Research Article

Hong Zhang*, Rui Li, and Chuncheng Shi

Deep learning technology of Internet of Things Blockchain in distribution network faults

Abstract: Nowadays, the development of human society and daily life are inseparable from the power supply. Therefore, people also put forward higher requirements for the reliability of distribution network, but power companies can only passively deal with distribution network failures, which is a bottleneck for the improvement of distribution network reliability. The Internet of Things (IoT) is the best solution for online equipment status monitoring and basic data sharing for large, widely distributed, relatively fixed, and large numbers of equipment. The construction of the IoT for power distribution equipment faces many important problems, including the selection of networking, equipment selection, and interaction standards. When researching the implementation plan, research on the distribution of IoT market was carried out. Based on the grid, the idea of optimizing the investment selection plan of the power distribution using IoT was discussed, and a result verification model was established. After the completion of the theoretical part, a case study of medium-voltage distribution grid equipment management and medium-voltage distribution network equipment management based on the grid was carried out by taking a real enterprise application situation as an example. Realizing fault diagnosis of distribution network will not only provide decision support for operation and maintenance of distribution network for power companies, but also have great economic and social benefits. Aiming at the shortcomings of single data mining method in distribution network fault diagnosis, hybrid data mining method is proposed. First, rough set theory is used to reduce the original fault data and form a simplified rule set. Because of the non-linearity of distribution network fault and the strong learning ability, adaptability, and robustness of Bayesian network, Bayesian network can be used to classify distribution network faults. Therefore, a simplified fault diagnosis system is established in this paper, and its correctness is confirmed. Then, the learning and training are carried out by using Bayesian network to call the simplest rule set, which has the characteristics of short learning and training time and high diagnostic accuracy.

Keywords: distribution network fault diagnosis, data mining, information entropy, Bayesian network

1 Introduction

When the distribution network fails, especially when there are many uncertain factors such as information loss due to channel interference, protection, switch malfunction, rejection, and so on, the response of

* **Corresponding author: Hong Zhang,** School of Electrical Engineering and Information Technology, Changchun Institute of Technology, Changchun 130000, Jilin, China, e-mail: 0202055@ccit.edu.cn
Rui Li: Maintenance Company, Maintenance Company of Jilin Electric Power Co., Ltd, Changchun, 130000, Jilin, China, e-mail: 43323590@qq.com
Chuncheng Shi: Safety Supervise Department, Power China Sepco1 Electric Power Construction Co. Ltd, Jinan, 250102, Shandong, China, e-mail: scc9040@sepco1.com

power system will be complicated, which will cause more difficulties [1,2]. Therefore, developing a fault diagnosis method of distribution network [3–5], which is sensitive to error information, strong fault-tolerant, fast, and accurate is important. Interaction and integration are the ability to liberate the power grid's human management mode and realize the ultimate intelligent management. Interaction contains two connotations. On the one hand, it is the interaction between grid operators and grid equipment. Through wearable smart devices, grid personnel can "communicate" with the grid to handle affairs without leaving home. On the other hand, it is the interaction between the power users and the grid. Users use their own smart appliances or distributed power sources to "communicate" with the grid on transaction and service issues. Integration is to optimize and integrate the management process and classified business of power grid enterprises to realize all-round intelligence in production, marketing, and scheduling [6,7]. This is the ultimate goal of the power distribution Internet of Things (IoT).

At present, the distribution network is closely related to human life, and every household cannot do without the power system. Therefore, once the distribution network fails, it will directly affect social production and people's lives, and bring huge losses to social production. Fault diagnosis of distribution network with high automation level [8–10] puts forward a correct and effective regional blackout recovery strategy to help dispatchers to accurately locate the fault location, isolate the fault area, and quickly restore the power supply in the fault area. Some faults will inevitably occur in power system due to long operation period, heavy maintenance workload, and aging of line insulation [11,12]. It is necessary to quickly monitor and eliminate faults, by realizing fault location and quick fault recovery of distribution network [13–16], and guarantee the normal life of human beings. In addition, due to the shortcomings of some existing fault diagnosis methods, it is difficult to achieve satisfactory results, so the improvement of distribution network fault diagnosis methods is imminent. Although the current domestic power distribution IoT technology research has been relatively mature, the management problems faced by this field are also very prominent. Since most of the power distribution IoT construction in various places is in the planning stage, it is still difficult to have a model to follow for the investment selection and supplier selection. From a technical point of view, there are a wide range of smart devices for power distribution IoT in the market, and the prices are relatively expensive, but there are few practical cases. So how to complete the equipment selection, how to choose the installation point, and what installation measures to take at different installation points are the issues to be discussed in management [17–20].

Data mining technology is a combination of database technology and artificial intelligence [21–24], and the research field is more active in the international community [25]. Data mining involves a large number of disciplines and methods. The commonly used data mining methods include neural network method [26], genetic algorithm, rough set method, and decision tree method. In data mining, decision tree classification method represents a classification [27–29].

It is necessary to determine the complex relationship between fault and its influencing factors, and then build a model to diagnose the distribution network fault. By using data mining technology to diagnose faults at feeder level [30–32], it can fine manage the patrol work of feeder, optimize the emergency repair station of distribution network, and provide guidance for the operation and maintenance work of feeder [33].

Aiming at eliminating the shortcomings, such as poor fault tolerance, low diagnosis speed and the uncertainty of the processed data, of single data mining method, new breakthroughs have been made in diagnosis, to ensure power supply, save human resources, and make rational use of various resources of distribution network, so as to maximize the economic and social benefits of distribution network.

2 Proposed method

2.1 Fault diagnosis of distribution network

The work of the distribution network includes four links: "generating, transforming, transmitting and distributing". The working condition of distribution network has a great influence on the quality of power

to the consumers, and its importance is getting higher and higher. Therefore, it is very important to improve the speed and accuracy of finding faults in a relatively short time, and to find out the fault area accurately, so as to maximize and guarantee the normal power supply to users. At the same time, the rapid and accurate treatment of the fault can also reduce the damage of the equipment and narrow the scope of the fault.

2.1.1 Fault location based on recloser and segmenter

The fault location method based on recloser and segmenter is realized by making full use of the action relationship between recloser and segmenter. By setting the action time and time of recloser and segmenter, the fault location can be accurately located. Therefore, the whole process is simple and easy to implement. However, different types of power grids have different configuration requirements for reclosers, and although the application of reclosers and sectional switches still has advantages in a certain environment, it has its own shortcomings, and the fault location processing time is too long; therefore, it is difficult for the network structure to achieve the best state.

2.1.2 Fault location based on feeder terminal unit (FTU)

In normal operation of distribution network, FTU can collect operation information corresponding to line switches to realize online monitoring of distribution network. In case of failure, FTU can upload relevant information to the control system, and finally find out the fault location through corresponding analysis, and determine the optimal recovery scheme.

According to the current information of each node in the fault zone, the network matrix of the fault zone can be judged and searched out. With the deepening of the research of matrix, the speed of matrix algorithm has been greatly improved, and its application scope is more and more extensive.

2.1.3 Fault current discrimination method

Based on graph theory, this method realizes fault diagnosis according to the expansion model of distribution network. This method can finally get a fault judgment matrix. The concrete steps are to first write the network description matrix according to the expansion model of distribution network. Then, according to the information matrix and the network description matrix, the rough description matrix is normalized, and finally the rough description is obtained. This method is based on the change in power flow. The structure and parameters of the system will change when faults occur. This change makes the calculation and analysis of power flow take more time, which will affect the speed of diagnosis and the quick recovery of faults. In addition, in normal operation, the power flow value of some lines is close to zero, such as online operation, so even if the power flow is used to judge the fault, the accuracy of diagnosis cannot be ensured.

2.2 Data mining

Database technology, artificial intelligence, and mathematical statistics are three powerful technical pillars of data mining research.

2.2.1 Introduction to data mining

Noisy and incomplete data. The essence of data mining is to filter the intuitive information in mass data and discover the inherent unknown information. It often requires constant adjustments to obtain these

unknowns from the data. At the same time, data mining is developed on the basis of multi-disciplinary technology integration, including database, machine learning, statistics, high-performance computing, and visualization.

2.2.2 Data mining process

2.2.2.1 Data preparation

Data mining will mainly extract knowledge from some data, that is, it extracts data related to data mining technology from the database according to user's requirements. In this process, in order to form a real database, some database operations will be used to process the data.

1. In this stage, the actual mining operations are carried out, including the key points of deciding how to generate hypothetical knowledge by selecting appropriate tools to discover knowledge and verify the discovered knowledge.
2. The results are analyzed, explained, and evaluated. Generally, according to the data mining operation, we decide which analysis method to use.

2.2.3 Application of data mining technology in distribution network faults

More and more power supply equipment are diversified, and the detection of various equipment by power supply companies is becoming more and more perfect. With more and more diagnostic information, the application of data mining technology also has a solid foundation.

Because of channel failure, monitoring equipment failure, and other reasons, some parameters of distribution network equipment are lost or even wrong. Traditional expert system cannot use historical data rules to analyze, process, and give the final decision. It is very difficult to find the model that is consistent with the current fault in the complicated historical data. The following processing is needed to determine the power network fault. Cluster analysis is clustering analysis of faults, wherein each fault satisfying the similarity between any two data samples (similarity can be calculated by specific values of sample description attributes, usually expressed by sample distance) is relatively high. Based on cluster analysis, the similarity of distribution network faults is processed and analyzed, and the faults are classified to lay the foundation for the next correlation analysis. Association rule analysis of faults refers to the analysis of the influence and association relations among different data in historical data. By integrating fault areas with certain association rules, a concise expression can be obtained. By comparing the support degree and reliability index under association rules, we can judge the usefulness of association conclusion between faults.

Diagnosis process is a decision-making process. The basis of knowledge will affect the correctness of decision-making. The richer and more reliable the knowledge is, the more correct the decision-making will be. It is difficult for humans to discover the hidden knowledge and rules in front of them, and data mining can make up for the shortcomings of humans in this regard. Therefore, data mining technology is needed in fault diagnosis of distribution network. It is precisely because of this demand that data mining technology is developing continuously. Data mining theory can integrate these methods effectively and compare and analyze them. Thus, the system, function, and process of information mining are defined, which lay a good foundation for the application of various advanced theories and technologies in condition monitoring and fault diagnosis, and makes the whole process clear and easy to implement.

2.3 Application of data mining technology based on information entropy in distribution network fault diagnosis

2.3.1 Information entropy

Shannon refers to the concept of thermodynamics and the average amount of information excluded from redundancy as "information entropy," and gives the mathematical expression for calculating information entropy. There is a measure of the value of information, and more inferences can be made about the flow of knowledge.

Essentially, information entropy reflects the information content represented by events. Therefore, the greater the uncertainty of variables, the more information they can represent, and the greater the value of information entropy:

$$H(x) = -\sum_{i+1}^{n} p(x_i) \ln(p(x_i)).$$

(1)

In formula (1) $p(x_i)$ is the probability of x_i occurrence. At present, information entropy has been widely used in various fields. Entropy is an important choice for constructing evaluation convergence function in path search. When more and more information is known, the probability of the remaining unknown events is lower, and once these unknown events occur, they represent a new search direction. In the aspect of fault diagnosis, different faults are expressed differently at different measuring points. When a fault is expressed differently from other faults at one measuring point, that is, the change caused by the fault at that measuring point is not found at other measuring points, then it shows that the fault can be distinguished accurately by the measuring point, and the occurrence of the fault can be identified when the specific value is detected by the measuring point.

1. Non-negative: $H(X) \geq 0$
2. Certainty: $H(1,0) = H(1,0,0) = H(1,0,0,0) = ... H(1,0,...,0) = 0$
3. $H(XY) = H(X) + H(Y)$
4. Extremum: $H(p_1, p_2,..., p_n) \leq H\left(\frac{1}{n}, \frac{1}{n},..., \frac{1}{n}\right) = \log n$

The calculation of information entropy is very complicated. The information with multiple preconditions can hardly be calculated.

2.3.2 Information entropy introduction

A large number of alarm signals are generated, some of which are related and some are independent. Independent features can provide complementary information, so they should be retained. Relevant features can generate redundant information and increase the computational workload, so they need to be eliminated. Due to the complexity of rule extraction, in order to find the optimal reduction from the reduction combination, the information entropy reduction algorithm is used.

However, outlier detection algorithms based on distance and bias require parameters to be determined in advance, which can lead to wrong conclusions if the parameters are not selected properly.

Since information entropy only depends on the probability of each attribute in the record, the value of the attribute can be either numeric or non-numeric (such as character). That is to say, information entropy can integrate numerical attributes and nominal attributes in a unified framework for processing.

2.4 Bayesian network

Bayesian network technology and Bayesian network have overcome many conceptual and computational difficulties of rule-based systems.

Edge is used to represent the dependency of each variable. Each attribute node in the structure diagram networks represented by directed acyclic graphs can be viewed from both quantitative and qualitative perspectives, which can be more intuitive qualitative analysis of the impact of those variables on its final results. Quantitatively, Bayesian networks can express the dependence of different variables on their parent nodes by using conditional probability distribution, and the relationship of each node in Bayesian networks can be expressed by using joint probability distribution. For example, if X_1, X_2, ..., X_n is used to represent variable nodes in a network, the joint probability distribution of each variable can be obtained by multiplying the conditional probability of each variable using formula (2):

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\pi(X_i)). \tag{2}$$

3 Experiments

When power system fails, it will produce a series of effects, such as abnormal drop in voltage, instantaneous increase in current, and change in public power. After the fault information is transmitted to the fault diagnosis system, various data are transmitted and the transmitted data are analyzed and processed, and finally the location where the fault occurs is diagnosed. The general steps of applying data mining technology based on information entropy to distribution network fault diagnosis are as follows:
1. Original decision table is established.
2. Based on the original decision table of fault diagnosis, an identifiable matrix is established according to formula (3).

$$c_{ij} = \begin{cases} \{a \in A : a(x_i) = a(x_j)\}, & D(x_i) \neq D(x_j), \\ 0, & D(x_i) = D(x_j), \\ -1, & a(x_i) = a(x_j), D(x_i) \neq D(x_j), \end{cases} \tag{3}$$

3. Find out the core attributes and get the reduction results.
4. The average mutual information in attribute reduction results is obtained by using the method of solving average.

$$I(X;Y) = \sum_{x,y} p(x, y) \log_b \frac{p(x, y)}{p(x)p(y)}. \tag{4}$$

5. According to the actual situation, the joint probability of conditional attributes in different fault areas is obtained by using formulas (5) and (6). According to the calculation results, the fault area is determined and compared with the predetermined fault area.

$$P(X|C_k) = \prod_{i=1}^{n} P(x_k|C_k), \tag{5}$$

$$P(X|C_k) = \frac{\frac{1}{N}}{N_{ck} + \frac{N_{xi}}{N}}. \tag{6}$$

4 Discussion

4.1 Fault diagnosis model of distribution network

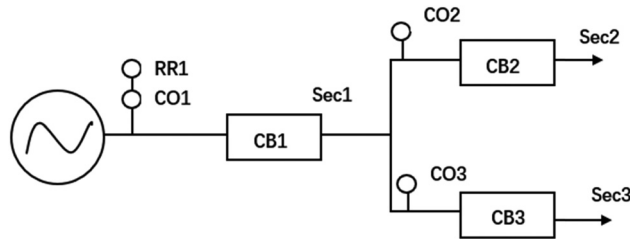Figure 1 is a simple distribution network model with three areas, Sec 1, Sec 2, and Sec 3.

Figure 1: Simple distribution network model.

4.2 Action principle of relay protection

The basic principles of relay protection: 1. When each electrical component has internal and external faults (including normal operation), the difference in current phase or power direction between two sides can constitute the protection of various differential principles, such as longitudinal differential protection, phase difference high frequency protection, directional high frequency protection, and so on. 2. The protection of differential principle can only operate when the internal fault of the protected component occurs, but does not respond to the external fault. So, it is considered to be absolutely selective. When various relay protection devices are constructed according to the above principles, their parameters can be reflected by the current and voltage in each phase, or only by the current and voltage of one of the symmetrical components (negative, zero, or positive sequence).

4.3 Fault diagnosis model of distribution network

Formula (3) is used to form the rules for the establishment of recognizable matrices, i, j = 1,2, ..., n. Conditional attributes are represented by A, while $a(x_i)$ and $a(x_j)$ represent elements of row I and column J in the set of conditional attributes A, respectively. According to the formation rules of the discernible matrix, the $c_{ij} = c_{ji}$ in the above definition matrix is a symmetric matrix, so the upper triangle or the lower triangle is needed. According to the definition, when $i = j$, $c_{ij} = 0$, the diagonal elements of the identifiable matrix are all 0.

Appeal rules can give an identifiable matrix, which is based on the original decision table of fault diagnosis, and the corresponding identifiable matrix can be derived from the rules of the formation of identifiable matrix. According to the rules of core attribute extraction, the elements containing carbon dioxide (CO_2) and carbon trioxide (CO_3) in the discernible matrix are removed, and the retained elements are carbon monoxide (CO_1), risk ratio (RR_1); central cannabinoid (CB_1), RR_1; CB_1, CO_1. Combining the above three sets of elements with CO_2 and CO_3, respectively, the attribute reduction result of the element decision table is achieved.

In general, the number of attribute reduction combinations is not always certain. When there is only one attribute reduction combination extracted, this is the best attribute reduction combination. But when there are multiple combinations in the attribute reduction results, it needs to be operated to find the simplest attribute reduction combination. In this paper, the results of attribute reduction include the values of CO_1, CO_2, and CO_3, which can also be calculated by formula (3). Because the association degree of these three attribute reduction combinations is different, if we need to use Bayesian network model reasoning, we

must assume that each conditional attribute is independent of each other. In order to minimize the influence of independence hypothesis on the results of probabilistic reasoning, it is necessary to find the attribute reduction combination with the least dependency among attributes. Results can be obtained by using the method of solving the average mutual information [formula (4)]. The combination of the simplest attribute reduction is the one with the smallest average mutual information.

According to the formula: (1) CO_1, RR_1, CO_2, and CO_3 are 0.0869; (2) CB_1, RR_1, CO_2, and CO_3 are 0.1192; and (3) CB_1, CO_1, CO_2, and CO_3 are 0.072. It is concluded that the combination of the simplest attribute reduction is CB_1, CO_1, CO_2, and CO_3. A new decision table is generated after the combination of the simplest attribute reduction is obtained, as shown in Table 1.

Table 1: Decision table after attribute reduction

Sample number	CB_1	CO_1	CO_2	CO_3	Fault area
1	1	1	0	0	Sec 1
2	0	0	1	0	Sec 2
3	0	0	0	1	Sec 3
4	1	0	1	0	Sec 2
5	1	0	0	1	Sec 3
6	1	0	0	0	Sec 2/Sec 3
7	0	0	0	0	NO

4.4 Bayesian network model

From Figure 1, a Bayesian network model with five nodes can be built, as shown in Figure 2.
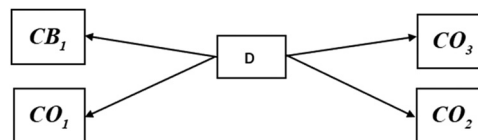

Figure 2: Bayesian network structure diagram.

Next the prior probability of each fault area is calculated based on the existing fault information, as shown in Table 2.

Table 2: Prior probability of each fault region

Fault area	Sec 1	Sec 2	Sec 3	Sec 2/Sec 3	NO
Prior probability	1/7	2/7	2/7	1/7	1/7

According to the Bayesian theorem, in the Bayesian network model, every child node can find the parent node which has a direct relationship with it. In order to simplify the operation process, this article regards each child node and parent node as independent. The joint probability of each conditional attribute taking a specific value under a certain fault condition can be expressed as the product of the conditional probability of each attribute taking a corresponding value under such fault condition, such as formula (5). In the formula, the fault type is C_k, and the condition attribute is x_k. $P(X|C_k)$ can be obtained from the sample of historical fault information record. If there is no fault type C_k and condition attribute x_k in the record,

formula (6) is used. The conditional probabilities of CB_1, CO_1, CO_2, and CO_3 in different fault areas can be calculated from the formulas (5) and (6), as shown in Table 3.

Table 3: Conditional probability of CB_1, CO_1, CO_2, and CO_3 in different fault areas

	Sec 1	Sec 2	Sec 3	Sec 2/Sec 3	NO
$CB_1 = 0$	1/10	1/2	1/2	1/10	1
$CB_1 = 1$	1	1/2	1/2	1	1/11
$CO_1 = 0$	1/13	1	1	1	1
$CO_1 = 1$	1	1/15	1/15	1/8	1/8
$CO_2 = 0$	1	1/19	1	1	1
$CO_2 = 1$	1/9	1	1/16	1/9	1/9
$CO_3 = 0$	1	1	1/19	1	1
$CO_3 = 1$	1/9	1/16	1	1/9	1/9

According to the probability of occurrence of conditional attributes under different fault types, Sec 1 is set as fault area, and the conditional attributes of short circuit protection are $CB_1 = 0$, $CO_1 = 1$, $CO_2 = 0$, $CO_3 = 0$. Condition attributes of the fault are assumed to be $CB_1 = 0$, $CO_1 = 1$, $CO_2 = 0$, $CO_3 = 0$, and then calculated according to the method, the most likely area is Sec 1. According to the calculation results, it can be judged that the fault area is Sec 1, which is exactly the same as the predetermined fault area. The diagnostic error curve of the Bayesian network model is shown in Figure 3.
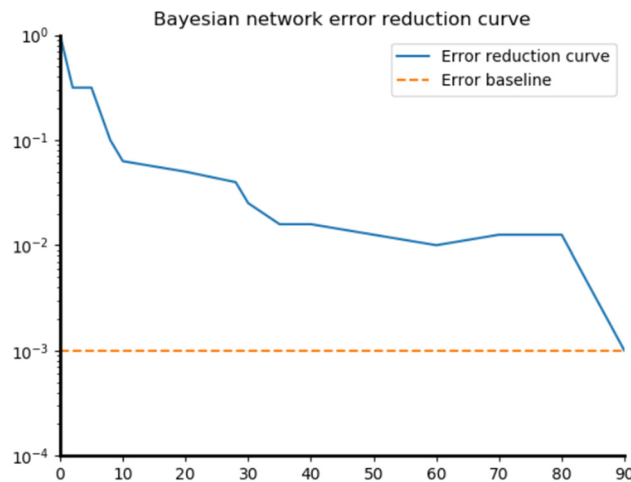

Figure 3: Diagnostic error curve of Bayesian network model.

First, this study compares the shortcomings of various existing methods in practical application. Aiming at the problem of diversification of results, a method of combining rough set theory based on information entropy and Bayesian network is proposed. Table 3 is used as the main tool. Then, the average mutual information of information entropy is used to reduce the feature parameters of samples, solve the problems of too large scale of network, and too slow speed of classification and recognition, so that the structure of Bayesian network is simplified, the training time of network is shortened, and the accuracy of fault identification is improved. The distribution network model is established, and the experimental results are summarized and analyzed. The characteristics of this method are more intuitively displayed. In this study, according to the characteristics of distribution network, Bayesian network is selected as the

classifier. The design and training process of the network are introduced in detail, which was verified by an example and good results were achieved. The sampling frequency is 4 kHz. The number of measurement nodes of the PMU device is related to the compressed sensing algorithm. In order to recover the original signal more accurately, the voltage amplitude changes before and after the fault are calculated, combined with the node impedance, and the reconstruction result is shown in Figure 4.
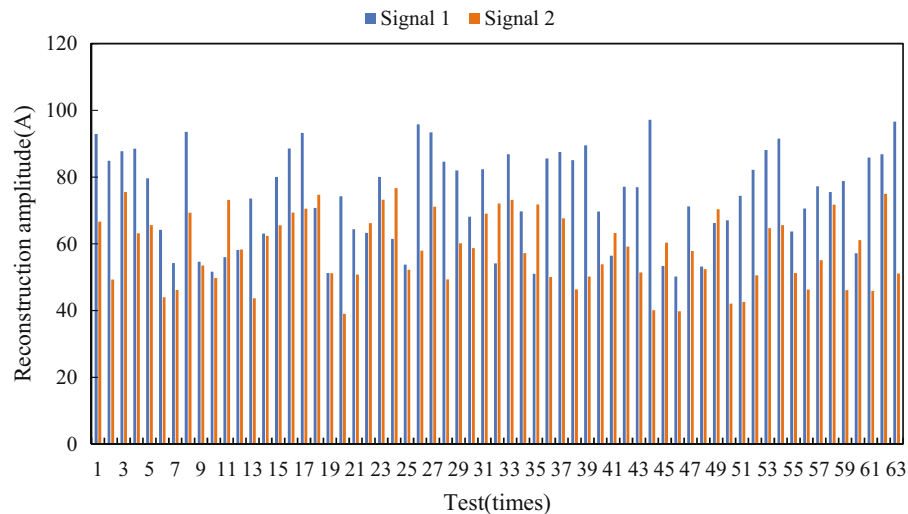

Figure 4: The reconstruction results.

When a fault occurs, the protection and circuit breaker action information is shown in Table 4.

Table 4: Protection and circuit breaker action information

Serial number	Time	Action information
1	16:00:09:307	L19-20 left end current I stage protection start
2	16:00:09:343	L19-20 right-side current stage I protection starts
3	16:00:09:393	CBz0-19 trip
4	16:00:10:089	L2-19 left end current stage II protection start
5	16:00:10:124	CB-21 trip

According to the method of fault location based on the multi-sensor data fusion algorithm based on the enhanced belief divergence measure, the fusion results are shown in Table 5.

Table 5: Fusion results

Element	A1	A2	A3
B3	0.015	0.957	0.028
B4	0.082	0.915	0.003
B5	0.283	0.632	0.085
B6	0.986	0.013	0.001
B7	0.074	0.854	0.072

In the calculation example, the interruption time of all ILs is mainly concentrated in the time period of 7:45–13:15; L#13 was interrupted in the time period of 19:30–24:00; L#23 was interrupted in the time period of 17:00–22:30; L#29 was interrupted at 16:45–23:30. The time of failure is shown in Figure 5.
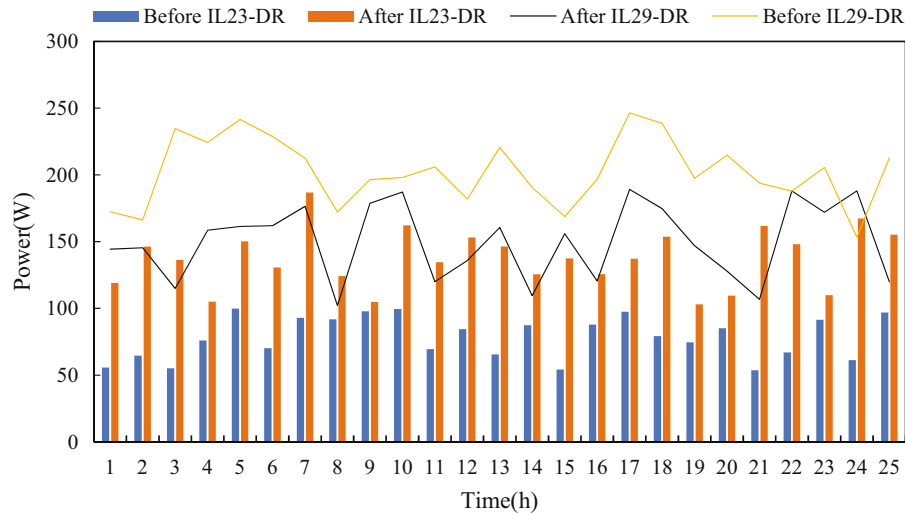

Figure 5: The time of failure.

Reconstruction amplitude of each bus and line within the fault range is shown in Table 6.

Table 6: Reconstruction amplitude of each bus and line within the fault range data

Element	Reconstruction amplitude
BI	83
B2	192
B3	120
B19	482
B20	501
B21	254
L12	141
L23	189

Compared with the original fault conditions, the fault recovery control strategy generated by the method in this study can greatly restore the system load. The control strategy that takes into account the interaction of supply and demand further reduces the load loss on the basis of only considering the network reconfiguration, which proves the generation strategy of this study. The effectiveness of active distribution network resilience under extreme disasters is improved. The total loss of system load before and after the fault recovery control is shown in Figure 6.

5 Conclusion

When the power system fails, it will cause a blackout of large area. However, the faults of the power system, especially the power grid, are inevitable. In order to quickly monitor and eliminate the faults, and quickly
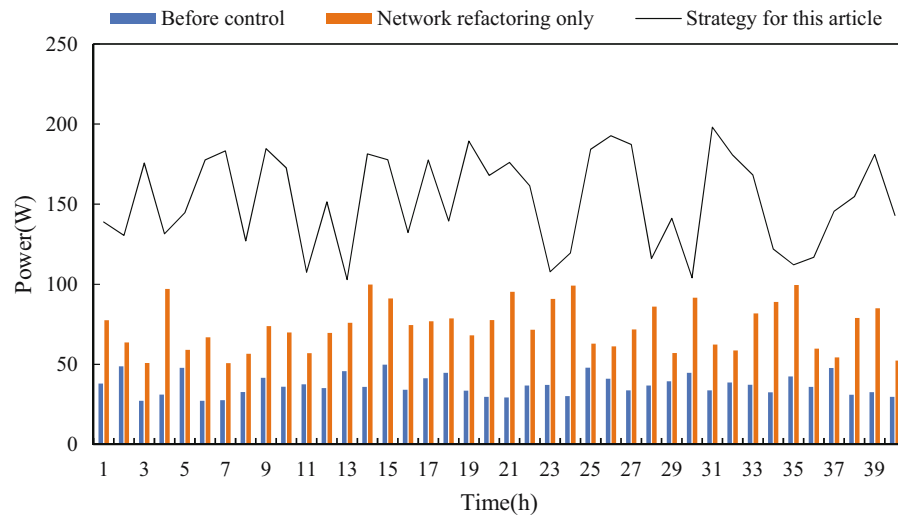
Figure 6: The total loss of system load before and after the fault recovery control.

realize the fault location and fault type identification of the transmission network, so as to facilitate the quick recovery after the overhaul and accident, a high-quality fault diagnosis system is needed.

After the application of the IoT has been popularized in recent years, the number of interface devices used in the sensors of animals, plants, and machinery will far exceed the number of mobile phones. This will further promote the economic development and progress, and create another industry with unlimited potential to promote economic benefits for the society. According to the current demand for the IoT in various fields of the society, in recent years, a large number of sensors and electronic meters used in the IoT will need to be produced, and the production of these requires a lot of manpower and can provide a lot of jobs for the people of the society. The production of these meters at the same time can speed up the production of information technology components. To truly build an efficient and useful IoT, it must have the necessary components. This study tried to apply the methods of information entropy, Bayesian network, and decision tree commonly used in data mining to the above methods through simulation examples. Through the application of the above methods, fault diagnosis becomes more accurate and effective, and the following conclusions are drawn.

By installing more precise and optimized engine equipment in the smart grid, power providers can break the limitations of "traditional" networks and build smart grids that can more actively manage power failures. Power suppliers can use these engine equipment to achieve more optimized management of power failures, manage the complex topology and resource constraints in the grid, and at the same time, through optimized engine equipment and IoT, power generation equipment can be identified more quickly and accurately. In this way, the power supplier can more efficiently arrange the sequence of power failure detection and power equipment restoration. Thus, at least 30% of the power outage time and frequency of power outages can be saved, and the economic losses caused by power outages can be reduced accordingly, and reducing the time and frequency of power outages increases the reliability of the power grid and the satisfaction of customers. Information entropy is a high mathematical method. Through the application of this method, the time of diagnosis process is reduced, and good results are achieved. Bayesian network is used to achieve better fault classification. By training the Bayesian network with samples, choosing an optimal network to realize classification, and by calling the toolbox method, the output of fault report is realized. The application of multi-variable decision tree in distribution network fault diagnosis can effectively simplify and describe knowledge, reduce the redundancy of diagnosis information, greatly improve the diagnosis efficiency, and the results are easy to understand.

Blockchain technology has unique advantages in terms of security. Its high redundancy and data are not easy to be tampered with. It coincides with the strict requirements of the distribution network protection

and control system for information security. Each node in the blockchain network saves complete data, and through a carefully designed consensus mechanism, ensures the consistency of each node's data. The newly generated data are associated with existing data blocks in the form of blocks, forming a chain structure, so that all historical data can be traced and cannot be tampered with. Blockchain technology was first applied in the field of digital encryption currency, and its high security based on cryptographic principles has been widely recognized. With the improvement of power supply quality requirements of users, more scholars and experts will pay attention to the fault diagnosis of distribution network. With the rise of data mining methods in artificial intelligence, more attention will be paid to the application of data mining in distribution network. In this study, although the information entropy is used to carry out profound research on the fault troubleshooting of the distribution network, there are still many short-comings. The depth and breadth of the research in this article are not enough. In the process of this research, the selection and acquisition of experimental data are absolutely ideal. The completeness and effectiveness are not enough. Our academic level research is also limited. This research is still in the preliminary stage. In the future work, we will study suitable methods from more perspectives based on the existing technology and level, and means to continuously improve the effectiveness of the method.

Conflict of interest: Authors state no conflict of interest.

References

[1] Akhavan-Rezai E, Haghifam MR, Fereidunian A. Data-driven reliability modeling, based on data mining in distribution network fault statistics Powertech. IEEE Buchar. 2009. doi: 10.1109/PTC.2009.5281796.

[2] Shu H, Cao L, Xu Z, Liu K. The research of multidimensional information decision mining based on information entropy. Int Forum Inf Technol Appl. 2009. doi: 10.1109/IFITA.2009.559.

[3] Cui S, Shang Z, Wang Q, et al. Research of fault indication system based on WIA industrial wireless technology for distribution network. Control & Decision Conference; 2013.

[4] Adil M, Ali J, Attique M, Jadoon MM, Abbas S, Alotaibi SR, et al. Three byte-based mutual authentication scheme for autonomous Internet of Vehicles. IEEE Trans Intell Transport Syst. 2021;23(7):9358–69.

[5] Adil M, Khan MK, Jamjoom M, Farouk A. MHADBOR: AI-enabled administrative distance based opportunistic load balancing scheme for an agriculture internet of things network. IEEE Micro. 2021;42(1):41–50.

[6] Adil M, Jan MA, Mastorakis S, Song H, Jadoon MM, Abbas S, et al. Hash-MAC-DSDV: mutual authentication for intelligent IoT-based cyber-physical systems. IEEE Internet Things J. 2021. doi: 10.1109/JIOT.2021.3083731.

[7] Farouk A, Alahmadi A, Ghose S, Mashatan A. Blockchain platform for industrial healthcare: vision and future opportunities. Computer Commun. 2020;154:223–35.

[8] Liao H, Yang L, Geng X. Application research of new distributed data mining model based on intelligent agent and web service in digital gas fields. International Conference On Computational & Information Sciences; 2011

[9] Chen S, Jia HE. Research on cost management system of distribution network construction projects based on data mining china. International Conference On Electricity Distribution; 2013.

[10] Li XM, Liu H, Wang WX, Ye Z, Lv HB, Lv ZH. Big data analysis of the internet of things in the digital twins of smart city based on deep learning. Future Gener Computer Syst. 2021;128:167–77.

[11] Li Y, Zuo Y, Song H, Lv Z. Deep learning in security of internet of things. IEEE Internet Things J. 2021;99:1–1142.

[12] Zhang Z. Application and research of data mining technology in communication network environment. Adv Mater Res. 2014;989–994:3814–7.

[13] Zheng CF, Jiang L, Jiang LQ, Wu ZJ. Application and research of Bayesian network in data mining. Adv Mater Res. 2012;532–533:738–42.

[14] Liu M, Qu M, Zhao B. Research and citation analysis of data mining technology based on Bayesian algorithm. Mob Netw Appl. 2016;22(3):1–9.

[15] Liang GY, Gao Y, Zhang L, Xu AR. Research of state estimation in power network based on data mining technology. Appl Mech Mater. 2013;448–453:2594–8.

[16] Yu DR, He HX, He HX, Zhang XX. The research of visualization data mining in industrial data. Appl Mech Mater. 2013;274:5.

[17] Zhao Q, Wu S, Ying C, Jia X. The research of information filtering technology based on Bayesian network. Procedia Environ Sci. 2011;11(Part B):545–51.

[18] Sui JY, Wang YN, Ren Z. The distribution network reconfiguration based on spatial data mining. Appl Mech & Mater. 2013;427–429:2748–51.

[19] Zhou J, Zhang X, Jia C. Application of fault location method based on multi-source fault information in repair of distribution network. North China Electr Power. 2015;1070–1072:6.

[20] Gao Z, Peng Z, Gao N, Chen B. A distribution network fault data analysis method based on association rule mining. Power & Energy Engineering Conference; 2015.

[21] Osamh. IK, Ghaida MA. Energy efficient routing and reliable data transmission protocol in WSN. Int J Adv Soft Comput Appl. 2020;12(3):45–53.

[22] Khalaf OI, Abdulsahib GM, Sabbar BM. Optimization of wireless sensor network coverage using the bee algorithm. J Inf Sci Eng. 2020;36(2):377–86.

[23] Rajendran S, Khalaf OI, Alotaibi Y, Alghamdi S. MapReduce-based big data classification model using feature subset selection and hyperparameter tuned deep belief network. Sci Rep. 2021;11:24138. doi: 10.1038/s41598-021-03019-y.

[24] Khalaf OI, Abdulsahib GM. Optimized dynamic storage of data (ODSD) in IoT based on blockchain for wireless sensor networks. Peer-to-Peer Netw Appl. 2021;14:2858–73. doi: 10.1007/s12083-021-01115-4.

[25] Zhang HP, He HY, Gu MX. Research of fiber-optical fault diagnosis based on support vector machine (SVM) mining. Fifth International Conference on Intelligent Systems Design & Engineering Applications; 2014

[26] Tao Z, Pan YM, Guo KY, Wang ZP. Fault location of distribution network based on immune algorithm. Power Syst Prot Control. 2014;42(1):77–83.

[27] Shu HC, Sun XF, Da-Jun SI. A rough set approach to distribution network fault location based on fault complain call information. Power Syst Technol. 2004;28(1):64–8.

[28] Sheng S, Ma Z, Wu J, Gao N. Distribution network fault restoration based on improved adaptive genetic algorithm. Algorithm Second International Conference on Intelligent Computation Technology & Automation; 2009.

[29] Guo K. Research on location selection model of distribution network with constrained line constraints based on genetic algorithm. Neural Comput Appl. 2020;32(6):1679–89.

[30] Yan W. Application research of data mining technology about teaching quality assessment in colleges and universities. Proc Eng. 2011;15:4241–5.

[31] Wang L, Tong YF. Application research of data mining technology in power dispatching management system. International Conference on Smart Grid & Electrical Automation; 2016.

[32] Zhang WQ. Application research of data mining technology on growth management of forestry. Adv Mater Res. 2014;846–847:995–8.

[33] Zhu Q, Wang Y, Li J. Public transport IC card data analysis and operation strategy research Based on data mining technology. Int Forum Computer Sci Tech Appl. 2009;1422–5; doi: 10.1109/IFCSTA.2009.352.