, it is counted as 100. The probability of repeated power outages is divided into high probability (>70) and general probability (40-70), low probability (20-40) and low probability (<20) four levels, according to the size of the comprehensive probability value. A comprehensive assessment of the risk of power outage can be achieved.

| Table Ⅰ Meteorological influence factor |
|----------------------------------------|
| Type         | Normal | Typhoon | Thunderstorm and Gale | High temperature | Freeze |
| Factor value | 1      | 1-1.5   | 1-1.5                  | 1-1.5            | 1.2    |

Remarks: Typhoon yellow warning is 1.2. Orange warning is 1.3. Red warning is 1.5. Lightning gale yellow warning is 1.2. Orange warning is 1.3. Red warning is 1.5. High temperature orange warning is 1.2. Red warning is 1.5

| Table Ⅱ Repeated power failure factor |
|---------------------------------------|
| Type                  | 0 outages in two months | 1 outage within two months | 2 outages within two months |
| Factor value          | 1.0                      | 1.1                        | 1.2                        |
4. Grading early warning for repeated power outages

4.1 Consequences of repeated power outages

When the risk assessment result of repeated power outages is "very likely", an early warning should be issued. This paper proposes to quantify and categorize early warning in combination with impact and consequences. Three factors (the importance of the power outage load that affects the consequences of repeated power outages, the rate of repeated power outage complaints and the scope of the power outage) are used as indicators. If the proportion of important load is larger, the loss caused by power outage will be greater. The important factors of the load will be determined according to the proportion of the I and II load. At the same time, this paper also takes into account the regional characteristics, and selects four types of areas: downtown, urban, urban, and rural areas. Each type of area corresponds to a load geographic characteristic factor value to represent the load importance factor. And there are more complaints about power outages. If repeated power outages occur again, it will lead to more complaints. Therefore, we consider the area’s repeated power-on complaint rate index to characterize the area’s tolerance to repeated power outages. The power outage range is expressed by the number of users affected by repeated power outages.

The subjective weighting method is also used to determine the index weights, and power experts’ understanding of the importance of the risk evaluation index for repeated power outages in the distribution network is combined. And then through corresponding mathematical calculations, the weights of each index are obtained. The weight of the importance of the load is 0.3, the weight of historical power outage complaint information is 0.3, and the weight of the number of users affected by repeated power outages is 0.4. After obtaining the scores of each index, the calculation method is calculated layer by layer according to the analytic hierarchy process, and the consequences of repeated power outages are worthy of points.

4.2 Early warning of failure risk classification

This paper classifies the risk according to the consequences of power outages. According to the usual distribution network risk classification theory, this project divides the distribution network repeated power outage risk early warning into three levels: level I, II, III, which corresponding to major losses (>70) and larger losses respectively (40-70) and general loss (<40) repeated power outages affect the consequence value. This processing method is equivalent to the establishment of a two-dimensional risk decision matrix with the possibility of failure and the severity of the consequences of the failure as the coordinate axes. However, considering that the risk value obtained by this method has numerical continuity, the numerical interval processing method is adopted.

5. Example application

In this paper, one year's power outage data in a certain area of Chongqing is used as a data for method application. First, we use the OFI-based FP-Growth algorithm to mine the data table. The default minimum support is 20% and the minimum credibility is 50% initially. Due to space limitations, only display Part of the content is shown in Table III. Then we use the fuzzy analytic hierarchy process of this article to analyse the main factors. Table IV shows the specific circumstances of repeated power outages in this area. User internal failures, design and construction, and high-voltage equipment failures are serious factors that have caused repeated power outages in this area this year.

Then, the comprehensive risk assessment of repeated power outage is carried out for the distribution network in the region. Firstly, the weight value of impact index of repeated multiple blackouts in the region based on subjective and objective factors is obtained, as shown in Table V. The final benchmark probability is 42.291. Combined with the actual meteorological factors and repeated outage factors, the final comprehensive assessment results of outage risk can be obtained. On this basis, the hierarchical warning of power failure is realized.
Table III Some strong association rules in the region

| Rule content | Scene description |
|--------------|-------------------|
| \{R1, D1, T2, L3, M2, E12\} \(\Rightarrow\) C1 | In the plain city at night, a pole failure caused by external forces |
| \{S2, L1, W4, E12\} \(\Rightarrow\) C5 | In summer with strong winds, the towers were blown down by the strong winds. |
| \{D3, S2, L6, E1\} \(\Rightarrow\) C25 | Improper design of overhead lines in mountainous areas, excessive sag of crossing lines, and line collisions |
| \{L6, W6, E14\} \(\Rightarrow\) C7 | After light rain, a flashover occurred inside the arrester along the surface of the valve plate. |
| \{R1, M2, L5, W4, E1\} \(\Rightarrow\) C7 | The trees in the urban area were not managed in accordance with the regulations and contacted with the overhead lines |

Table IV Degree of influence of the main factors of repetitive power outages

| Reason for power outage                | Serious impact | General impact | Slight impact |
|----------------------------------------|----------------|----------------|--------------|
| External factors                       | 29.87%         | 50.58%         | 19.55%       |
| Meteorological factors                 | 32.20%         | 32.40%         | 35.40%       |
| High voltage equipment failure         | 42.32%         | 40%            | 17.68%       |
| Design and construction                | 54.49%         | 32.51%         | 13%          |
| Platform equipment failure             | 32.29%         | 40.58%         | 27.13%       |
| User internal fault                    | 63.31%         | 26.05%         | 10.64%       |
| Operation and Maintenance              | 22.03%         | 34.57%         | 43.41%       |
| Overload                               | 18.92%         | 29.31%         | 51.79%       |

Table V Probability index weight table

| Classification index | Specific indicators                          | Weights   |
|----------------------|----------------------------------------------|-----------|
| Equipment level      | Lightning protection level \(0.14\)          | 0.0574    |
|                      | Insulation level \(0.24\)                    | 0.0984    |
|                      | Proportion of old equipment \(0.58\)         | 0.2378    |
|                      | Distribution network protection equipment configuration \(0.04\) | 0.0164    |
| Operation and maintenance level | Technical level of uninterrupted operation \(0.32\) | 0.1184    |
|                      | Repair capability \(0.37\)                    | 0.1369    |
|                      | Online monitoring capability \(0.31\)        | 0.1147    |
| Grid structure level  | Line N-1 pass rate \(0.16\)                  | 0.0352    |
|                      | Power supply radius \(0.3\)                  | 0.066     |
|                      | Line segmentation \(0.54\)                   | 0.1188    |

6. Conclusion

Aiming at the characteristics of repeated power outages in the distribution network, this paper proposes to use OFI-based FP-Growth mining algorithm and fuzzy analytic hierarchy process to obtain the strong association rules and main reasons of repeated power outages in combination with actual grid statistics. The relevant indicators are formulated. The improved analytic hierarchy process combining subjective and objective is used to comprehensively evaluate power outages and provide subsequent hierarchical early warning. This has certain practical significance for the subsequent solution of the problem of repeated power outages in the distribution network.

References

[1] M. S. Bashkari, A. Sami and M. Rastegar, “Outage Cause Detection in Power Distribution Systems Based on Data Mining,” [J]. IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 640-649, Jan. 2021.
[2] A. Arif, Z. Wang, J. Wang and C. Chen, “Power Distribution System Outage Management With Co-Optimization of Repairs, Reconfiguration, and DG Dispatch,” [J]. IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4109-4118, Sept. 2018.
[3] Q. Zhang, “Analysis on Repeated Power Outages and Overload of Distribution Transformers in the Construction of Shenzhen Distribution Network,” [J]. Technological Innovation and Application, 2019, (01).
[4] W. J. Zhang, “Research on Evaluation Index System and Evaluation Method of Power Distribution Network Failure Outage Risk” [D]. Baoding: North China Electric Power University, 2014.
[5] S. H. Chen, P. Sun and C. Q. Zhang, “Identification and Evaluation of Operational Risks in Distribution Networks” [J]. East China Power, 2011, 39(4).