Power System Data Application Based on Data Association Rules

Yuelin Zou1,*, Tao Wang1, Ayidos Jaynes1, Rong Ma1
1State Grid Xinjiang Information & Telecommunication Company, Urumqi, Xinjiang, 830000, China

*Corresponding author e-mail: zouyuelin@sgcc.com.cn

Abstract. With the continuous development and progress of society, electricity has been integrated into people's lives. However, power transmission is a very complex process that requires power transmission and power system conversion. A large amount of data will be generated during the operation of the power system. Through these data, we can use electricity better and more efficiently. This paper aims to study the power system data application based on data association rules. Based on the analysis of data association rules related algorithms and the application of data association algorithms in the power system, a power failure prediction system is designed and the performance of the system is analyzed. The test results show that the system has a very high transaction success rate, the longest response time does not exceed 20 seconds, and the CPU is operating normally, reaching the expected expectations.

Key words: Data Association Rules, Power System, Data Application, Fault Prediction

1. Introduction

Electricity, as an important energy pillar of our country, has always been the most important factor affecting our country’s economic and social development and people’s quality of life [1-2]. How to conduct safe, stable and reliable power production will determine whether it can play a key supporting role in the growing economic and social environment. Therefore, to ensure the safe and stable operation of the power system, there will be very huge economic and social benefits [3-4].

In recent years, with the development of big data and computer technology, the application of data association rules in power systems has attracted the attention of more and more researchers at home and abroad. Finding and extracting hidden relationships between database items or features is based on the features that appear at the same time of the data items. The corresponding discovery rules help individuals make decisions and operate the system. At present, the application of data mining based on association rules to power systems has achieved some research results [5-6]. Some researchers proposed an improved April algorithm based on association rule mining, analyzed the data characteristics of the three grid states, created a test database, improved the traditional April algorithm, and now has a support number, accept and improve the current support number [7-8]. Some researchers also apply the association rules of data mining technology to the relevant analysis of
power transmission and transformation device failures and weather factors, and use the Apriori algorithm to obtain a large amount of relevant historical data, obtain one-dimensional extreme climate events, and establish multi-dimensional association rules[9-10]. Some researchers use the main method of program data analysis to extract important indexes from massive data, and analyze the relevant four-dimensional index system that affects the reliability of distribution network operation, use autonomous program analysis to mine key evaluation indexes from massive data, and analyze relevant influencing factors; use the correlation rule model to dig out the main reliability index, and the strong correlation rules between each influencing factor to obtain the main influencing factors [11-12]. Finally, through artificial neural networks, these important influencing factors are used as predictable inputs, and the main index is used as output, so as to estimate the operational reliability index for a certain period in the future.

On the basis of consulting a large number of references related to "data association rules" and "power system", this paper combines data association rules related algorithms and the application of data association algorithms in power systems to design a power failure prediction system, which mainly includes four modules. The four modules are system management module, information viewing module, predictive function module and maintenance module. The performance of the system is tested after the above modules are implemented. The test results show that the system has reached the expected value of this article.

2. Power System Data Application Based on Data Association Rules

2.1. Data Association Rules

(1) Association rules
The connection rule is mainly used to find the implicit type in the data, such as \(X \Rightarrow Y\). There are many such implicit expressions in transaction data, and not every one of them is what scholars want. People filter some highly relevant items based on the number of items in the data set. Related products must meet specific requirements. One is to meet the minimum volume of the entire transaction. The minimum quantity is usually measured with minimum support. Second, the number of two highly correlated items that appear simultaneously in a transaction must reach a certain percentage in each transaction. This ratio is usually measured with the least support and the least confidence. The support degree indicates the frequency of items set across transactional databases. Assuming that the transaction database contains \(N\) pieces of data, the supporting formula is as follows:

\[
Sup(x) = \frac{Sup(x)}{N}
\]  

(1)

Confidence indicates how often a particular rule appears in the transaction database. For the confidence level, if \(X\) is included in the total transactions, the value represents the percentage of transactions that include \(Y\). Assuming there are \(N\) transactions in the transaction database, the formula for calculating trust is:

\[
Conf(X \Rightarrow Y) = \frac{Sup(X \cup Y)}{Sup(X)}
\]  

(2)

(2) Apriori algorithm
The Apriori algorithm is the most classic algorithm for mining connection rules. The algorithm is easy to understand at the beginning of the algorithm, and it is simple and convenient to implement the algorithm. The principle of the Apriori algorithm is that the more two elements displayed in pairs in the transaction data, the greater the correlation between the two elements, and the stronger the correlation between the two elements. The database must be scanned multiple times to implement the Apriori algorithm. The first scan counts the number of views for each item in the data, and deletes some items that do not meet the minimum support requirements. Before the second scan, the data
obtained from the first scan must be paired together. The second data scan then counts the number of combinations that appear, removes some combinations that do not meet the minimum support, and repeats the combination and scanning until a new combination is created and the Apriori algorithm is completed.

(3) FP-Growth algorithm
The Apriori algorithm needs to save the results every time it scans the database, and it also needs to scan the database multiple times. Scanning multiple data can be time-consuming, but if the database is large, this algorithm consumes a lot of memory. Therefore, people proposed another classic FP-Growth related algorithm. The algorithm mainly compresses and stores data by constructing an FP tree, and only needs to scan the database twice, which improves the extraction efficiency. The FP-Growth algorithm mining process scans the database to obtain support for each component. Items that do not meet the support level will be deleted and sorted according to the data support level. After that, build the FP tree and populate the items in the database until all items are populated. The general data set can be extracted from the FP tree. This completes the FP-Growth algorithm of the entire association rule.

2.2. The Specific Application of Data Association Rules in the Power System

(1) Visualized monitoring of grid operation
There are many devices and detection points in the entire transmission and power conversion network. Equipment operation indicators and power transmission and transformation voltage, current, and load monitoring indicators are collected in real time from each device through sensors and processed by multiple real-time processing platforms. The operation data of various equipment and monitoring points are extracted, processed, and integrated, and displayed in real-time on a large display screen to provide early warning and processing of abnormal equipment data and test points. At the same time, it organizes and analyzes the operating data and equipment maintenance data, uses big data processing and data mining technology to predict the equipment life cycle, proposes abnormal equipment maintenance processing methods, forms a knowledge base for predicting the potential equipment maintenance cycle, and Big data support for power equipment asset management, equipment operation testing management, equipment technology management, and technology management.

(2) Failure prediction
There is a system master station and several sub-stations in the entire network system. The main function of the sub-stations is to collect system field data, make preliminary analysis and processing of these data, and transmit the data analysis results to the system master station. The main function of the master-slave station is to receive different data signals from different sub-stations, analyze these data signals and perform data processing. Therefore, data preprocessing is an important link for the system to reduce and prevent a large amount of invalid or junk data information from flowing into the main site of the system, and to efficiently manage a large amount of invalid data information. Obtaining, exchanging, clearing and monitoring are the basic steps of data preprocessing. Data extraction and cleaning are mainly based on the different requirements of the work process. If all data extraction and processing of these processes are completed, the system will perform fault monitoring and prediction. Its main function is to monitor and predict any abnormal changes in power system data. Its main function is to monitor and predict any abnormal changes in power system data. If abnormal data is found, timely use technical means to deal with the big data error prediction characteristics.

(3) Energy efficiency analysis and demand response
The power grid management system collects a large amount of data from users. But at present, this part of data is still only used as shallow data statistics, and its detailed use value has been rarely studied. Using big data analysis technology, through the large statistics and analysis of user energy
consumption on the user side, it is reasonable and effective to guide the user's charging resources into the top cut and the bottom of the grid, thereby reducing the peak load pressure of the grid and effectively improving the utilization of the grid efficiency and economy. At the same time, through big data analysis and data mining of users' electric energy, it can analyze the load characteristics and actual power supply conditions of typical industries under typical power supply scenarios, deeply explore the influence mechanism and behavior characteristics of power supply activities, and establish a peripheral application profile system to improve service levels.

3. Experiment

3.1. System Management Module Design
The system management module also includes authorization management and application management, which refers to the authorization management and authorization of applications used by the system. The user login module is the system login, and the user must confirm the account and password when logging in. The system acts as a system administrator, protector, etc., invites users of different levels to browse different browsable web pages, and uses the same authority information given by the system as an identity. After the user is registered, the system can query and modify the user's personal information.

3.2. Information Viewing Function Module
After the administrator logs in, the administrator can check the status of the power lines in the local area under their management. After receiving the fault prediction information, they can check and troubleshoot the maintenance personnel closest to the fault location. Once connected, maintenance personnel can also view circuit information and failure prediction information.

3.3. Prediction Function Module
The prediction function is the main function of the system. The system collects current, voltage, and resistance information for transmission and power terminals. If failure information is found in the model results, the background will notify the administrator of the error information. In the most important error prediction function, the sensor sends the collected input and output information, current and resistance information to the model, receives the classification results, and stores the classification results in the server database. The database sends data through JAVA+JSP and displays it dynamically on the HTML page.

3.4. Maintenance Function Module
After the maintenance person receives the maintenance information notification from the administrator, he can choose to accept the maintenance and go to the troubleshooting location to solve the potential error. After the troubleshooting, click Finish. The administrator's point of failure indicates that the editing has been completed. The performance of the maintenance work mainly depends on the submission of the maintenance personnel. After requesting information on the customer page, the maintenance personnel can click the order to use the database. When the maintenance personnel complete the maintenance, the maintenance personnel send a complete maintenance key to update the database information. These two operations are used for data transmission through JAVA+JSP.

4. Discussion
LoadRunner is a widely used system control tool. Test system performance and understand system operation problems by simulating user interaction on the system. The test process can use a combination of manual and system to run tests.

Table 1. Transaction analysis of the first scene and the second scene

|                  | the first scene | the second scene |
|------------------|----------------|-----------------|
| Through affairs  | Failed affairs | Through affairs  | Failed affairs |
Table 1 and Figure 1 show the results of the transaction graph research in the two scenarios of the performance test. Among the 50 users in scenario 1, 7,321 transactions were successful and 5 failed. Among 100 users, 8652 transactions have been successful and 7 have failed. In scenario 2, 50 users succeeded 11284 transactions, 3 failed, 100 users, 12447 succeeded, and 5 failed. The failed transaction is due to the response timeout. The server timeout was previously set to 5 seconds, and after the current setting is 10 seconds, there are no failed transactions.

**Figure 1.** Transaction analysis of the first scene and the second scene
Figure 2. Response time of the first scene and the second scene

Figure 2 shows a simulation of concurrent transactions operated by 50 users. The longest and average response time reached 9.8s and 3.2s, respectively, and the CPU usage rate reached 60%. Simulating the concurrent transaction operation of 100 users, the longest and average response time were 13.5s and 6.7s, respectively, and the CPU usage rate reached 68%. From this result, it can be known that the system has a very high transaction success rate, the longest response time does not exceed 20 seconds, and the CPU is operating normally, meeting the expected expectations.

5. Conclusions

The operation of the power system directly affects the economic and social structure and the lives of individuals and groups. It has a close relationship with people and is naturally indispensable. The dynamic development of the operation of the power system, the expansion of the amount of database information, and the rational use can all enhance the stability of the power system to a certain extent. The data association rule technology has never been paid attention to until now it is indispensable. As mentioned above, people can pay more attention to specific applications, improve assumptions and make better use of data association rule technology. Make the business management process more optimized and make users more comfortable. However, we still need to continue to explore analysis methods, improve the technology of big data association rules, and contribute more to the national power industry.

References

[1] James, Sauers, Griffin, et al. Investigation of S/sub 2/F/sub 10/ production and mitigation in compressed SF/sub 6/-insulated power systems[J]. IEEE Electrical Insulation Magazine, 2018, 9(3):29-40.

[2] Dosoglu M K , Guvenc U , Duman S , et al. Symbiotic organisms search optimization algorithm for economic/emission dispatch problem in power systems[J]. Neural Computing and Applications, 2018, 29(3):721-737.
[3] Amini M, Orkun K. Optimal Operation of Interdependent Power Systems and Electrified Transportation Networks[J]. Energies, 2018, 11(1):196-.

[4] Setiadi H, Krismanto A U, Mithulananthan N, et al. Modal interaction of power systems with high penetration of renewable energy and BES systems[J]. International Journal of Electrical Power & Energy Systems, 2018, 97(APR.):385-395.

[5] Ziping W U, Gao W, Gao T, et al. State-of-the-art review on frequency response of wind power plants in power systems[J]. Journal of Modern Power Systems & Clean Energy, 2018, 6(1):1-16.

[6] Wang X, Blaabjerg F. Harmonic Stability in Power Electronic Based Power Systems: Concept, Modeling, and Analysis[J]. IEEE Transactions on Smart Grid, 2018, PP(99):1-1.

[7] Dickes R, Dumont O, Guillaume L, et al. Charge-sensitive modelling of organic Rankine cycle power systems for off-design performance simulation[J]. Applied Energy, 2018, 212(FEB.15):1262-1281.

[8] Meng W, Wang X, Fan B, et al. Adaptive Nonlinear Neural Control of Wide-area Power Systems[J]. IET Generation Transmission & Distribution, 2018, 11(18):4531-4536.

[9] Mondejar M E, Andreasen J G, Pierobon L, et al. A review of the use of organic Rankine cycle power systems for maritime applications[J]. Renewable and Sustainable Energy Reviews, 2018, 91(aug.):126-151.

[10] Qiao L, Gao W, Zhang H W, et al. Consensus-Based Distributed Economic Dispatch Control Method in Power Systems[J]. IEEE Transactions on Smart Grid, 2019, 10(1):941-954.

[11] Netto M, Mili L. Robust Data-Driven Koopman Kalman Filter for Power Systems Dynamic State Estimation[J]. IEEE Transactions on Power Systems, 2018, PP(6):1-1.

[12] Branco H, Castro R, Lopes A S. Battery energy storage systems as a way to integrate renewable energy in small isolated power systems[J]. Energy for Sustainable Development, 2018, 43(apr.):90-99.