Power Environment Warning Prediction Model Based on Big Data Association Rules

Dongfang Zhang*, Liangliang Yu, Ou Wang and Liang Ning
State Grid Corporation of China, Liaoning Electric Power Supply Co., LTD, Shenyang 110000, Liaoning, China

*Corresponding author e-mail: 1027445580@qq.com

Abstract. With the rapid development of power communication network, power production and operation put forward higher requirements for the stability of power communication network. The power environment system can collect the operation status of communication power supply, equipment and other related alarm information as the basis for troubleshooting. However, due to the failure of a device, there will be some columns of linkage alarm, which increases the difficulty of fault location. In order to solve the above problems, this paper uses the method of big data association rules to carry out data deep mining, and makes in-depth analysis on the location and prediction of power environment fault alarms. Through the correlation between power environment alarms, this paper puts forward the analysis and prediction method of power environment alarms, and realizes the prediction of power environment alarms through the analysis of big data association rules. Based on the in-depth analysis of the Apriori algorithm, this paper designed an improved Apriori algorithm which is suitable for power environment alarm prediction, and greatly improves the efficiency of the algorithm.

Keywords: Big data, Association Rules, Power Environment, Alarm Prediction

1 Current situation of power environment alarm

During the operation of the power environment system of power communication network, a large number of alarms are generated at every moment. Because the network management systems of different manufacturers are different, their alarm formats and information are also different. In addition, one kind of fault alarm may cause other equipment to give alarm. Therefore, the number level of alarm is very large, and the relationship between them is very complex, and it is impossible to completely rely on manual maintenance. In order to effectively reduce the workload of operation management personnel and improve the work efficiency, it is necessary to analyze the correlation between alarm data, and use appropriate data mining methods to find association rules, on the basis of ensuring the accuracy of the rules, to achieve alarm filtering and accurate fault location.

When the power environment equipment gives an alarm, it means that there may be, but not necessarily a fault. In fact, the alarm event contains the information of the abnormal state in the current network, which represents the response of the equipment to the abnormal event. When the power environment is abnormal, all kinds of equipment will generate a series of alarms. However, not all alarms can clearly indicate the cause of the fault, so it is necessary to analyze the correlation of alarm data generated by the system to determine the root cause of the fault. At the same time, with the development of big data research, we can use big data association rule analysis algorithm to predict power environment alarm and improve the operation level of power information communication network.
2. Big Data Association Rule Algorithm

2.1 Research Status at Home and Abroad

Association rule algorithm was first proposed by Agrawal et al. In the following decades, after continuous improvement and improvement, it has developed a variety of algorithms, which has been the research hotspot of experts and scholars at home and abroad. At present, in the field of big data association rule mining, there are three popular association rule algorithms: Apriori algorithm based on frequent item set, FP growth algorithm derived from it, and parallel partition algorithm based on partition. Among them, the most influential algorithm is Apriori algorithm, for which people have done a lot of research to improve the efficiency of the algorithm.

In the aspect of applying association rule algorithm to network alarm prediction, more and more professionals are attracted to study. In literature [1] [2], malicious attacks on network are detected by association algorithm; in literature [3] [4], sequence mining mode is proposed based on association mining, which can better predict the occurrence of alarm. However, the research on the application of correlation algorithm to the prediction of power environment alarm is rare, and the distinction and pertinence of different alarm types are not considered basically. In this paper, through the analysis of different aspects of power environment alarm, we can have a more accurate understanding of the alarm, analyze and obtain the strong correlation alarm with alarm symptoms, and propose a big data association rule. Then the improved Apriori algorithm of dynamic environment alarm prediction is proposed.

2.2 Definition of Association Rules

Data association analysis involves the following definitions: dataset, transaction, project, item set, support, frequent item set and confidence.

1. Dataset, transaction, project
The initial data collected is filtered, and the resulting database is dataset $D$.

$$D = \{t_1, t_2, ..., t_k, ..., t_n\} ,$$  \hspace{1cm} (1)

2. $t_k = \{i_1, i_2, ..., i_p\} (k=1,2,...,n)$;  \hspace{1cm} (2)

2. Itemsets
$I = \{i_1, i_2, ..., i_n\}$ is the set of all items in transaction set $D$. any subset $X$ of $I$ is called the item set in $D$, when $|X| = k$ is called Set $X$ a set of Kterms.

3. Support, Frequent itemsets
The transaction count of item set $X$ contained in data set $D$ is called the support number of item set $X$, and the percentage of the support number of $K$ item set to the number of transactions in $D$ is called the support degree

$$\text{support}(X) = \frac{\delta_X}{|D|} \times 100\%$$ \hspace{1cm} (3)

In the process of data mining, the threshold of support will be adjusted according to the needs of the situation and the actual results, which is called minimum support. If the support of itemset $X$ meets the specified minimum support, itemset $X$ is called frequent itemset; if $|X| = k$, itemset $X$ is called frequent itemset.

4. Confidence
Confidence refers to the percentage of transactions with $B$ in transaction set $D$ containing $A$ in transactions with $A$, i.e. conditional probability:

$$\text{confidence}[A \rightarrow B] = \frac{\text{support}_C(A \cup B)}{\text{support}_C(A)}$$ \hspace{1cm} (4)

Support stands for the universality of association rules in transaction sets, indicating that rules do not happen by chance; confidence stands for the frequency of occurrence of $B$ in transactions containing $A$, that is, the reliability of rules in transaction sets.
There are many kinds of association rules: Boolean association rules, quantitative association rules, single dimension association rules, multi unique association rules, etc. the analysis of alarm problem in this paper belongs to the mining of Boolean association rules. In this paper, the algorithm of mining alarm association rules is divided into two steps:

1. Generating frequent itemsets: finding all itemsets that meet the minimum support threshold, i.e. frequent itemsets;
2. Generate alarm association rules: extract rules greater than the confidence threshold from the set of frequent items found in the previous step.

2 Alarm correlation analysis of Power environment based on improved Apriori algorithm

In the data preprocessing stage, the alarm mining data is analyzed and extracted, and transformed into a data format suitable for data mining algorithm processing. The main work includes data cleaning, data extraction, data processing and data conversion.

1. Data cleaning
   Data cleaning will not be suitable for data mining noise data removal, such as artificial alarm data, noise data, etc., only retain the real useful alarm data for data mining.
2. Data extraction
   Alarm table (including historical alarm table and active alarm table) contains many alarm related fields. Alarm data mining does not need all fields, just extract the key fields related to mining. For example, alarm name, alarm object, alarm time, alarm title and other fields.
3. Data processing
   The main work of this stage is to further process the alarm data extracted in the previous stage to prepare for generating alarm transaction data.

3 Simulation experiment and result analysis

227 fault records are collected in this simulation, and a mining database with 200 records and a test database with 27 records are established. Simulation experiments are carried out on the computer. The structure of the mining database and the testing database is completely consistent, the difference is that the mining database is grouped according to the fault nature in the preprocessing, which is used for mining and obtaining rules, while the testing database has no fault nature information. Used to verify the validity and generality of rules.

| Number of alarm records | Number of alarm types | Algorithm time /s | Number of generated rules |
|-------------------------|-----------------------|-------------------|--------------------------|
|                         |                       | Apriori           | improved Apriori         | Apriori       | improved Apriori |
| 300                     | 18                    | 16.457            | 0.159                    | 37           | 18              |
| 600                     | 20                    | 132.842           | 0.847                    | 39           | 18              |
| 1000                    | 27                    | 457.528           | 2.915                    | 44           | 18              |

The above experiments show that: (1) The efficiency of the improved Apriori algorithm is greatly improved, and the generated rules are more concise. (2) It has the best robustness. The number of mining rules is mainly related to the credibility index. In the case of missing information, the effective rules can still be mined by reducing the credibility properly.

4 Conclusions
In the process of power environment alarm prediction, the Internet of things technology, artificial intelligence technology and big data technology are combined to realize the analysis and research of power environment alarm prediction. In this paper, the big data technology is applied to power environment health monitoring and diagnosis by means of power environment alarm monitoring, and the Association rule algorithm is used to realize the prediction and analysis of power environment alarm.

Acknowledgements

This work was supported by State Grid Corporation of China, Liaoning Electric Power Supply Co., LTD2019 Mass Entrepreneurship Project

References

1 Tjhai G C, Furnell S M, et al. A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algotithm[J]. Computers & Security, 2010, 29(6): 712-723.
2 Yu M, Li W, Li-jin W C. A practical scheme for MPLS fault monitoring and alarm correlation in backbone networks[J]. Computer Networks, 2006, 50(16): 3024-3042.
3 Chao C S, Yang D L, Liu A C. An automated fault diagnosis system using hierachical reasoning and alarm correlation[J]. Journal of Network and Systems Management, 2001, 9(2): 183-202.
4 Chen Y, Lin Q, Tu Z. Pruning Redundant Alarm Correlation Patterns[J]. Journal-CHINA Universites of posts and telecommunication, 2002, 9(2): 45-48