Alarm Correlation Analysis Method for Smart Power Distribution and Utilization Communication Network Based on Bayesian Networks

Wei lei[1, a]**,**Lu Min[1]**,** Zhu Hong[1]**,**Sun Li-qian[2,b]

[1]State Grid Jiangsu Electric Power Company, No. 215 Shanghai Road, Nanjing, China

[2]Beijing University of Posts and Telecommunications,Beijing,China

[a]38362751@qq.com, [b]liuyqpaper@163.com

Keywords: Electric power communication network, Bayesian network, alarm correlation

Abstract. With the rapid development of smart grid, the alarm management method used in the existing smart power communication network cannot meet the logical requirements of the new business in the network. So this paper introduces an alarm correlation analysis method based on Bayesian network. Firstly, the structure, training method and the learning mode of Bias network are analyzed, and the method of parameter learning is given, structure learning and updating learning of Bayesian network is proposed. Secondly, this paper models smart power distribution network fault and alarm, give the process of analysis, and describes each stage of the Bayesian model. Finally, the high efficiency and reliability of the application of Bayesian network in smart power distribution communication network alarm correlation analysis is verified by the simulation experiment which using JAVA language simulates application scenarios. The experimental results show the superiority of the proposed method.

Introduction

Power distribution and utilization network is the primary way of State Grid for power transmission and the core part of smart grid. The related established power distribution and utilization communication network provide information transmission for the power grid, and help power grid to achieve healthy operation, rapid fault treatment, rational use of equipment, rapid demand satisfaction and sustainable development of electric power. With the rapid development of smart grid, power distribution and utilization communication network not only to inheritance and to perform traditional monitoring, intelligent business, also need to support the new business logical requirements within the network, such as environmental monitoring, operation condition judgment, and expand the business of other types of power utilization process.

Fault management is one of the core functions of the network management system. When the network alarm, the root alarm will trigger a series of related equipment alarm, which makes the surge in the number of alarms, resulting in a large number of useless repeated alarms. A large number of alarms information not only cannot help operation and maintenance staffs locating alarm source and determine the cause of the fault, but will cause the alarm information flooded. In order to solve the above problems, we need to find an alarm correlation analysis method for smart power distribution and utilization communication network. However, the alarm management method of the current smart power communication network is only the simple presentation of the alarm information or quantitative compressed in accordance with the severity of the alarm, cannot help operation and maintenance staffs to intuitively analysis and locate the fault. Therefore, the theory research and practice of the smart Power distribution and utilization network alarm correlation analysis technology is important and necessary.

Related work

A fault within the network may cause our network management system to receive a large number of alarms. The cause of the correlative alarm is the same alarm repeated many times and a number of

different alarms in the system caused by the single fault. At present, the alarm correlation analysis is mainly divided into the following categories[1].

Analysis method based on rule reasoning[2] This method is close to the mode of thinking, more intuitive, easy to understand. But when the network topology changes or add a new network equipment, this method cannot automatically adapt to these changes and it need to artificially add a new rule for analysis. This will increase the dependence of the system on network management and improve the cost of system operation. Especially when the rules are huge, it is difficult to manually sum up the comprehensive rules to adapt to the huge number of alarms. Therefore, the making and maintenance of rule is the main short board for the Analysis method based on rule.

Analysis method based on case-based reasoning When new problems arise, the system can deal with the problem by analyzing the existing similar problems, and after solved the new problem, it will take the new problem as a new case to solve the problems in the future. This method does not require the network management and maintenance staffs to have the specific professional background knowledge. But this method is similar with the method based on rule, which cannot be timely and effective adapt to the changes of network. Once the network structure changes, the accuracy of the case analysis will be likely to decline even completely cannot forecast the analysis.

Analysis method based on model reasoning [3] This is a kind of correlation analysis method based on knowledge model, which is depend on the network topology and the diffusion characteristics of alarm information in the network, through the establishment of accurate network model and the warning diffusion model, can quickly and accurately deal with the alarm information in the target model, and give the correlation analysis results. However, with speed up construction of power grids, it has been unable to establish an accurate and stable topology structure and alarm diffusion model, which makes the method based on model reasoning is not suitable for power grid field.

Analysis method based on Association Rule Mining This model refers to the correlation between the huge data in the analytic database. The mining process is generally divided into two steps, the first step is to find all frequent item sets, then deduce the strong association rule which meet the requirement from frequent item sets. The commonly used algorithms of association rule mining are Apriori algorithm, FP-growth (Frequent-Pattern tree growth) algorithm and so on[5]. FP-growth can produce a huge condition tree in the intermediate steps, and the space complexity is also higher.

Analysis method based on Bayesian network[6] Bayesian network is a combination of traditional concept and graph theory, is related to statistical knowledge to solve the uncertain problem, and is the method to solve uncertain problems based on the relevant knowledge of Statistics and an important way to solve the problem of uncertain data. The deep research of Bayesian statistics and graph theory provides a sufficient theoretical support for this method, which makes the Bayesian network widely used in fields such as AI and ML. Analysis method based on Bayesian network has a strict probabilistic theory foundation, which is an effective method to solve the uncertain problems.

In this paper, the following sections are organized as follows: Section 3 describes the network problems such as fault causes and location. Section 4 established a two layer Bayesian network model of the alarm correlation analysis. Section 5 gives the analysis steps and algorithm of the alarm correlation. Section 6 is the simulation experiment. Section 7 is the conclusion.

Problem description

The fault of smart power distribution and utilization communication network usually need to be displayed through some phenomena or alarm, and cannot be directly observed by the network management system. With the increasing scale of the network, the fault correlation problem in power distribution and utilization communication network showed intricate uncertain and other characteristics. How to use the efficient fault correlation analysis method to filter the useful information from the large number and wide range of alarms to realize the timely and accurate fault location has become the focus of the research on the current fault correlation problem.

The current fault location technology is through the analysis of the existing fault case data and usually set up a combination model based on fault-alarm event tree and code thinning model to solve

the problem. The method of code thinning put the alarm as the password to recognize the fault at this time the steps of fault location is actually the steps of alarm data decoding. However, code thinning method requires a deterministic model structure between faults and symptoms, which is difficult to achieve in the complex smart power distribution and utilization communication network. Therefore, the above methods have certain limitations and uncertainties, and cannot get a higher degree of accuracy of the fault location results. In order to solve the problem, this paper uses the Bayesian fault correlation analysis method to realize the location of analysis of the fault causes.

Alarm Correlation Analysis model based on Bayesian network

The fault of smart power distribution and utilization communication network usually need to be displayed through some phenomena or alarm, and cannot be directly observed by the network management system. With the increasing scale of the network, the fault correlation problem in power distribution and utilization communication network showed intricate, uncertain and other characteristics. How to use the efficient fault correlation analysis method to filter the useful information from the large number and wide range of alarms, to realize the timely and accurate fault location has become the focus of the research on the current fault correlation problem.

In order to solve the above problems existing in the power distribution and utilization communication network, this paper establishes an alert correlation analysis of two layer Bayesian network model, as shown figure 1.
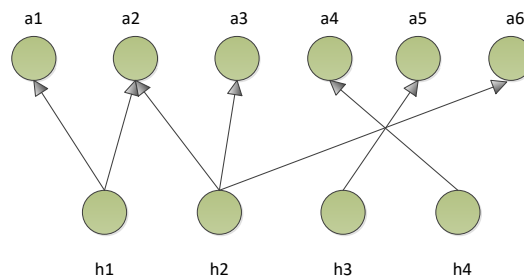

Fig.1 Sketch map of hierarchical model

Model contains two sets, network fault set and alarm set, fault set is the set of all faults in power distribution and utilization communication network; alarm set is the set of all alarms caused by the fault in power distribution and utilization communication network, and using probability weighted to express the relationship between alarm and fault. The relationship between the two sets, which is the relationship between the alarm and the fault, is represented by a priori probability.

Define the fault set $A = \{a_1, a_2, \ldots, a_m\}$, in which m represents the number of faults in the network.

Define the alarm set $H = \{h_1, h_2, \ldots, h_n\}$ in which n represents the number of alarms in the network.

Define the correlation analysis model as S= {H, A}.

Using the above definition of the fault set and alarm set, the fault set $H = \{h_1, h_2, \ldots, h_n\}$, in which m represents the number of faults in the network; alarm set. $A = \{a_1, a_2, \ldots, a_n\}$, in which, n represents the number of alarms in the network.

By the Bayesian formula, the conditional probability of the alarm a_i caused by the fault h_i is

$P(H = h_i \mid A = a_i)$

$$P(H = h_j \mid A = a_i) = P(h_j)P(a_i \mid h_j) \Big/ P(a_i) \qquad (1)$$

Among them, $P(h_j)$ represents the prior probability of the alarm corresponding to the fault h_j;

$P(a_i \mid h_j)$ represents the conditional probability of a_j, $P(a_j)$ represents occurrence probability of a_i.

However, taking into account the calculation of $P(H = h_i \mid A = a_i)$, $P(a_j)$ are the same for the same alarm, therefore, this value does not need to be calculated in comparison.

For the calculation of conditional probability table, assuming that $\{d_1, d_2, \ldots, d_n\}$ is independent and identically distributed, λ is the parameter, then the model can be expressed as

$$g(d_1, d_2, \ldots, d_n | \lambda) = g(d_1|\lambda) \cdot g(d_2|\lambda) \cdot \ldots \cdot g(d_n|\lambda) \tag{2}$$

Likelihood function is

$$L(\lambda | d_1, d_2, \ldots, d_n) = g(d_1, d_2, \ldots, d_n | \lambda) = \prod_{i=1}^{n} g(d_i | \lambda) \tag{3}$$

Log on both sides，

$$\ln L(\lambda | d_1, d_2, \ldots, d_n) = \sum_{i=1}^{n} \ln g(d_i | \lambda) \tag{4}$$

$$\hat{l} = \ln L / n \tag{5}$$

The left end of the formula 3-5 is the log likelihood; the left end of the formula 3-6 is the average log likelihood.

$$\bar{\lambda} = \arg \max \bar{l}(\lambda | d_1, d_2, \ldots, d_n) \tag{6}$$

$\bar{\lambda}$ is the parameter values estimated by the parameter learning.

Calculate the conditional probability of the alarm a_1 caused by the fault h_1, and use the existing historical data in the database as the training set, assume alarm is binomial distributed and mutual independent, let estimated parameter is α, then the maximum likelihood function is:

$$L(\alpha \| d_1, d_2, \ldots, d_n) = \alpha^a (1-\alpha)^b \tag{7}$$

The number of alarms caused by this fault is a, the number of alarms not caused by this fault is b. The conditional probability table of the network can be improved by the similar computing process, and the network parameters learning can be realized.Final determination of suspected fault is to take the highest probability.

Algorithm design based on Bayesian network

After the system collects the alarm data, the alarm correlation analysis is conducted. The analysis steps are as follows:

1)According to the relationship between the fault set and the alarm set, select and determine the possible corresponding fault sub sets of all alarms in the list of the current warning data.

2)According to each alarm, introducing the Bayesian theory, in turn, the prior probabilities of all candidate faults transform into the conditional probability of the current situation.

3)After analyzing all the transformation results, the fault of the maximum posterior probability is the correlative fault, and the results of the correlation analysis are obtained。

4)The analysis results are recorded for parameter learning and updating the priori probability.

Simulation and experiment

The reasoning tool used for Bayesian network is Bayesian Network tool, which is used to carry out correlation analysis of network alarm data. Before using Bayesian network tools for correlation analysis, the first step is to construct a Bayesian network to express the dependent relationship of the power grid topology node variables.

The simulation uses Java language to simulate alarm reporting of the factory network management system, docking with the collection module of this system, after updated in accordance with the original time sequence, interval, part of the historical data obtained from the factory network management will be send to the alarm correlation analysis module for the real data test.

The test content includes two items; both of them are horizontal contrast to the original rule-base method. It uses 7532 alarm extracted from the database of this correlation analysis system for the test, the original data corresponds a total of 306 real faults.

The accuracy of correlation analysis for different number of nodes in the network.

Time consuming of association analysis for different number of nodes in a network.

Table 1 Comparison of the accuracy of different nodes

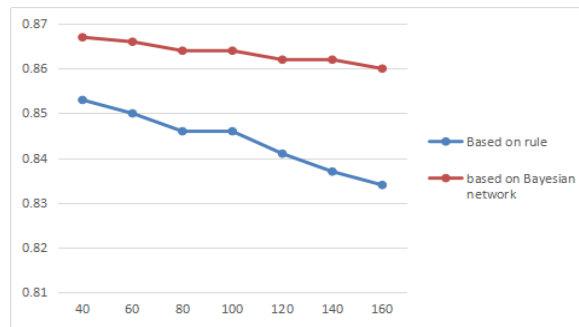| Node number accuracy rate	40	60	80	100	120	140	160
Based on rule	0.853	0.850	0.846	0.846	0.841	0.837	0.834
based on Bayesian network	0.867	0.866	0.864	0.864	0.862	0.862	0.860

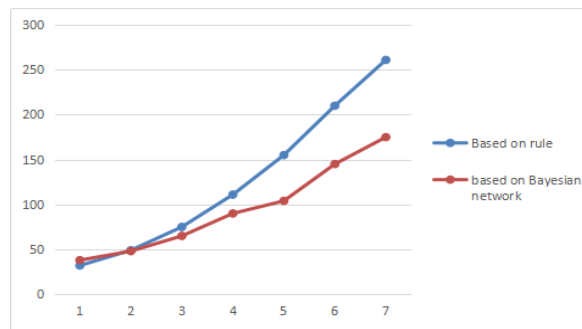Fig. 2 Comparison of the accuracy of different nodes

Fig. 3 Time-consuming comparison of operation with different nodes

Table 2: Time-consuming comparison of operation with different nodes

Node number Time	10	20	30	40	50	60	70
Based on rule	32	49	75	111	155	210	261
based on Bayesian network	38	48	65	90	104	145	175

Conclusion

At first, this paper study Bayesian network structure and training method, presents the learning method of Bayesian network parameter learning, structure learning and update leaning, establishes the alarm correlation analysis model based on Bayesian network. Based on the analysis model, the alarm correlation analysis algorithm based on Bayesian network is designed, the correlation between the alarm phenomenon and the reason of the fault is calculated, and thus the fault location is realized. At last, the effectiveness and accuracy of the algorithm are verified by simulation test. It provides a theoretical basis for subsequent system implementation.

Reference

[1] Deng Xin. The Research of Key Technologies and Applications of Alarm Correlation in Optical Transport Networks [D]. Beijing University of Posts and Telecommunications, 2008.

[2] Hu Yi-fei. Research on Alarm Data Processing Technology in Computer Network [J].FUJIAN COMPUTER, 2006 (11): 32-33.

[3] Bi Hong-yan, Fang Ge, Wang Gui-qin and so on, Mental Model Theory and Related Research in Deductive Reasoning [J]. Psychological Science,2001, 24(5): 595-596.

[4] Yu Ping，Wang Ji-wen, Research of an Improved Apriori Algorithm [J]. JOURNAL OF LANGFANG TEACHERS COLLEGE: natural science edition, 2009, 9 (4):18-19.

[5] An Huan, Research of Communication Network Alarm Correlation Based on Data Mining [D]. Beijing University of Posts and Telecommunications, 2012.

[6] Tang Ai-guo, Wang Ru-long, Hu Chun-hua. Application of Bayesian Network in Software Project Risk Assessment [J]. Computer Engineering and Applications, 2010, 46(7):62-65.