International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5741-5753

https://doi.org/10.70517/ijhsa463494

Research on Power Restoration and Intelligent Operation and Maintenance Equipment Scheduling under Flooding Disaster Based on Decision Tree Algorithm

Hao Dai[1], Guowei Liu[1,*], Lisheng Xin[1], Longlong Shang[1], Qingmiao Guo[1] and Hao Deng[1]
[1] Shenzhen Power Supply Co., Ltd., Shenzhen, Guangdong, 518000, China
Corresponding authors: (e-mail: LiuGuowei202401@163.com).

Abstract The frequent occurrence of extreme rainstorms leads to an increase in urban flooding disasters, and flooding damage to power system equipment can cause large-scale power outages, affecting social production and life. This paper discusses the power system recovery and intelligent operation and maintenance equipment scheduling after extreme rainstorms triggering flooding disasters. By constructing a grid fault diagnosis alarm model based on decision tree algorithm, a hierarchical fault information diagnosis algorithm is applied to realize rapid power system recovery. The study utilizes ID3, CART, and C4.5 decision tree algorithms to construct the grid scheduling alarm model, and the experiments show that the C4.5 decision tree model has an accuracy of 96.91% on the training set, which is 2.66% higher than that of the ID3 algorithm; the weighted W-C4.5 decision tree's accuracy rate for the “safe” state can reach 98.11%, and the recall rate for the “risky” state can be increased to 79.65%. The recall rate for “risky” states increased to 79.65%. The single-factor analysis shows that the load has the greatest influence over the number of risks, with an average K-value of 0.4578. This study provides theoretical foundation and technical support for power system recovery decision-making and intelligent operation and maintenance equipment scheduling under the flooding disaster, and is of great significance for improving the post-disaster recovery capability of urban distribution networks.

Index Terms flooding disaster, decision tree algorithm, power restoration, intelligent operation and maintenance, fault diagnosis, grid scheduling

I. Introduction

With the growing problem of global climate change, there has been a significant increase in the frequency of extreme natural disasters, and the resulting large-scale power system failures have occurred frequently, leading to severe shortages in energy supply [1], [2]. Taking extreme rainstorms as an example, heavy rainfall can cause waterlogging disasters in cities, resulting in flooding damage to transmission and substation equipment, and such systematic failures not only destroy the power network topology, but also cause regional large-scale persistent power outages, which affects industrial production, residential life, and commercial activities [3]-[5]. Extreme rainstorms in Henan in July 2021 caused severe flooding in several cities, including Zhengzhou, leading to large-scale failures in the local power system [6]. In July 2023, under the influence of exceptionally heavy rainfall, a total of 70 distribution lines of 10 kV and above were damaged and out of service in the city of Beijing, and more than 9,600 power distribution equipment was damaged [7]. In September of the same year, heavy rainfall led to extensive flooding in the city of Shenzhen, power supply and communications and other urban lifeline projects suffered from the impact of flooding of one in 200 years, and flooding failures occurred in a number of power distribution houses [8].

China is at the peak of urbanization, especially in coastal cities, where not only the frequency of extreme rainstorms is higher, but also the risk of urban flooding is exacerbated by the greater hardening of urban surfaces, which leads to a reduction in the amount of rainwater infiltration in urban areas [9]-[11]. While the urban distribution network has a large number of important loads concerning socio-economic stability, such as key municipal infrastructures like medical institutions, administrative command centers, and 5G base station clusters, the mid-range of its power supply can result in the dysfunction of the social operation mechanism, and even a crisis of confidence in the public service may also occur [12], [13]. Therefore, improving the resilience of urban distribution grids in the face of extreme flooding disasters is an urgent issue that needs to be addressed.

Currently, there has been some progress in the study of enhancing power resilience of distribution networks through distribution network reconfiguration under extreme disasters.Duffey, R. B conducted a comprehensive analysis of power resilience prediction under extreme disaster events and quantified the benefits and impacts of deploying standby or emergency power systems for nuclear reactors [14]. Duffey, R. B studied power restoration after catastrophic damage caused by various extreme events and disasters and came up with trends in restoration time and probability of non-recovery dependent on the level of damage and the level of social disruption [15]. Moglen, R et al. proposed a methodology for optimal restoration of power grids in the context of disasters with environmental hazards, i.e., generating optimal restoration strategies by mixed integer linear programming (MILP) and comparing it with heuristic strategies [16]. Che, L et al. proposed an adaptive microgrid formation strategy with mobile emergency resources to improve the resilience of critical power service restoration under extreme disaster conditions, which consists of a load-switching sequence to ensure the dynamic performance of the system during the restoration process [17]. Chen, C et al. proposed an integrated solution in the form of a decision-support tool and utilized the grid to improve distribution system restoration and achieve grid resilience to extreme weather events [18]. Poudel, S and Dubey, A proposed a method for restoring critical loads in distribution networks after a major disaster using distributed energy resources (DERs) with the goal of maximizing the reliability and duration of the restoration while restoring as many critical loads as possible [19]. Arab, A et al. used a post-disaster decision model to search for optimal maintenance schedules, unit commitment solutions, and system configurations to improve the resilience of the grid by responding quickly and effectively to damage caused by natural or man-made disasters [20]. However, the fault environment after inland flooding is complex, and there are uncertainties on both sides of the repair work, such as the uncertainty of travel demand in the transportation network, the uncertainty of repair time of the damaged roadway and the fault point of the distribution network, and other factors [21], [22]. Existing studies, most of which only consider the impact of unilateral uncertainties, are insufficient in terms of comprehensive analysis of multiple uncertainties in coordinated emergency repair on both sides of power traffic [23]-[25].

As the global climate change problem becomes more prominent, extreme natural disasters occur frequently, resulting in a significant increase in large-scale power system failure events. In particular, urban flooding disasters triggered by extreme rainstorms have caused transmission and substation equipment to suffer flooding damage, resulting in persistent regional power outages.2021 Extreme rainstorms in Henan Province in July 2021 led to large-scale failures of the power system in Zhengzhou and many other cities; in July 2023, extraordinarily heavy rainfall in Beijing caused the shutdown of 70 10 kV and above distribution lines across the city; and in September of the same year, heavy rains in Shenzhen triggered extensive flooding across the city, which In September of the same year, heavy rainfall in Shenzhen triggered extensive flooding in the city, which led to the failure of power distribution houses in many places. Urban distribution networks contain many important loads, such as medical institutions, administrative command centers, and 5G base stations and other critical municipal infrastructures, whose power supply disruptions will seriously affect the normal operation of the society. Most of the existing researches only consider the influence of unilateral uncertainties, and there is insufficient comprehensive analysis of multiple uncertainties in collaborative emergency repair on both sides of power traffic. This paper proposes to establish a grid scheduling alarm model through decision tree algorithm, combined with hierarchical fault information diagnosis algorithm and alarm information knowledge graph construction method, to realize rapid diagnosis and processing of power system faults under flooding disaster. Firstly, we analyze the causes of flooding disaster and its impact mechanism on the power system, and then we construct the alarm information analysis framework based on knowledge graph and the grid data association rule optimization algorithm, and then we establish the grid scheduling alarm model. The performance of different decision tree algorithms in fault diagnosis is verified through experiments, and the key factors affecting power restoration are identified using single-factor K-value analysis.

II. Flooding

II. A.Flooding

Flooding is a phenomenon that occurs when the intensity of rainfall exceeds the existing drainage capacity of a town, resulting in waterlogging in the town area. Although the flooding disaster is almost synchronized with the rainstorm events, but not the occurrence of heavy rainfall events will necessarily lead to the occurrence of flooding, so the occurrence of flooding disaster has a certain degree of randomness and unpredictability, but in time has a certain degree of concentration. At the same time, the flooding disaster also has a certain rapidity, short calendar time rainstorms can cause waterlogging of many roads, resulting in traffic disruption, sunken overpasses and underground parking lots and other low-lying areas in a short period of time to become a country of zephyr [26], [27].

The determination of the level of flooding disaster is directly related to the depth of waterlogged roads. In this paper, based on the reference to the relevant literature on the classification of flooding disaster, the flooding disaster is divided into three levels. The evaluation criteria for the grading of flooding disaster are shown in Table 1.

Table 1: Evaluation criteria for classification of flood disaster

Urban flood level	Evaluation criteria	
	Water depth/cm	Hazard level
A	0~14	Mild waterlogging, a slight effect on the water area,The speed of the car fell slightly, and a small number of cars chose to go around.
B	15~29	Moderate waterlogging, most of the vehicles choose to go around, small partsThe vehicle is driven by a low speed.
C	⩾30	Severe internal water and water, the operation function of the water road is failed,The vehicle chooses the adjacent section of the road.

II. B.Analysis of the causes of heavy rainfall and flooding disasters

II. B. 1) Natural factors

Since the middle of the twentieth century, human activities have led to an increase in greenhouse gas (GHG) emissions, thus contributing to the global warming trend. The 6th Assessment Report of the United Nations Intergovernmental Panel on Climate Change (IPCC) indicates that global surface temperatures between 2011 and 2020 will increase significantly by 1.1°C compared to the 1850-1900 baseline period, and that this global warming trend is altering the global water cycle, with the effects reflected in accelerated hydrological cycle processes and increased evaporation from the oceans.

At the same time, as atmospheric temperatures continue to rise, the water-holding capacity of the atmosphere increases significantly, a phenomenon that suggests that the atmosphere needs to accumulate more water vapor in order to reach a saturated state, which in turn triggers precipitation.

Since a moist and warm atmosphere is less stable, it is highly susceptible to the formation of heavy rainfall processes. This change makes precipitation events, when they do occur, more intense than in the past, thus increasing the frequency and intensity of extreme weather events. Therefore, in the context of global warming, climate anomalies are becoming more prominent, extreme weather events are occurring frequently, and meteorological disasters are occurring more frequently, especially the increase in heavy rainfall processes and the severity of rainfall and flooding disasters, which have become a major challenge to be addressed on a global scale.

II. B. 2) Social factors

In cities, the frequency of heavy rainfall and flooding disasters is inextricably linked to human activities. With the acceleration of urbanization, the urban environment has undergone significant changes, exacerbating the potential for storm waterlogging disasters.

First, rapid urbanization has led to a significant reduction in the area of green space, which has been replaced by artificial surfaces such as buildings and impervious paved surfaces. This change directly reduces the permeable area of the ground, which in turn reduces the amount of rainwater infiltration, while increasing the amount of runoff generated, creating conditions for the frequent occurrence of heavy rainfall and flooding.

Secondly, the density of high-rise buildings in the city has created an “urban heat island” effect. The “urban heat island” effect exacerbates the development of thermal convection over the city, creating a difference in air pressure from the surrounding area that favors the formation and intensification of extreme weather events, such as heavy rainfall.

Furthermore, cities emit large quantities of pollutants that create a “turbidity island” effect in the air. The “turbid island” effect promotes the formation of condensation nuclei, which results in higher precipitation in urban areas than in surrounding areas, further exacerbating the risk of flooding from heavy rainfall.

In summary, social factors such as land hardening, increased impermeability of the subsurface, and the city-specific “heat island” and “turbidity island” effects are also contributing to the high frequency of urban storm flooding, with a particular impact on urban transportation.

II. B. 3) Urban stormwater system factors

With the continuous optimization and expansion of urban layout, the pattern of urban drainage system gradually reveals the problem of difficulty in matching the needs of urban construction, which in turn exacerbates the vulnerability of the drainage system. As a complex system, the urban stormwater system usually consists of three key components: the source reduction system, the stormwater drainage system and the flood prevention system.

However, in current practice, there are significant irrationalities in the interface and coordination between these three systems, leading to a series of drainage problems. For example, the interface and coordination between pipelines and river drainage is not perfect, and frequent changes in drainage paths, disorganization of drainage patterns, and the inadequacy of the drainage system itself are becoming more and more prominent. These problems are particularly prominent when the area lacks an effective source abatement system and flood prevention system.

In such cases, rainfall runoff from road surfaces is mainly diverted by the stormwater drainage system, which undoubtedly increases the pressure on the operation of the stormwater drainage system. In addition, due to insufficient routine maintenance and frequent occurrence of extreme rainfall events, the existing stormwater system is already struggling to meet the increasing drainage demand. This has led to more and more serious waterlogging on roads, which in turn has led to flooding, causing serious impacts on urban transportation and residents' lives.

III. Use of decision tree algorithms in grid scheduling

III. A. Construction of decision tree algorithm

The basic principle of the ID3 algorithm is to filter the features according to the criterion of information gain [28]. If the objective is a categorized result and its result corresponds to values $0,1,\sim K-1$, let m be the proportion of K class observations in the node, and if it is the most terminal node refer to Eq:

$$p_{mk}=1/N_m\sum_{y=Qm}I(y=k) \tag{1}$$

In the above equation m is the number of nodes, Q_m is the region of observations and N_m is the number of observations. Typically in the ID3 algorithm, the information entropy of the set D is defined as the share of the current sample set D occupied by the samples of the K th class is P_k:

$$Ent(D)=-\sum_{k=1}^{|v|}P_k\log_2 P_k \tag{2}$$

Suppose the current discrete attribute a has V possible values $[a1,a2,....,aV]$. Then the set of samples of values of attribute a is defined as D^v, then the division of a will produce an information gain see Eq:

$$Gain(D,a)=Ent(D)-\sum_{k=l}^{|v|}\frac{|D^v|}{|D|}Ent(D^v) \tag{3}$$

The C4.5 algorithm is improved from the ID3 algorithm, which uses information gain as a filtering rule, and the result of using information gain is that it is easy to filter those attributes with more values to be used as nodes. The C4.5 algorithm does not take the information gain but uses the information gain rate as a filtering condition, i.e., the ratio of the information gain to the information entropy in the attributes [29]. Namely:

$$Gain\_ratio(D,a)=\frac{Gain(D,a)}{IV(a)} \tag{4}$$

where the information gain entropy of the attribute is:

$$IV(a)=-\sum_{i=1}^{n}\frac{|D^i|}{|D|}\log(D^i) \tag{5}$$

III. B. Flow of diagnostic alarms for power grid faults

Grid fault diagnosis warning system, using automation, intelligent means, rapid search to determine the equipment that has failed, and give processing recommendations. The means and key technologies include fault diagnosis algorithms, fault inference and analysis library (expert knowledge base), inference control strategy, fast search and localization algorithms, and comprehensive display.

In the process of reprocessing and reanalysis of real-time data, the grid fault diagnosis and alarm system has to discover the correlation between signals by means of specialized signal processing and analysis software for discrete grid signals. Intelligent grouping of signals and inductive reasoning, and the design of a specialized fault analysis knowledge base for the capture of grid fault analysis, to give the possible causes of faults and processing recommendations.

This new approach can provide timely early warning of defects and faults in the grid, thus effectively assisting dispatchers in detecting and handling grid faults and defects. On this basis, statistics of alarm information and fault retrieval function can be provided to realize automatic recording and query management of power grid faults.

The data processing flow is shown in Fig. 1. After receiving and preprocessing the signals (after filtering out the signals of maintenance, debugging, cold standby, blocking, etc.), the regional grid fault diagnosis and alarm system carries out the rule-based analysis and matching and draws conclusions. Rule-based analysis is the main core point that distinguishes this system from the traditional alarm signal processing function, which not only contains a rule logic library, but also contains improved algorithms to support logic matching as well as control strategies.

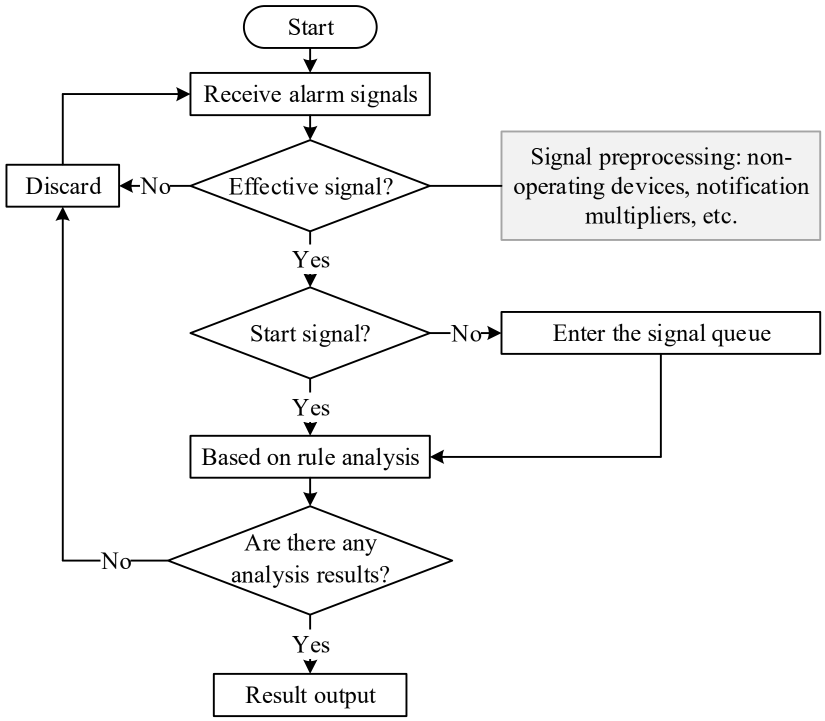

Figure 1: Data processing process

III. C. Hierarchical fault information diagnosis algorithm

The equipment position signals, protection information and telemetry information directly uploaded from the remote management machine or the letter-protecting sub-station are characterized by real-time accuracy. While the fault recording system can record the detailed information at the time of fault, its timeliness is poor, and the waveform files generated by the recording device can generally be used for multi-factor combination analysis, or after-the-fact cause analysis. Based on the characteristics of power grid equipment signals, the use of hierarchical fault information diagnosis algorithm can be a good solution to the contradiction between “flexible and fast” and “accurate and reliable” diagnostic methods. Layered fault diagnosis algorithm is studied as follows:

1) Grid fault diagnosis and processing strategy is divided into three layers

(1) Fast fault diagnosis layer, (2) protection logic diagnosis layer, and (3) fault combination diagnosis layer.

2) Fast fault diagnosis layer

The principle of the rapid fault diagnosis layer is: when the circuit breaker tripping alarm signal is received, based on the grid hierarchical model quickly utilizes and initiates the check of the grid topology connectivity relationship, and based on the topology charged state and the circuit breaker's switching and closing information, the faulty equipment can be quickly localized to find the faulty equipment. This strategy is simple and efficient. Assuming that the fault point is unique, the fault diagnosis and analysis can be completed by this strategy. If the fault point is not unique, then turn on layer (2), that is, the protection logic diagnosis layer.

3) Protection Logic Diagnosis Layer

The principle of protection logic diagnostic layer is: in the power system, protection devices and primary equipment have close logical correlation, in the same time period, the use of circuit breaker equipment and the action of the protection device with the relationship between the regional power grid can be based on the protection action alarm information for fault diagnosis. This strategy has the technical characteristics of accurate fault location and complex logic, and can complete the analysis and diagnosis of most faults.

4) Fault combination diagnosis layer

The principle of fault combination diagnostic layer is that the waveform file of fault recording device, the change information of circuit breaker and the action information of protection device are used to carry out combination diagnosis and identify the faults, which involves the complex logic relationship between switch displacement, safety automatic device and protection equipment. As the efficiency of the generation, summoning and parsing process of the fault wave recording file is relatively low, and its real-time performance is poor, it is generally used to deal with the fault diagnosis of the complex system is very high.

III. D. Construction of Knowledge Graph of Alarm Information

III. D. 1) Alarm information knowledge graph construction process

Alarm information is the basic information collected by grid dispatch through SCADA system, which can indirectly describe various abnormal events in the system. Aiming at the characteristics of the text of alarm information, this paper makes the following modifications when constructing the knowledge graph of alarm information:

(1) This paper focuses on the study of the adjacency of the fault event description clauses in the alarm information, each fault clause is an entity, and a single piece of alarm information generally contains no less than three entities, and the connection between the entity objects forms a characterization of the fault event.

(2) Alert information is the description of grid operation events in the grid dispatch operation EMS system, the knowledge graph of alert information belongs to the domain knowledge graph, the lexical meaning of entities is limited to the electric power domain, with a clear naming specification, entity disambiguation and co-reference disambiguation problems basically do not exist, without knowledge fusion step.

(3) In the relationship extraction, according to the neighboring frequency between the previous fault participle and the next fault participle, the fault event description participle correlation matrix is established, and the neighboring frequency between the fault participles is used as the relationship between entities.

(4) Perform data integration to organize a triad of head entity-adjacent frequency relationship between entities-tail entity (entity 1-relationship-entity 2) type. Utilize Neo4j graph database to build knowledge graph based on alarm information. The construction process of the knowledge graph of alarm information is shown in Figure 2.

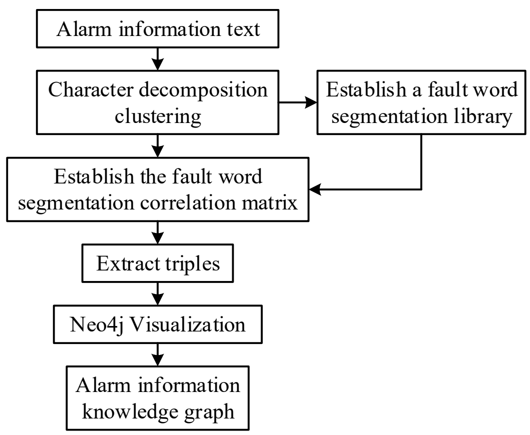

Figure 2: The construction process of the information knowledge map

III. D. 2) Entity extraction

The main task of entity extraction is to extract the fault event description participle in the text of the alarm message as an entity and number the obtained fault event description participle. The specific steps are:

(1) Firstly, the text of the alarm information is processed by word division, which is based on the conventional dictionary and viterbi algorithm, and then corrected by manual intervention using customized dictionaries.

(2) De-weighting and filtering the text of the alarm information after word separation, and constructing a lexicon of fault event descriptions of the alarm information. Based on the fault event description sub-word dictionary, the words are numbered according to the order of appearance of the fault event description sub-word in the dictionary.

III. D. 3) Relationship extraction

To obtain the fault event description clauses correlation matrix, the obtained fault clauses are numbered and the fault clauses correlation matrix is created by using the numbers as the rows and columns of the matrix:

$$A_{nn}=\begin{array}{c} \\ 1 \\ 2 \\ \cdots \\ i \\ \cdots \\ n \end{array}\begin{array}{c} \begin{array}{cccccc} 1 & 2 & \cdots & j & \cdots & n \end{array} \\ \begin{bmatrix} cor_{11} & cor_{12} & \cdots & cor_{1j} & \cdots & cor_{1n} \\ cor_{21} & cor_{22} & \cdots & cor_{2j} & \cdots & cor_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ cor_{i1} & cor_{i2} & \cdots & cor_{ij} & \cdots & cor_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ cor_{n1} & cor_{n2} & \cdots & cor_{nj} & \cdots & cor_{nn} \end{bmatrix} \end{array} \tag{6}$$

A_{nn} denotes the fault event description participle correlation matrix, where the diagonal elements are all zero. $i,j=1,2,\ldots,n$ denotes the number of the fault participle in the fault event description participle dictionary, and cor_{ij} denotes the number of times the fault participle j follows the fault participle i. The magnitude of the value of how often two fault participles appear next to each other determines the correlation between them, the larger cor_{ij} is, the stronger the correlation between the fault participles. The A_m is an asymmetric matrix, i.e., $cor_{ij}\neq cor_{ji}$.

The operation is initialized by initializing the matrix so that $cor_{ij}=0(i,j=0,1,\ldots,n)$. Iterate through the text of the warning message after the disambiguation, and when the fault event description disambiguation numbered j occurs once after the fault event description disambiguation numbered i, perform the following operation:

$$A_{nn}=A_{nn}+E_{ij} \tag{7}$$

Among them:

$$E_{ij}=\begin{array}{c} \\ 1 \\ 2 \\ \cdots \\ i \\ \cdots \\ n \end{array}\begin{array}{c} \begin{array}{cccccc} 1 & 2 & ... & j & ... & n \end{array} \\ \begin{bmatrix} 0 & 0 & ... & 0 & ... & 0 \\ 0 & 0 & ... & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ n & 0 & ... & 0 & ... & 0 \end{bmatrix} \end{array} \tag{8}$$

III. E. Grid dispatch alarm modeling using decision tree algorithm

The decision tree algorithm is used to optimize the association rule generation algorithm for grid data based on knowledge graph, and then the optimized algorithm is used to analyze the grid dispatch alarm signals and construct the grid dispatch alarm fusion set.

In the process of data mining, there is an association between each attribute within the decision tree. If the rule attributes in the traditional association rule algorithm are centrally trained using the decision tree structure, the association relationship between each attribute is analyzed and simplified, the computing efficiency of the association rule algorithm can be effectively improved under the premise of ensuring its classification accuracy. According to the above idea, the association rule generation algorithm for grid data based on knowledge graph is optimized and designed as follows:

Step 1: Based on the C4.5 decision tree structure, construct the decision tree model of association rule R as S.

Step 2: In the decision tree S the rule R is a rule path from a root node to a leaf node in the decision tree, i.e.:

$$R:\{r_i \mid \text{ if } Cond_1\wedge Cond_2\wedge\cdots\wedge Cond_i \text{ then class } C_x\} \tag{9}$$

where "$\wedge$" is the "per-positional different-or" operator.

Step 3: Simplify the rules R in the decision tree S by combining the decision rule parameters τ for each association rule r_i. The different constraints $Cond_i$ and $Cond_j$ corresponding to i and j commands correspond to the attributes A_v,A_s, respectively. When there is no association rule property between A_v,A_s, then $Cond_i$ and $Cond_j$ are not kept. If there is association rule data between A_v,A_s, the attribute relationship probabilities are calculated respectively.

Step 4: Based on the size of the obtained probabilities, determine whether $Cond_i$ and $Cond_j$ are retained or not. The decision tree association rule threshold $P(Cond_{i+1}\wedge Cond_i)$ is obtained, and the rule set is reconstructed by the newly designed association rule threshold to obtain a new dataset T'' with association attributes.

Through the above knowledge graph-based grid data association rule optimization algorithm, the grid dispatch alarm signals are analyzed and the grid dispatch alarm signal association rule fusion set is constructed with the following expression:

$$x'(t) = \frac{1}{2}\sqrt{x(t)+v} - h(t) \tag{10}$$

where $x(t)$ is the correlation parameter of the alarm information of grid dispatch. $h(t)$ is the fault information characteristic value of grid dispatch. v is the command-assisted decision coefficient.

Then the knowledge graph grid data layer decision tree data set is:

$$s_i = Hx'(t)^Q \tag{11}$$

where Q is the fault point being monitored weight.

According to the constructed knowledge graph data layer, the fault information feature association rules in the alarm information are mined and analyzed, and the feature offset coefficient of the grid dispatch intelligent alarm is obtained as:

$$s'(t) = G + \sum_{t=1} \frac{h(t)+S(t)}{2} \tag{12}$$

where $s'(t)$ is the offset coefficient of the grid dispatch alarm information. $S(t)$ is the monitoring quantization index of grid dispatch smart alarm. t is the alarm time.

Under the condition of ensuring stable operation of the grid, the resonant frequency $S_r(t)$ is obtained as:

$$S_r(t) = \frac{S(t)}{h(t)} + \frac{H}{n_s(t)} \tag{13}$$

where $n_s(t)$ is the virtual alarm value of grid scheduling.

In the region where the alarm fault exists, the region-wide calculation function for the existence of the grid dispatch fault point is obtained by using the normal grid dispatch information value α_k obtained above as:

$$Q_s = \alpha_k + \frac{p(f_1)}{2} - \sum_{k=1}^{K} \alpha_k \tag{14}$$

where $p(f_1)$ is the fault region matching function for grid dispatch. f_1 is the current frequency of the fault region, and K is the fault distribution coefficient.

Through the above-determined fault characteristic quantities of grid scheduling information, as the association rules of intelligent alarm, the intelligent alarm model of grid scheduling is constructed in the grid data layer of the knowledge graph:

$$\rho = \sum_{k=1} [\alpha_k + S(t)]^2 - \frac{\sin k}{\sin[z(k)]} + R \tag{15}$$

where $z(k)$ is the density distribution function of the grid dispatch intelligent alarm.

By calculating the resulting grid scheduling intelligent alarm model, combined with Eq. (8), the grid scheduling intelligent alarm fault detection output function is obtained:

$$H_k = G(U \mid \mu_k, k) - \frac{S(t)}{2\tau_i(t)} \tag{16}$$

Among them:

$$\tau_i(t) = \alpha_k / 2 + \sqrt{S(t)} \tag{17}$$

where $G(U \mid \mu_k, k)$ is the set of intelligent alarm fault features for grid dispatch. $\tau_i(t)$ is the attribute relationship function between the monitoring quantization indices.

The grid scheduling intelligent alarm fault output function is detected to complete the accurate judgment of the location of the grid scheduling alarm fault point, and finally realize the grid scheduling alarm.

IV. Experiments and simulations

IV. A. Data preparation

The data source of the experiment is a large amount of power data in the database of OPEN3000 power monitoring system of a power supply bureau of the Southern Power Grid.

After the preliminary cleaning of the data, 14 features including "operation type", "voltage level", "equipment type", "equipment (line, transformer, generator) load rate", and "equipment exceeding limits" were selected to construct the dataset. Five experts and dispatch personnel were invited to add risk assessment labels based on the local power grid's topology and overall safety situation, in order to align with the actual safety risk status of the power grid.

Since real power grids are generally in normal operation and dangerous accidents rarely occur, there are almost no negative samples in the data. When used for training, although the correct rate is very high, but due to the lack of negative samples can not distinguish the dangerous situation better, so we need to construct negative samples based on real data to meet the real situation.

In order to simulate the real situation and test the training effect of unbalanced samples, the final construction of the dataset has a large gap between the proportion of “safe” samples and “dangerous” samples, which is about 11:1.

IV. B. Analysis of results

The experiment constructed a total of 2,000 pieces of data, which were randomly divided into training set and test set, of which 80% as the training set, 20% as the test set, and the training set and the test set of various types of samples accounted for basically the same ratio. In the test set, there are 320 “safe” samples, 60 “risky” samples and 20 “dangerous” samples, which are unbalanced samples.

The decision trees were constructed respectively using the ID3 algorithm, the CART algorithm and C4.5. Firstly, ten-fold cross-validation and grid search are utilized for "max_depth" (the maximum depth of the tree), "min_samples_split" (the minimum number of samples required for splitting nodes), "min_samples_leaf" (the minimum number of samples required for leaf nodes), and "max_leaf_nodes" The four parameters (the maximum number of leaf nodes) are tuned based on the accuracy rate of the training set. After obtaining the optimal parameters, they are used to train the decision tree, and the training effect of the weighted decision tree (represented by "W-") is compared with that of the general decision tree.

For the random forest algorithm, the parameter “n_estimators” (number of spanning trees) also needs to be trained and the random forests based on general decision trees and weighted decision trees are compared. The training results are shown in Figure 3.

ID3 algorithm, CART algorithm, and C4.5 constructed decision trees, among which the accuracy of C4.5 decision tree model is better. On the training set, the accuracy of C4.5 decision tree model is 96.91%, which is 2.66% and 1.37% higher than ID3 algorithm and CART algorithm, respectively. After the decision tree was weighted, the model accuracy of W-C4.5 decision tree increased to 97.15%.

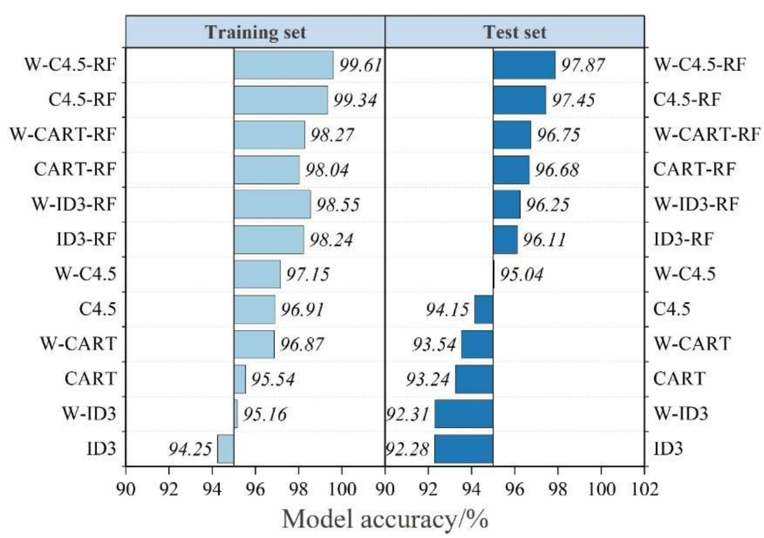

Figure 3: Training results

The accuracy rates of the models for the other metrics on the test set are shown in Figure 4. The accuracy of the decision trees for each category is higher for “Safety” than for “Risk” and “Hazard”. The accuracy of W-C4.5 decision tree for “safety” is 98.11%.

For general decision trees, the CART algorithm outperforms the ID3 algorithm, and the C4.5 algorithm outperforms both.

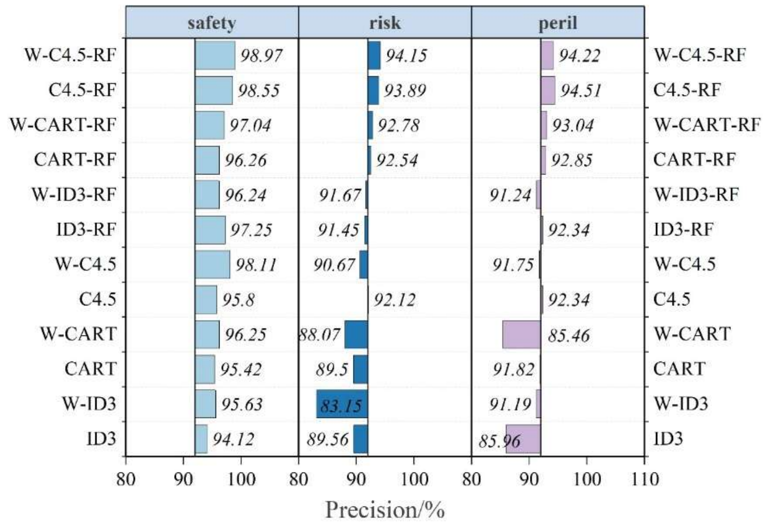

Figure 4: The accuracy rate of other indicators in the test set is counted

The recall of each model for other metrics on the test set is shown in Figure 5. The models have higher recall for “safety” and “hazard” and lower recall for “risk”. The ID3 algorithm has a 66.33% recall of “Risk”, which is lower than the C4.5 algorithm's recall of “Risk”. When the C4.5 algorithm is trained with a weighted decision tree, the recall rate of the W-C4.5 decision tree model for “risk” can be increased to 79.65%.

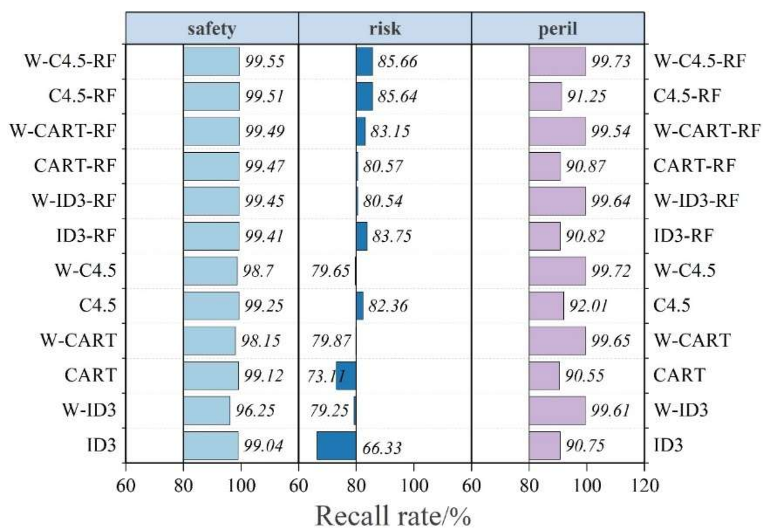

Figure 5: The recall rate of each model in the test set

IV. C. Grid fault alarm simulation

The correlates and risk value sign-control relationships are shown in Table 2.

Table 2: Correlation factors and wind risk values

Embody	Source	Factor composition	Field name (n=1,2,3,4)
Factor	Factor 1	Controlled risk value	CF_n
		Total length of repair	TR_n
		Power failure controlled risk value	C_n
		Total length of power failure	T_n
	Factor 2	Affect the total number of households	N_n
		Release the supermarket	OT_n
		Whether or not	D_n
	Factor 3	The risk of load ratio	R_{xn}
Dependent variable	-	Guarantee number	y_{xn}
Results	Calculate	Forecast risk	V_{xn}

Adjusting for a single factor of 0 and leaving the other variables unchanged yields y'_{xn}. The slope is calculated as shown in equation (18):

$$K_n = \frac{y'_{xn} - y_{xn}}{0 - x} \tag{18}$$

Judgment is made by K value, which has a greater impact on the number of guarantees, and 6 hours are taken as a time period, and the day is divided into 4 time periods, and the value of risk change is calculated for each time period.

The factors that have an impact on the number of guarantees are represented by the 4 time periods, including the fault controllable risk value, the number of load ratio risks, and the outage controllable risk value. The minimum 3 factors in the first, third, and fourth time periods are whether the release is overtime, the total number of customers affected, and the total duration of outages.

Simply put, whether the grid company can release blackout information according to the set time at 6:00 ~ 12:00, the user perception is more sensitive, if there is no timely release of blackout information, the user is unprepared when the power outage of the station area will increase the number of times the user's security.

The single-factor K-value analysis is shown in Table 3. It can be seen that after averaging the slopes of the fourth time period for each factor, the overall average K-value for all factors is obtained. Load has the highest K-value than the number of risks, with an average K-value of 0.4578, which is ranked first, indicating that load has a greater influence than the number of risks.

Table 3: Single-factor k analysis

Factor	First time section K value	Second time section K value	Third time section K value	Fourth time section K value	Mean K value	K sort
The risk of load ratio	0.4780	0.4553	0.4419	0.4558	0.4578	1
Fault controlled risk value	0.2311	0.3212	0.3723	0.4171	0.3354	2
Total length of repair	0.0133	0.0211	0.0242	0.0235	0.0205	4
Power failure controlled risk value	0.0410	0.0362	0.0395	0.0366	0.0383	3
Total length of power failure	0.0012	0.0010	0.0012	0.0031	0.0016	6
Affects the total number of users	0.0008	0.0007	0.0007	0.0017	0.0010	8
Whether or not to issue timeout	0.0002	0.0022	0.0001	0.0000	0.0006	7
Whether or not	0.0099	0.0261	0.0173	0.0103	0.0159	5

The distribution and importance of the single time period of each factor in the number of times of impact coverage is analyzed by sorting to determine whether there is a significant difference and to determine whether the impact of load is stronger than the number of times at risk. For the user, if all the station equipment fails, the user has to bear the added risk of power outage, and if the power outage is perceived, the user will immediately report to the grid company.

Through the establishment of the decision tree-based grid scheduling alarm model, it is clear to know how each factor affects the value of fault risk, so as to provide a reference for the maintenance of the power sector and the decision-making of the power system. Combined with the results of simulation experiments using the decision tree-based grid scheduling alarm model, it can be applied to flooding and other disasters to improve the efficiency of the power sector maintenance work.

V. Conclusion

In this study, we constructed a grid fault diagnosis and warning model based on decision tree algorithm for the problem of power restoration and intelligent operation and maintenance equipment scheduling under flooding disaster. The experimental analysis shows that the C4.5 decision tree model outperforms ID3 and CART algorithms, with an accuracy of 96.91% on the training set, which is 2.66% higher than ID3 algorithm and 1.37% higher than CART algorithm. After weighted processing, the accuracy of W-C4.5 decision tree model is further improved to 97.15%. On the test set, the accuracy of the W-C4.5 decision tree for the “safe” state reaches 98.11%, and the recall rate for the “risky” state increases to 79.65%, which is significantly better than the unweighted model. The results of single-factor K-value analysis show that, among the factors affecting the number of times of power security, the load ratio has the greatest impact on the number of times of risk, with an average K-value of 0.4578; followed by the controllable risk value of faults, with a K-value of 0.3354; and the third is the controllable risk value of power outages, with a K-value of 0.0383. The experiment also reveals that, in different time periods (from 6:00 a.m. to 12:00 p.m.), the users are more sensitive to the timely release of outage information by the power grid company. The perception is more sensitive. The hierarchical fault information diagnosis algorithm combined with the alarm information knowledge graph construction method effectively improves the accuracy and efficiency of grid fault diagnosis.

The study provides theoretical support for maintenance decision-making and operation and maintenance scheduling of the power sector under the flooding disaster, which is of positive significance for improving the disaster prevention and mitigation capability of urban distribution networks. Future research can further explore the optimized scheduling strategy under the synergistic effect of multiple factors to enhance the resilience of the power system under extreme weather conditions.

Funding

This research was supported by the Science and Technology Project of China Southern Power Grid Company Limited (Project No. 090000KKK52222158).

References

[1] Wang, Y., Chen, C., Wang, J., & Baldick, R. (2015). Research on resilience of power systems under natural disasters—A review. IEEE Transactions on power systems, 31(2), 1604-1613.

[2] Bhusal, N., Gautam, M., Abdelmalak, M., & Benidris, M. (2020, August). Modeling of natural disasters and extreme events for power system resilience enhancement and evaluation methods. In 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS) (pp. 1-6). IEEE.

[3] Xiang, Y., Wang, T., & Wang, Z. (2022). Risk prediction based preventive islanding scheme for power system under typhoon involved with rainstorm events. IEEE Transactions on Power Systems, 38(5), 4177-4190.

[4] Mohamed, M. A., Chen, T., Su, W., & Jin, T. (2019). Proactive resilience of power systems against natural disasters: A literature review. Ieee Access, 7, 163778-163795.

[5] Watson, P. L., Spaulding, A., Koukoula, M., & Anagnostou, E. (2022). Improved quantitative prediction of power outages caused by extreme weather events. Weather and Climate Extremes, 37, 100487.

[6] Wu, P., Clark, R., Furtado, K., Xiao, C., Wang, Q., & Sun, R. (2023). A case study of the July 2021 Henan extreme rainfall event: From weather forecast to climate risks. Weather and Climate Extremes, 40, 100571.

[7] Zhang, H., Wang, J., Liu, Y., & Cao, Y. (2025). Uncertainty Analysis of the Extreme Rainstorm in Beijing on July 2023 Based on the Convection-Allowing Ensemble Forecast System of North China. Journal of Hydrometeorology, 26(4), 415-429.

[8] Qiang, Y., Xiao, T., He, J., & Zhang, L. (2024). Multi-hazard stress testing under extreme rainstorms in the Shenzhen metropolitan area. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 1-19.

[9] Sun, S., Zhai, J., Li, Y., Huang, D., & Wang, G. (2020). Urban waterlogging risk assessment in well-developed region of Eastern China. Physics and Chemistry of the Earth, Parts a/b/c, 115, 102824.

[10] Tang, X., Shu, Y., Lian, Y., Zhao, Y., & Fu, Y. (2018). A spatial assessment of urban waterlogging risk based on a Weighted Naïve Bayes classifier. Science of the total environment, 630, 264-274.

[11] Lin, T., Liu, X., Song, J., Zhang, G., Jia, Y., Tu, Z., ... & Liu, C. (2018). Urban waterlogging risk assessment based on internet open data: A case study in China. Habitat International, 71, 88-96.

[12] Shi, J., Wang, H., Zhou, J., & Zhang, S. (2024). Assessment and Improvement of Emergency Rescue Service Accessibility under Urban Waterlogging Disasters. Water, 16(5), 693.

[13] Kong, F., Sun, S., & Lei, T. (2021). Understanding China’s urban rainstorm waterlogging and its potential governance. Water, 13(7), 891.

[14] Duffey, R. B. (2019). The risk of extended power loss and the probability of emergency restoration for severe events and nuclear accidents. Journal of Nuclear Engineering and Radiation Science, 5(3), 031601.

[15] Duffey, R. B. (2019). Power restoration prediction following extreme events and disasters. International Journal of Disaster Risk Science, 10(1), 134-148.
[16] Moglen, R., Leibowicz, B. D., Kwasinski, A., & Cruse, G. (2024). Optimal restoration of power infrastructure following a disaster with environmental hazards. Socio-Economic Planning Sciences, 95, 101974.
[17] Che, L., & Shahidehpour, M. (2018). Adaptive formation of microgrids with mobile emergency resources for critical service restoration in extreme conditions. IEEE Transactions on Power Systems, 34(1), 742-753.
[18] Chen, C., Wang, J., & Ton, D. (2017). Modernizing distribution system restoration to achieve grid resiliency against extreme weather events: An integrated solution. Proceedings of the IEEE, 105(7), 1267-1288.
[19] Poudel, S., & Dubey, A. (2018). Critical load restoration using distributed energy resources for resilient power distribution system. IEEE Transactions on Power Systems, 34(1), 52-63.
[20] Arab, A., Khodaei, A., Khator, S. K., & Han, Z. (2016). Electric power grid restoration considering disaster economics. Ieee Access, 4, 639-649.
[21] Sun, S., Li, G., Bie, Z., Dong, Z. Y., Liu, J., Chen, M., ... & Xu, M. (2025). Resilience Assessment of Power Distribution System against Rainstorm-Induced Urban Waterlogging. IEEE Transactions on Industry Applications.
[22] Li, X., Liu, Y., Su, S., Li, J., & Li, B. (2024, October). Flood risk analysis of urban power distribution terminals under the influence of rainstorm and waterlogging. In Ninth International Symposium on Advances in Electrical, Electronics, and Computer Engineering (ISAEECE 2024) (Vol. 13291, pp. 554-560). SPIE.
[23] Dai, H., Liu, D., Liu, G., Deng, H., Xin, L., Shang, L., ... & Chen, C. (2025). A Method for Restoring Power Supply to Distribution Networks Considering the Coordination of Multiple Resources Under Typhoon-Induced Waterlogging Disasters. Energies, 18(5), 1284.
[24] Ma, N., Xu, Z., Wang, Y., Liu, G., Xin, L., Liu, D., ... & Chen, C. (2024). Strategies for Improving the Resiliency of Distribution Networks in Electric Power Systems during Typhoon and Water-Logging Disasters. Energies, 17(5), 1165.
[25] Cao, Y., Zhou, B., Chung, C. Y., Li, J., Sun, Y., Gan, L., & Yang, H. (2025). Integrated Robust Planning of Photovoltaic-Penetrated Power and Drainage Networks Under Prolonged Waterlogging Risk. IEEE Transactions on Sustainable Energy.
[26] Pei Xing,Ruozi Yang,Wupeng Du,Ya Gao,Chunyi Xuan,Jiayi Zhang... & Feilin Xiong. (2024). Risk Assessment of Community-Scale High-Temperature and Rainstorm Waterlogging Disasters: A Case Study of the Dongsi Community in Beijing. Atmosphere,15(9),1132-1132.
[27] Jian Zhang,Bin Pan,Wenxuan Shi,Yu Zhang,Shixiang Gu,Jinming Chen & Quanbin Xia. (2024). Construction of a High-Resolution Waterlogging Disaster Monitoring Framework Based on the APSIM Model: A Case Study of Jingzhou and Bengbu. Remote Sensing,16(14),2581-2581.
[28] Huang Ke & Wang Tingxuan. (2024). Optimized Application of the Decision Tree ID3 Algorithm Based on Big Data in Sports Performance Management. International Journal of e-Collaboration (IJeC),20(1),1-20.
[29] Chun Pan. (2024). Construction of Risk Prediction Models for Enterprise Finance Sharing Operations Using K-Means and C4.5 Algorithms. International Journal of Computational Intelligence Systems,17(1),208-208.