Investigating Correlations in Power Grid Weather Disasters Using an Advanced Apriori Algorithm

Dr.B.Ratnakanth[1], P.Sriramulu[2], Amaresh[3]

[1] Professor & HOD, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad
[2] Assistant Professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad
[3] Assistant Professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

Abstract: ***In recent years, due to the frequent occurrence of extreme meteorological disasters, the safe operation and production activities of power grid have been greatly threatened. If we can find out the relationship between different types of climate and power grid equipment failure, it can provide help for power grid disaster prevention and reduction. Therefore, this paper chooses to use the Apriori algorithm based on mapping and pre-pruning to analyze the relationship between meteorological disasters and power grid equipment faults. Compared with the classical Apriori algorithm, this algorithm reduces the number of scanning transaction databases and the amount of calculation, so as to improve the efficiency. The improved Apriori algorithm is used to find out the correlation between meteorological disasters and power grid equipment faults, so as to understand the influence scope and degree of meteorological disasters. Finally, the effectiveness of the proposed method is verified by experiments.***

Keywords: Apriori Algorithm, Power Grid, Meteorological Disasters, Mapping, Pre-pruning

1. Introduction

Although in recent years, with the improvement of power grid equipment level and operation and maintenance capacity, the stable operation of power grid has been greatly improved. However, since the 21st century, with the further intensification of the trend of global warming, extreme weather events such as typhoon, thunderstorm, rainstorm, heat wave, hail and haze have occurred frequently, which have posed new challenges to the stable operation of the power grid [1].

In July 2012, affected by the hot and humid weather, India was forced to withdraw from operation due to excessive equipment load, resulting in large-scale power failure, resulting in 600 million people unable to maintain their normal lives due to power failure [2]. On the afternoon of September 28, 2016, affected by a strong typhoon in southern Australia, wind turbines were disconnected on a large scale, resulting in a 50 hour power outage and heavy losses. In China, heavy losses are caused by meteorological disasters almost every year. Typhoon "Rainbow" in October 2015 and typhoon "moranti" in September 2016 successively attacked many eastern provinces, causing great harm to power grid equipment [3]. In August 2019, typhoon "lichima" hit Fujian, Zhejiang, Jiangsu, Shandong and other provinces in China, causing many places to be cut off from water and power, and causing damage to a large number of power equipment [4]. At the same time, according to the 2020 national annual report on power reliability issued by China's national energy administration, we can find that the proportion of equipment unplanned outage time caused by meteorological disasters ranks the top three among all reasons, and the proportion of user power outage caused by natural factors is the highest, reaching 34.33% [5].

This paper expects to analyze the relationship between meteorological disasters and power grid equipment faults through Apriori algorithm, so that according to the analysis results, when a certain disaster occurs, we can clearly know which equipment is prone to faults, and focus on the inspection according to the probability of faults. This can greatly improve the maintenance efficiency of the power grid and minimize the unplanned outage time of the power grid.

In short, by exploring the relationship between meteorological disasters and power grid equipment failures, we can intuitively judge the impact scope and degree of disastrous weather, and evaluate the disaster situation of important power facilities in the power grid, so as to provide decision-making basis for power grid staff. In this way, preventive measures can be taken in advance, response preparations can be made, emergency disposal time can be shortened, and accidental power failure can be reduced.

This paper is divided into five parts. The specific content and organizational structure are as follows:

- The first part is a brief introduction, which mainly introduces the research background, research purpose and the importance of power grid disaster prevention and reduction.
- The second part is the introduction of related work, which mainly introduces the research status.
- The third part is the introduction of key technologies. This part mainly introduces the algorithm used in this paper. This algorithm is an Apriori algorithm based on mapping and pre pruning. Compared with the classical Apriori algorithm, this algorithm greatly reduces the amount of calculation and improves the efficiency of the algorithm.
- The fourth part is the experiment, which verifies the

feasibility of this idea, including the accuracy of association rule analysis results and the effectiveness of the algorithm.

- The fifth section is the summary and outlook of this paper.

2. Related Work

With the development of science and technology, people's demand for electricity is also increasing day by day. Therefore, how to protect users' electricity has attracted more and more attention. As one of the main causes of power grid equipment failure, meteorological disasters have naturally attracted much attention. Nowadays, there are more and more researches on power grid meteorological disasters.

In the United States, some scholars estimate the maximum wind speed, the maximum and minimum temperature, the maximum and minimum precipitation, the thickness of ice and other values in a year based on the actual meteorological data monitored in history, and then correct the predicted values according to the actual meteorological conditions in the region, and compare the corrected values with the historical observed values, It provides a valuable reference for power grid planning and design [6]. Others assess the risk level of distribution network affected by hurricanes in a year by simulating the number, intensity, duration and other parameters of hurricanes in the northeast of the United States. In Japan, meteorological prevention and control is carried out by measuring and analyzing temperature, rainfall and ice thickness. A meteorological observatory is established on the line according to historical operation data, and the monitoring data is compared with the data released by the National Meteorological Administration. If the two are consistent, they are adopted, and if the difference is large, they are directly discarded [7]. According to the local geographical conditions, Italian scholars have developed a prediction and early warning system, which is composed of meteorological data integration interface and GIS visualization system, which provides a basis for the stable operation of power system [8].

In China, people mainly focus on the analysis and research of a specific meteorological disaster or specific electrical equipment by using fuzzy reasoning, grey correlation theory and state enumeration method, etc. Jin Yan, Yu Zhen and others established a comprehensive micro meteorological early warning system through the study of power micro meteorological disasters [9]. By analyzing the temporal and spatial distribution of the impact of freezing and high temperature disasters on the power grid, Hollin obtained the distribution map of freezing and high temperature disasters in Hunan Province, which provides an effective basis for power grid disaster prevention and reduction [10]. Liu Song and Han Jian put forward the simulation method of continuous time distribution function of transmission line fault rate, using Fourier function, Gaussian function and Weibull function continuous time model assumptions, and fitting the parameters of the hypothesis function with the actual fault samples of power grids in Central China and southern coastal areas, so as to evaluate the risk of transmission line [11]. Some scholars have combined the K-means nearest neighbor algorithm with LS-SVM to predict the icing of transmission lines [12].

In addition to the above studies, there are many related studies, but we can find that most of the studies focus on the impact of a specific meteorological disaster on the power grid, while there are few studies on the comprehensive analysis of the relationship between meteorological disaster and power grid equipment failure. Therefore, this paper is no longer limited to the impact of a specific meteorological disaster, but uses Apriori algorithm to comprehensively analyze and study the relationship between meteorological disaster and power grid equipment fault. Thus, we can intuitively judge the impact scope and degree of disastrous weather, evaluate the disaster situation of important power facilities, improve work efficiency and reduce the unplanned outage time of power grid.

3. Algorithm Introduction

The algorithm used in this paper is to add mapping and pre pruning on the basis of the classical Apriori algorithm, so as to reduce the amount of calculation. The algorithm will be introduced in detail below.

3.1 Introduction to Classical Apriori Algorithm

As the earliest algorithm for mining association rules, Apriori algorithm was proposed in 1994. It is the most influential algorithm for mining frequent itemsets of Boolean association rules. Its core is the recursive algorithm based on the idea of two-stage frequency set. By mining the k-order frequent item set through the iterative search of the target transaction database layer by layer, until the highest order frequent item set is found. Finally, the association rules are mined through the obtained frequent item set, so as to achieve the ultimate goal of mining the association relationship between the target data. The main strategy of Apriori algorithm is to obtain all the frequent item sets in the target transaction database according to the preset minimum support, and then quickly obtain the association rules according to the frequent item sets. The detailed excavation steps are shown in Table 3.1.

Table 3.1: Apriori algorithm steps

①Given minimum support min_ sup.
②Traverse the target transaction library for the first time. Find the first-order frequent itemset L_1.
③L_{k-1} ($k \geq 2$) is connected by itself to generate k-order candidate itemset C_k.
④According to the fact that any subset of the frequent itemset is all frequent itemset, the k-order candidate itemset C_k can be pruned. Assuming that C_{k-1} is any (k-1) order subset of C_k, if $C_{k-1} \notin L_{k-1}$, then $C_k \notin L_k$, then the candidate item set is certainly not frequent, and the candidate item set can be deleted directly.
⑤Cycle steps 3 and 4 until a higher-order frequent item set cannot be obtained. Calculate the association rules that meet the requirements in all the obtained frequent item sets, and the mining process ends.

Through the above description, we can give the algorithm flow chart of classical Apriori algorithm, as shown in Figure 3.1.

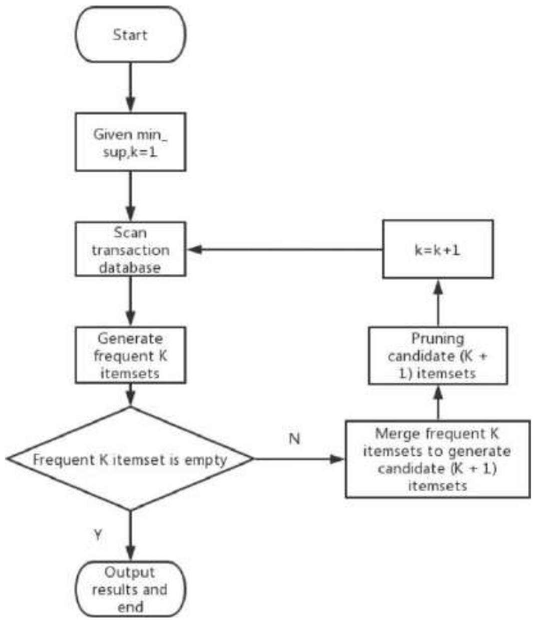

Figure 3.1: Flow chart of classical Apriori algorithm

We can find that the classical Apriori algorithm needs to scan the transaction database frequently, and will produce a large number of candidate item sets. This will increase the load of I/O system and make the amount of calculation become very huge, which will make the algorithm execution time become too long and inefficient. In view of these reasons, many improvement strategies have been proposed, such as improved methods based on transaction compression, vector group, hash table and matrix, etc [13]-[17]. In order to solve the disadvantage of large amount of calculation of classical Apriori algorithm, this paper chooses to improve the algorithm by using the data structure based on mapping, and pre prune the generated frequent itemsets, so as to reduce the workload of scanning transaction database and calculating frequent itemsets. In this paper, the algorithm is called MappingPre-purning_ Apriori algorithm (MP_Apriori algorithm for short).

3.2 Introduction to MP_Apriori algorithm

MP_ Apriori algorithm uses the data structure of mapping to save the target transaction database; so that each item in the candidate item set can be used as a key to quickly locate the array pointed to by the target. When it is necessary to calculate the frequency of candidate itemsets, we can quickly calculate the frequency of candidate itemsets by logical 'and' operation on the values of the corresponding positions of these target arrays, without querying the target transaction database D, and can significantly reduce the complexity of Apriori algorithm in calculating the frequency of candidate itemsets. Thus, the workload of the algorithm is greatly reduced and the efficiency is improved.

Let I={$I_1, I_2, I_3,\ldots, I_n$} be a set of all items, where I_k (k = 1, 2, 3,. . ., n) is called an item. A transaction T is an itemset, which is a subset of I, each transaction is associated with a unique identifier TID. Different transactions together form transaction set D, which constitutes the transaction database of the algorithm. According to the principle of mapping, this paper designs the storage structure of the target transaction database as shown in Figure 3.2.

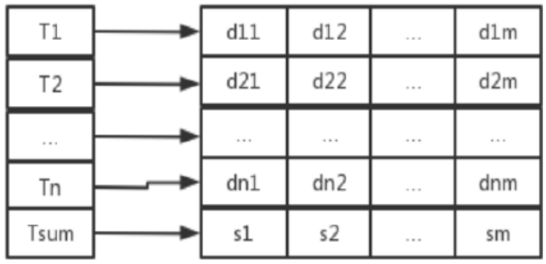

Figure 3.1: Storage structure diagram of target transaction Library

In Figure 3.2, T_k is a set composed of several I_k. The values corresponding to T_k and T_{sum} are one-dimensional arrays. When $I_i \in T_j$, $d_{ij} = 1$, otherwise $d_{ij} = 0$ (where $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$).$S_j$ is the number of itemsets contained in the transaction in the target transaction library, $s_j = \sum_{i=1}^{n} dij$.

For example, suppose there is an itemset I = {I_1, I_2, I_3, I_4}. The database contains four transactions, as shown in table 3.2. According to the above contents, the mapping storage structure of the database is shown in Figure 3.3.

Table 3.2: Transaction content of transaction Library

Tid	Itemset
T_1	I_1, I_4
T_2	I_2, I_3, I_4
T_3	I_2, I_4
T_4	I_2, I_3

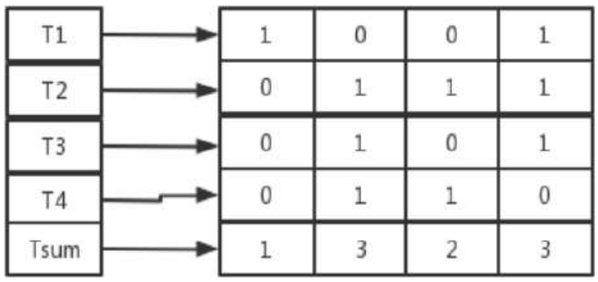

Figure 3.3: Target transaction Library

We know that when the candidate itemset C_k is obtained by self connection according to the frequent itemset L_{k-1}, it is necessary to ensure that the first (k-2) item must be the same and the last item must be different, which makes the calculation of this process particularly large. At the same time, after generating the frequent itemset C_k, according to the property that any subset of the frequent itemset must also be a frequent itemset, the amount of calculation when reducing the frequency of candidate sets is also considerable. Therefore, this paper chooses to pre prune L_{k-1} before obtaining the candidate itemset C_k, which can avoid a large number of comparisons and reduce the number of calculations as much as possible. The core idea of pre pruning is: the necessary condition that C_k generated by L_{k-1} self connection is a frequent itemset is that the frequency of the first I items of any itemset of L_{k-1} is at least k-i.

The specific algorithm implementation steps are shown in table 3.3.

Table 3.3: MP_Apriori algorithm steps

①Given the target transaction database D and minimum support min_ sup. ②The target transaction library is traversed for the first time to obtain L_1. At the same time, a set M that saves the item set and its occurrence times is obtained. ③Computing k-order frequent itemsets. First, pre prune L_{k-1} to obtain the result L'_{k-1}. Then the k-order candidate set C_k is generated. (The pre pruning steps are shown in table 3.4.) ④Calculate the frequency of C_k. Get the array of corresponding itemsets according to M, calculate the frequency through logical 'and', and compare it with the minimum support to judge whether it is a frequent itemset. ⑤Get all frequent itemsets.

Table 3.4: Pre pruning steps

①Given two cyclic variables x = 1, y = 1. ②Take out the x-th itemset X of L_{k-1}. The set X' composed of the first y items of X is used as the key to query the value from M. If yes, go to step 5, otherwise go to the next step. ③Initialize the cyclic variable r, obtain the r-th itemset of L_{k-1} respectively, and store the frequency of X' in M. ④Let the value of X' be p. If $p \geq k\text{-}y$, go to step 5, otherwise delete the r-th itemset starting with X' in L_{k-1}. ⑤x = x + 1, if $x \geq \mid L_{k-1} \mid$, the algorithm ends. Otherwise, make y = y + 1 and judge that y > k-1. If it is true, make y = 1 and skip to step 2. If it is false, skip to step 2 directly.

4. Experiment

The data used in this paper is the power grid maintenance report of some cities in Ningxia Hui Autonomous Region in recent ten years.

4.1 Running results of MP_Apriori algorithm

According to the description in Section 3, we know MP_ Apriori algorithm needs to set the minimum support first, and different values of the minimum support will lead to different results. In this paper, the support is set to 0.3, 0.2, 0.15, 0.1 and 0.05 respectively. Their accuracy is shown in the table 4.1.

Table 4.1: Accuracy under different support

Minimum support	0.05	0.1	0.15	0.2	0.3
Accuracy	94.7%	95.1%	93.8%	87.9%	79.3%

Convert the data in the table into a histogram. As shown in the figure 4.1.

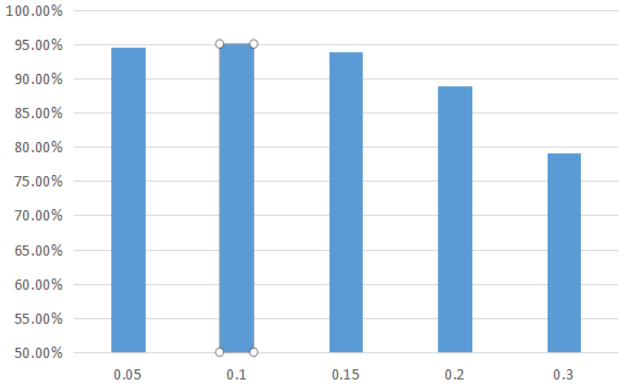

Figure 4.1: Accuracy under different support

Through figure 4.1, we can find that with the decrease of support, the accuracy of the algorithm will improve. When the support is positioned at about 0.1, the change of accuracy tends to be stable. This is because when the support is too large, some extreme weather that occurs less often will be filtered out, making the experimental results inaccurate. Therefore, the minimum support selected in this paper is 0.1. At this time, the accuracy is high, and the association rules discovered by the algorithm are more reliable.

When the minimum support is 0.1, the screenshot of some results of algorithm execution is shown in Figure 4.2.

Figure 4.2: Screenshot of experimental results

Figure 4.2 shows some of the results of the experiment. We can find that the support of rainstorm is 0.56 and that of thunder is 0.33. In case of rainstorm, pole switch, circuit breaker, lightning arrester, transformer and pole tower will be affected, and the probability of failure is 32%, 85%, 63%, 77% and 18% respectively. In case of thunder, the probability of phase fuse failure is 34%, the probability of circuit breaker failure is 83%, the probability of relay protection device failure is 69%, and the probability of simultaneous failure of relay protection device and circuit breaker is 71%. From the experimental results, we can find some unknown connections. For example, heavy rain sometimes causes damage to towers, and the probability of simultaneous failure of relay protection devices and circuit breakers in case of thunder is high. In this way, we can focus on these devices in case of rainstorm and thunder through association rules. Carry out safety inspection before extreme weather and repair in time after it occurs.

In a word, we can clearly know the impact of different meteorological disasters on power grid equipment according to the mined association rules, so that we can selectively and emphatically inspect and repair power grid equipment in case of extreme weather. Then improve the work efficiency of power grid staff and reduce the unplanned outage time of power grid.

4.2 Comparison between classical Apriori algorithm and MP_Apriori algorithm

According to the theory of the third part, it is obvious that compared with the classical Apriori algorithm, MP_ Apriori greatly reduces the workload of scanning transaction database. At the same time, pre pruning L_{k-1} can reduce many unnecessary work and improve the efficiency of the algorithm. In a word, compared with the original algorithm, MP_Apriori algorithm has a great improvement in performance.

This paper counts the execution time of Apriori algorithm and MP_Apriori algorithm under different support. The execution results are shown in table 4.2 and the comparison is shown in Figure 4.3.

Table 4.2: Execution schedule of two algorithms under different support (Unit: Second)

Algorithm	0.05	0.1	0.15	0.2	0.3
Apriori	20456.2	15974.5	8673.9	4945.2	1556.9
MP_Apriori	3210.1	2689.4	1803.3	1300.6	1087.1

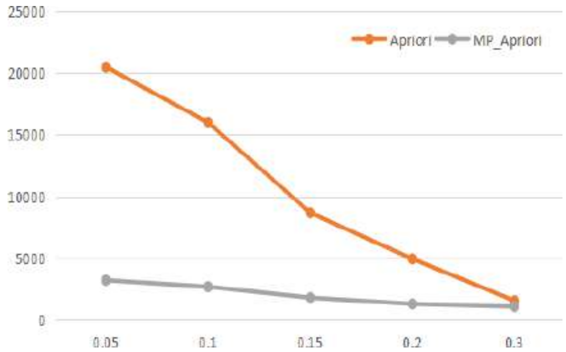

Figure 4.3: Comparison of execution time between two algorithms

Through figure 4.3, we can find that when the support is large, the execution time gap between the two algorithms is small, and when the support is getting smaller and smaller, the execution time gap between the two algorithms will become large. It can be seen that the efficiency of MP_Apriori is significantly higher than that of Apriori.

5. Conclusion

This paper proposes to use the association rule algorithm to mine the relationship between meteorological disasters and power grid equipment faults, so as to judge the impact scope and degree of disastrous weather and evaluate the disaster situation of specific facilities according to the contact assistant staff. This can reduce the troubleshooting time of maintenance personnel, improve work efficiency and reduce the unplanned outage time of power grid.

Compared with other studies, this paper is not limited to the impact of a specific weather on power grid equipment, but comprehensively analyzes and studies the relationship between meteorological disasters and power grid equipment to find out the universal law, which can be applied to any scene. At the same time, the Apriori algorithm based on mapping and pre pruning selected in this paper is more efficient and effective than the classical Apriori algorithm.

The future work of this paper is to study the setting of minimum support. From the experimental results, we can find that there will be a considerable gap between the experimental results obtained by different minimum support. Therefore, how to find the most appropriate minimum support is a very important research direction. At the same time, there is still much room to improve the efficiency of the algorithm. I will try to improve the efficiency of the algorithm.

References

[1] Li Xiaoyan, Ding Ming, Qi Xianjun. Reliability interval evaluation of transmission network considering weather factors and its affine algorithm [J]. Power system protection and control, 2016, 44 (16): 8-13.

[2] Zeng Ming, Li Na, Dong Jun, et al. Key technologies of power grid operation and management based on the concept of great security--Thoughts on the blackout in India [J]. Power system automation, 2012, 36 (16): 9-13.

[3] Chai Qingfa. Research on emergency material prediction method for power grid meteorological disasters [D]. Shandong University, 2020.

[4] Zhong Kaibin. Review and Prospect: Construction of China's emergency management system [J]. Political science research, 2009 (1): 77-88.

[5] National Energy Administration, China Electric Power Enterprise Federation.2020 national annual report on power reliability [R].2021.8.

[6] Tan P C, Loh P C, Holmes D G. Optimal impedance termination of 25-kV electrified railway systems for improved power quality [J]. IEEE Transactions on Power Delivery, 2005, 20 (2): 1703-1710.

[7] Ni M, Mccalley J D, Vittal V, et al. Software Implementation of Online Risk-Based Security Assessment [J]. IEEE Power Engineering Review, 2007, 22 (11): 59-59.

[8] Allan R N, Grigg C H, Al - Shakarchi M R G. Numerical techniques in probabilistic load flow problems [J]. International Journal for Numerical Methods in Engineering, 2010, 10 (4): 853-860.

[9] Jin Yan, Zhang Wei, Yu Zhen, Zhao Weimei. Research on power micro meteorological disaster monitoring and early warning technology [J]. Power information and communication technology, 2015, 13 (04): 11-15.

[10] Huo Lin, Tan Ping, Zhang Tingting, Li Caixia. Temporal and spatial distribution characteristics and impact analysis of power meteorological disasters [J]. Southern agriculture, 2017, 11 (20): 83-84.

[11] Liu Song, Han Jian, Zhong Lu, Zheng qiongling, Li Tao, Xiao Chun. Power grid risk assessment based on meteorological statistical failure rate of transmission line [A]. Jiangxi Society of electrical engineering. Proceedings of the 2018 annual meeting of Jiangxi Society of electrical engineering [C]: Jiangxi Institute of electrical engineering, 2019: 6.

[12] Huang Xiaoning, Xu Jiahao, Yang Chengshun, Wang Jiao, Xie Jiajia. Transmission line icing prediction based on data-driven algorithm and LS-SVM [J]. Power system automation, 2014, 38 (15): 81-86.
[13] Shao Wenshuang. Improvement of Apriori algorithm based on association rule mining [D]. Northeast University, 2015.
[14] Lin Jiaxiong, Huang Zhan. An improved Apriori algorithm based on array vectors [J].computer Applications and Software, 2005, 28 (5): 268-271.
[15] Zeng Lei. Research on Apriori algorithm in association rule mining [D]. Chongqing Jiaotong University, 2016.
[16] Yu Yanyan, Li Shaozi. Improved AprioriTid algorithm for association rules based on Hash [J].computer engineering, 2008, 34 (5): 60-62.
[17] Cheng Changpin, Wu Yilin, Jiang Yongsheng. Association Rule Mining Based on matrix Apriori improved algorithm [J]. Journal of Guangdong Second Normal University, 2019, 39 (05): 89-97.