

# Defect Analysis of Secondary Equipment Based on Power Dictionary and Apriori Algorithm

Jinshuang Mu<sup>1\*</sup>, Zhanjun Gao<sup>2</sup> and Fengyuan Zhou<sup>3</sup>

<sup>1</sup>Key Laboratory of Power System Intelligent Dispatch and Control, Ministry of Education Shandong University, Jinan, Shandong, 250061, China

<sup>1</sup>Key Laboratory of Power System Intelligent Dispatch and Control, Ministry of Education Shandong University, Jinan, Shandong, 250061, China

<sup>1</sup>Key Laboratory of Power System Intelligent Dispatch and Control, Ministry of Education Shandong University, Jinan, Shandong, 250061, China

**Abstract:** In order to improve the operation level and maintenance efficiency of the secondary equipment in the power system, based on the historical defect data, starting from the efficiency of data processing, power system need to build electricity dictionary. In the process of describing and processing the defect data, based on the electric power dictionary, the key characteristics of the defect data can be effectively extracted. From the perspective of data mining, this paper use Apriori algorithm to correlate and analyze the defect data, establish a analysis model for the secondary equipment defect data. Take a provincial electric power company's secondary equipment historical defect data mining as an example, describes the application process and analysis method of Apriori algorithm. The results show that the algorithm can effectively dig out familial defects and find the weakness of the equipment, it has a certain guiding role for the improvement of equipment performance and secondary equipment operation, maintenance and overhaul.

## 1 Introduction

On the one hand, during the long-term operation of power equipment, massive heterogeneous and polymorphic data will be generated in the process of power grid operation, maintenance and management [1]. These historical operating data have not been fully utilized, only as a guide to eliminate faults. Data mining and correlation analysis play an important role in improving the operation and maintenance level. However, the current data mining only exists in the simple induction, statistics stage still lacks theory. On the other hand, the power system has accumulated a large amount of text data in the process of maintenance and operation. From the perspective of data structure types, the data in the power system can be divided into purely digital structured data, sound, image, and text. Historical defect data is often recorded in text form. Therefore, it is necessary to establish a set of standard electric power dictionary, so it can effectively extract the feature quantity of text data.

At present, data mining technology has been widely used in various fields of electrical engineering, Reference [4] adopts the idea of data mining and typical scene simulation, proposes a network loss analysis method based on hybrid cluster analysis. However, data mining is not widely used in the field of secondary equipment, mainly focusing on the evaluation of the secondary equipment status. This paper proposes to use the Apriori

algorithm combined with the power dictionary method to conduct data mining and family defect analysis. This paper analyzed the basic idea of Apriori algorithm in familial defects, and based on Apriori algorithm, established an analysis model of secondary equipment familial defects. Taking the historical defect data of the secondary equipment as an example, this paper describes the data mining and defect analysis methods of the historical defect data of the Apriori algorithm.

## 2 Data preprocessing based on power dictionary

### 2.1 Building Power Dictionary

In the analysis process, the power dictionary was established by using the historical defect data of secondary equipment. After data preprocessing, the data sample contains 4808 sets of fault data. Using this data to find keywords, this paper set up power dictionary of secondary equipment. This is shown in Table 1.

\*Corresponding author's e-mail: Mjs0216@mail.sdu.edu.cn

**Table 1.** Partial data of power dictionary.

|      |                                  |  |      |                   |      |              |
|------|----------------------------------|--|------|-------------------|------|--------------|
| code | Equipment category               |  | code | Defect level      | code | Manufacturer |
| a1   | Transformer protection           |  | b1   | general           | e1   | A            |
| a2   | Shunt reactor protection         |  | b2   | serious           | e2   | B            |
| a3   | Operation box                    |  | b3   | urgent            | e3   | C            |
| a4   | Series compensation protection   |  |      |                   | e4   | D            |
| a5   | Circuit breaker protection       |  |      |                   | e5   | E            |
| a6   | Generator protection             |  |      |                   | e6   | F            |
| a7   | Traveling wave ranging           |  |      |                   |      |              |
|      |                                  |  |      |                   |      |              |
|      |                                  |  |      |                   |      |              |
| code | Cause of defect                  |  | code | Defect site       |      |              |
| c1   | CPU failure                      |  | d1   | CPU               |      |              |
| c2   | GPS module failure               |  | d2   | DPU               |      |              |
| c3   | Board failure                    |  | d3   | Protection screen |      |              |
| c4   | Acquisition card failure         |  | d4   | goose board       |      |              |
| c5   | Parameter tuning error           |  | d5   | GPS               |      |              |
| c6   | extended active duty             |  | d6   | Sampling board    |      |              |
| c7   | Program failure                  |  | d7   | operating system  |      |              |
| c8   | Printer failure                  |  | d8   | Operation box     |      |              |
| c9   | Wire failure                     |  | d9   | LCD panel         |      |              |
| c10  | Power failure                    |  | d10  | Control panel     |      |              |
| c11  | Unqualified raw material quality |  |      |                   |      |              |
| c12  | Software design is unreasonable  |  |      |                   |      |              |
| c13  | Component damage                 |  |      |                   |      |              |
| c14  | Strong electric field            |  |      |                   |      |              |
| c15  | Program firmware                 |  |      |                   |      |              |
| c16  | Soft fault                       |  |      |                   |      |              |

## 2.2 Preprocessing power equipment defect data

Different from texts in other fields, there are a large number of mixed Chinese, English, numbers and unit symbols in the defect text, and the length of the defect text

is different. Under these circumstances, it is not conducive to the extraction of text defect features. Therefore, based on the power dictionary, keyword query can quickly and effectively extract the key information characteristics of the historical defect data. According to the power dictionary, table 2 is that extracts the key feature information from the text-type historical defect data.

**Table 2.** Historical defect database D

| TID  | Equipment category         | Defect level | Cause of defect                  | Defect site   | Manufacturer |
|------|----------------------------|--------------|----------------------------------|---------------|--------------|
| 1    | Generator protection       | general      | Unreasonable software design     | CPU           | C            |
| 2    | Traveling wave ranging     | general      | Unqualified raw material quality | LCD panel     | A            |
| 3    | Transformer protection     | serious      | Soft fault                       | CPU           | D            |
| ...  | ...                        | ...          | ...                              | ...           | ...          |
| 4808 | Circuit breaker protection | urgent       | Strong electric field            | Control panel | F            |

Text data cannot be mined by using Apriori algorithm, shown in Table 3. it should use power dictionary to encode defect data, as

**Table 3.** Coded defect database D'

| TID | Equipment category | Defect level | Cause of defect | Defect site | Manufacturer |
|-----|--------------------|--------------|-----------------|-------------|--------------|
|-----|--------------------|--------------|-----------------|-------------|--------------|

|      |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|
| 1    | a1  | b1  | c6  | d1  | e3  |
| 2    | a10 | b1  | c9  | d4  | e1  |
| 3    | a2  | b2  | c2  | d1  | e4  |
| ...  | ... | ... | ... | ... | ... |
| 4808 | a6  | b3  | c7  | d7  | e6  |

### 3 Association Rule and Data Mining Algorithm

#### 3.1 Basic concepts of association rules

A transaction involved in an association rule is called a project. The collection of items is called item set I. The itemset composed of k transactions is k itemset, where k is the length of the itemset. The entire sample constitutes the sample database  $D'$ . Two important indicators for evaluating the pros and cons of as are Support and Confidence. Where support is the probability of rule occurrence, confidence is the degree of reliability of the rule. The association rule R can be expressed as:

$$R: A \rightarrow B \tag{1}$$

$$Support(A) = \frac{Count(A)}{D} \tag{2}$$

$$Support(A \rightarrow B) = \frac{Count(AB)}{D} \tag{3}$$

$S_{min}$  is minimum support for association rules, indicates the minimum requirements that need to be met in the process of mining frequent item sets.  $C_{min}$  is minimum confidence of association rules, indicates the minimum reliability that the association rules need to meet. If  $S \geq S_{min}$  and  $C \geq C_{min}$ , it is strong association rule.

#### 3.2 The basic principle of the Apriori algorithm

Apriori algorithm is a common algorithm for mining association rules. Its core is to generate candidate item sets and frequent item sets. Scan the database through a layer-by-layer search method, K itemset generate K+1 itemset. The main steps are divided into two parts:

- 1) Through the method of searching layer by layer, it generate frequent item sets according to the minimum support.
- 2) It retains the item sets that meet the minimum support and minimum confidence, generates strong association rules.

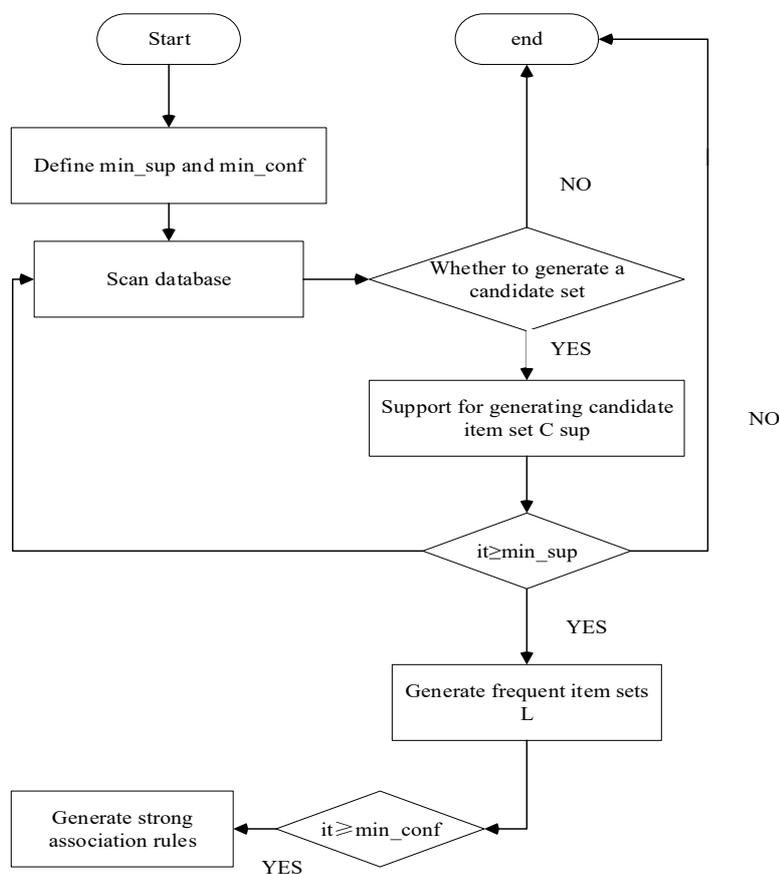


Figure 1. Apriori algorithm flow chart

## 4 Defect Analysis and Modeling of Secondary Equipment Based on Apriori Algorithm

### 4.1 Select the characteristic quantity of defect data

In the process of data mining, considering the manufacturer and equipment type of the secondary equipment can help analyze the family defects. Considering the causes and defects to find out the weaknesses of the equipment. It has a certain guiding effect on the maintenance of secondary equipment. Considering the defect level can distinguish the difference between different defect samples. Building item sets  $I$  for data mining to indicate the defect information of secondary equipment, as shown in formula (4):

$$I = (T, G, C, P, F) \quad (4)$$

Each defect sample is composed of these 5 types of defect information, mining association rules for defect sample data through Apriori algorithm, frequent item sets are at most 5-item sets. Taking the historical defect data of the secondary equipment as an example, it describes the

data mining and defect analysis methods of the historical defect data of the Apriori algorithm. And through further analyzing the results of the association analysis, it can obtain results that are instructive for condition evaluation of the secondary equipment. After simple processing of the power company's secondary equipment defect data, it can get 4808 sample data. The sample data group includes 30 manufacturer, equipment types include 21 categories such as transformer protection, fault recorder, etc. The reasons for the defects include 103 items such as liquid crystal display failure, component damage, strong electric field action, parameter setting error, etc. There are 108 kinds of defective parts including CPU board, liquid crystal display panel, power supply board and so on.

### 4.2 Analysis example of secondary equipment defect based on strong association rule

It uses the Apriori algorithm to mine the historical defect data with association rules, The minimum support is 1.5%, and the minimum confidence is 60%, then scanning the database, the final result contains 216 strong association rules. According to confidence, it is sorted from high to low, taking the top 10 strong association rules with as an example to analyze the results.

**Table 4.** Strong association rules based on Apriori algorithm

| code | Strong association rules   | support/% | confidence/% |
|------|--|-----------|--------------|
| 1    | Manufacturer D→fault recorder                                      | 8.91      | 98.34        |
| 2    | LCD display failure→LCD panel                                      | 12.87     | 97.46        |
| 3    | Manufacturer D, components damage→CPU board                        | 6.93      | 94.62        |
| 4    | Bus bar protection, component damage →manufacturer C               | 3.97      | 94.32        |
| 5    | LCD panel, manufacturer A→transformer protection                   | 7.92      | 91.28        |
| 6    | Transformer protection, LCD panel→ components damage               | 3.83      | 88.65        |
| 7    | Component damage, manufacturer B→fault recorder                    | 8.91      | 87.50        |
| 8    | Transformer protection, component damage, manufacturer A→LCD panel | 10.89     | 84.36        |
| 9    | Manufacturer C→components damage                                   | 12.65     | 83.31        |
| 10   | Manufacturer A, the LCD panel→components damage                    | 9.34      | 80.03        |

According to strong association rules 4 and 9, the equipment produced by manufacturer C is prone to component damage and defects, the equipment produced by this manufacturer has a considerable degree of component quality problems, the confidence is as high as 83.31%. This manufacturer C should pay attention to the quality of components when producing equipment.

According to strong association rule 8 and 10, the quality of the LCD panel produced by manufacturer A also has a certain degree of defects, there is a strong correlation between component damage and LCD panel, confidence degree reaches 83.31%. It can be judged that manufacturer A has certain familial defects.

## 5 Conclusions

This paper builds a local power dictionary based on the historical defect data. And it combines with the Apriori algorithm, studies the association rule mining technology of historical defect data. As form of numerical example

analysis, it applies the Apriori algorithm to the defect analysis. Through data mining, it obtains the following conclusions.

1) Through the mining of historical defect data, it can obtain familial defects. It has a certain guiding role in the maintenance and operation of the same manufacturer's equipment in the power grid. And it is helpful for manufacturers to concentrate on improving equipment performance.

2) Through the mining of the historical defect data, it can find the relationship between the cause of the defect and the defect location. It can identify the weakness of the equipment and be able to analyze the cause of the weakness. It has a certain significance for the key maintenance and overhaul of the equipment.

In this paper, combining with the local power dictionary is used to extract the feature quantity of the defect data. But it needs human participation to complete. In the face of a large database, the workload is large and the efficiency is not ideal, and it can reduce the ability of

defect analysis to a certain extent. In the next step, we will focus on solving the efficiency problem of massive data processing, use machine learning method to realize automatic feature extraction, and reduce workload to achieve the purpose of improving efficiency.

## References

1. Song, Y.Q., Zhou, G.L., Zhu, Y.L. (2013) Present Status and Challenges of Big Data Processing in Smart Grid. *Power System Technology*, 37(04):927-935.
2. Zhang, Y.L., Hu, C.C., Huang, S., Feng, S.Q., Lin, G.Q. (2017) Apriori Algorithm Based Data Mining and Analysis Method for Secondary Device Defects. *Automation of Electric Power Systems*, 2017,41(19):147-151+163.
3. Zhao, H.Y., Cai, L.C., Li, X.J. (2011) Overview of Association Rules Apriori Mining Algorithm. *Journal of Sichuan University of Science & Engineering (Natural Science Edition)*, 2011, 24(01):66-70.
4. Li, Y.H., Wang, J.X., Wang, X.L. (2016) A Power System Network Loss Evaluation Method Based on Hybrid Clustering Analysis. *Automation of Electric Power Systems*, 2016,40(01):60-65.
5. GUO, Y.F., GAO, H.L., TIAN, J. (2016) Photovoltaic output modeling by introducing clustering analysis and its application in reliability evaluation. *Automation of Electric Power Systems*, 2016,40(23): 93-100.
6. Li, X., GONG, Q.W., YANG, Q.Y., et al. (2011) Fault information management and analysis system based on data mining technology for relay protection devices. *Electric Power Automation Equipment*, 2011,31(9): 88-91.
7. Y, H., QU, G., ZHUANG, W.J., et al. (2014) Discussion on condition monitoring contents of secondary equipment in power grid. *Automation of Electric Power Systems*, 2014, 38 (12): 100-106.