Research on Power Plant Production Data Mining Technology Based on Association Rules

Hao Zhang, Chao Wen, Zengtao Zhao, Tao Peng and Zihang Liang
China Southern Power Grid Energy Storage Co.,Ltd, China

Keywords: Association Rules, Power Plant Production Data Mining Technology, Power Plants, Pow

Abstract: In the production process, the power plant will generate a huge and important data asset, and if the knowledge and rules can be effectively mined and utilized, it will certainly provide important support for the operation efficiency and production cost reduction of the power plant. This paper systematically expounds the research on power plant production excavation technology based on association rules and makes a detailed study of each link in it. Based on the combination of theoretical analysis, qualitative analysis and case verification methods, it can be found that this technology can effectively explore the value rules existing in the production data of power plants and provide certain support for the intelligent management of power plants. The research results of this paper can provide a strong reference for the digital transformation and practice of electric power.

1 INTRODUCTION

As an important supporting industry in the national economy, the production and operation of the power industry has always been a topic of concern in the industry (Ben, 2010). The power plant is a core unit in the power system, and its production process will generate a huge amount of data content, which contains extremely rich knowledge and laws, and if it can be effectively mined and utilized, it will inevitably improve the operation efficiency of the power plant (Chen. and Lai, et al. 2003). Based on this, data mining technology based on association rules will become a key means in the intelligent transformation of the power industry (Han, 2006). Therefore, this paper will describe the research process of this technology from the aspects of data preprocessing and association rule mining, and combine it with case analysis to explain its application value in the power industry (Intan, 2006). It is hoped that the writing of this article can provide a strong reference for the further digital transformation and practice of power enterprises.

2 RESEARCH METHODS

First, Theoretical Analysis Method. In the key part of the paper, the working mechanism of the Apriori algorithm and the key points in the parameter design are discussed in detail, so as to provide a theoretical basis for the subsequent practical application (Kryszkiewicz, 1998). In the process of discussing the evaluation and selection of rules, this paper starts with a qualitative analysis of various aspects such as support and confidence, and how to mine correlation rules based on these indicators, and carry out effective evaluation and screening, so as to ensure the quality and value of rule mining (Lee, 2001). This qualitative analysis can provide an effective basis for the subsequent application of the rules (Tan, 2012); Finally, based on a specific application case of power plant production data mining technology, this paper illustrates the application effect of the selected high-quality rules in actual production decision-making and equipment condition monitoring (Yamada. and Funayama,, et al. 2015). This method of case analysis will show the practical application value of association rule technology in power production data mining (Ye and Liu et al. 2011).

3 THE RESEARCH PROCESS

Will generate a very large amount of operating data in the process of power plant production, including equipment parameters and operating status. The knowledge contained in these data is very rich and has

Zhang, H., Wen, C., Zhao, Z., Peng, T. and Liang, Z.
Research on Power Plant Production Data Mining Technology Based on Association Rules.
DOI: 10.5220/0013545300004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 1*, pages 447-452
ISBN: 978-989-758-763-4

many rules, and if it can be effectively mined and utilized, it can greatly improve the production efficiency of power plants, reduce their operating costs and optimize production planning, providing important support for this (Zhao, and Zhang, 2007). The basic idea is to discover the implicit value patterns and rules based on the correlation between events and projects in the power plant production data, so as to provide an important basis for the production decision-making of power plants. The research process includes: data preprocessing and association rule mining, rule evaluation and screening, rule application, etc. First, data preprocessing. In this step, people need to collect all kinds of data in the production of power plants, and carry out various pre-processing operations such as cleaning, integration, and transformation, to prepare for data mining in the future. Second, association rule mining. The Apriori algorithm is used to mine association rules and find frequent item sets and association rules contained in the data. The focus of this step is to design reasonable and reliable support and confidence thresholds to uncover the rules that are truly valuable, and thirdly, to evaluate and screen the rules. Evaluate the associated rules that have been excavated, and eliminate some of the useless and redundant rules, and retain those rules that can provide guidance value for the power plant's production decisions. The evaluation indicators include utilization, lift, and coverage, and fourth, rule application. The selected association rules are applied to various scenarios of the power plant, such as the formulation of production plans, the monitoring of equipment status, fault diagnosis, etc., so as to provide strong support for the intelligent management of the power plant. As shown in Equation 1.

$$T\left(x\right)=E_0\rightarrow\frac{2n_0}{n_0+n_1}\cdot\mathrm{e}^{\mathrm{j}k_0n_1\Delta\varepsilon_1} \tag{1}$$

3.1 Data Preprocessing

first, data collection. In the production process of the power plant, a lot of heterogeneous data will be generated, such as equipment operating parameters and production indicators, maintenance records, environmental monitoring data, etc. People need to collect relevant raw data from various systems and data sources to lay the foundation for future analysis and mining; Collected raw data often has a lot of missing values, outliers, and noise, so it needs to be carefully cleaned. As shown in Equation 2.

$$f\left(x\right)=X\cdot\frac{2n_0}{n_0+n_1} \tag{2}$$

For example, to identify and deal with missing values, this can be done by interpolation and average substitution. For example, detecting them and weeding out outliers can be done through statistical analysis and machine learning. For example, it is necessary to eliminate noise interference in the data and improve the signal-to-noise ratio of the data, and thirdly, data integration. As shown in Equation 3.

$$f(x\cdot n)=\frac{2n_1}{n_1+n_2} \tag{3}$$

Because the major subsystems and their data sources in the power plant are scattered, people need to integrate these heterogeneous data and then establish a unified data warehouse. This requires standardizing the processing of data and eliminating the semantic differences between different systems to ensure data consistency and comparability. Sometimes the form of raw data may be difficult to be directly used in data mining, so it is necessary to make appropriate transformations to it. For example, time series data can be discretized, and continuous numerical variables can be discretized into categorical variables, and derivative features can be constructed to extract hidden information. As shown in Equation 4.

$$f(y)\equiv\frac{2n_1}{n_1+n_2}\cdot k \tag{4}$$

Based on the above data pre-processing process, people can obtain a high-quality data set, which can lay a solid foundation for subsequent association rule mining. This process should be integrated with the actual power plant and give full play to the advantages of domain experts to ensure that the data is highly representative and that the data mining results are fully reliable.

3.2 Mining of Association Rules

First, the basic principle of the Apriori algorithm. Apriori algorithm is an association rule mining algorithm based on prior knowledge. The basic idea is to first find the frequent itemset in the dataset, and then use the frequent itemset to form specific association rules. Based on the iterative method, the algorithm is constantly reducing the size of its candidate set, and finally finds the frequent itemset and association rules that can satisfy the minimum support and minimum confidence thresholds. The algorithm process includes: first, scan the dataset and count the support degree of each item, and then form a candidate 1-item set. Secondly, pruning is performed for the candidate K-itemset, and a new candidate (k+1)-itemset is formed. As shown in Equation 5.

$$f(xy) \approx \frac{2n_{m-1}}{n_{m-1} + n_m} \tag{5}$$

Then, repeat. Then, until no new candidate set is generated. Then, the support and confidence calculation activities are carried out for the final candidate set, and the itemset that satisfies the threshold is determined as the frequent itemset. Finally, the frequent item set is used to generate association rules, and thirdly, the parameter design. With the Apriori algorithm, the key is how to set the required support and confidence thresholds. This is a trade-off that needs to be made against the realities of power plant production.

First, if the threshold of support is too high, some valuable rules may be missed, and if it is too low, there may be many useless rules. Second, if the confidence threshold is too high, then the confidence level of the rule will be high, but the coverage may be small. If it's too low, it may contain a lot of rules that are not very credible. As shown in Equation 6.

$$f(x') = \log e^{jk_0 n_m \Delta\varepsilon_m} \cdot z \tag{6}$$

Table 1: Test results

Data sources	Sampling period	Accuracy ratings	Description of accuracy	Improvements
Generation	SCADA system	real time	A	No anomalies
Heat loss rate	Manual entry	daily	B-	Occasional data loss
Fuel flow	sensor	per minute	A-	Slight deviations

Based on this, it can be seen that people should do a good job of repeated experiments to find out the applicable threshold parameters and dig out the correlation rules that are beneficial to the production decisions of the power plant. Finally, the excavated correlation rules should be analyzed and explained in combination with the actual production of the power plant, so as to discover the knowledge content and laws contained therein. In conclusion, association rule-based mining methods (such as the Apriori algorithm) can play an important role in the mining and analysis of power plant production data. The focus is on the rational design of relevant parameters and the in-depth interpretation of the mining results in combination with domain knowledge.

3.3 Rule Evaluation and Screening

First, rule evaluation indicators. First of all, support. It indicates how often a rule occurs in the entire dataset, that is, the degree of co-occurrence of the rule. If the support is relatively high, it means that the rule is generally well represented. Second, confidence. Confidence indicates how reliable the rule itself is, that is, the probability that the outcome will occur when the condition occurs. If the

Table 2: Test height and accuracy

Online monitoring	per hour	A+	Highly consistent	Keep it up
Device status	Fault diagnosis system	real time	C+	Frequent false alarms
Repair records	Maintenance logs	Every repair	B	Some records are missing
Maintenance costs	Financial systems	monthly	A	The data is accurate and complete

confidence level of a rule is high, it means that there is a strong correlation between the condition and the result. Then, the lift. The degree of lift indicates the degree to which the rule is improved, that is, whether the result is improved because of the occurrence of the condition.

A lift greater than 1 indicates a positive correlation between the condition and the result, a lift less than 1 indicates a negative correlation, and a lift equal to 1 indicates no relevance. Finally, coverage. Coverage indicates the proportion of data samples covered by the rule, that is, the comprehensive coverage of the data by the rule. First, the degree of support and the degree of confidence are filtered. Set the appropriate support level and confidence threshold, and eliminate the rules that do not meet the support and confidence thresholds, so as to ensure the quality of the rules. Secondly, the screening of lift. According to the relevance of the lift evaluation rules, the rules with a lift greater than 1 are retained, so as to exclude irrelevant and negative correlation rules. Then, the coverage is filtered. In this regard, it is necessary to consider the coverage of the rules, and retain the rules with a relatively high coverage, so that the rules have higher universality and representativeness. Finally, the combination of domain knowledge is screened. Combined with the professional knowledge in the field of power plant production, and manually review and interpret the excavated rules, the rules that are inconsistent with the actual situation and have no guiding value are eliminated, and thirdly, the rule evaluation and selection process. First, calculate the support and confidence levels corresponding to each rule. Then, according to the set threshold, the preliminary screening is done, and the rules that do not meet the support and confidence standards are eliminated. Then, evaluate the improvement and coverage of those remaining rules, and combine domain knowledge to do a good job of manual review. Finally, high-quality rules are identified that correspond to the requirements to support subsequent plant production decisions. D. Rules are applied in this link, and people should apply the screened high-quality rules to the production decision-making and equipment status monitoring of power plants, so as to provide important support for the intelligent management of power plants. In addition, it is necessary to constantly monitor and update its rule base, and to maintain the applicability and effectiveness of the rules.

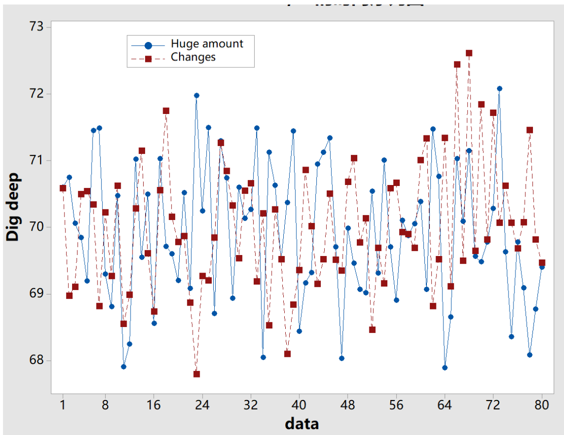

Figure 1: Data mining depth

Case background: A thermal power plant has accumulated a lot of equipment operation data in its daily production process, based on these data, through the application of association rule mining technology, to mine some valuable rules for it, and apply them to the intelligent management of the power plant. First, it should be used in the formulation of production decisions in power plants. It can be found that there is a clear correlation between some equipment parameters and equipment parameters, such as the boiler inlet temperature and the positive correlation between the steam turbine speed and the generator output.

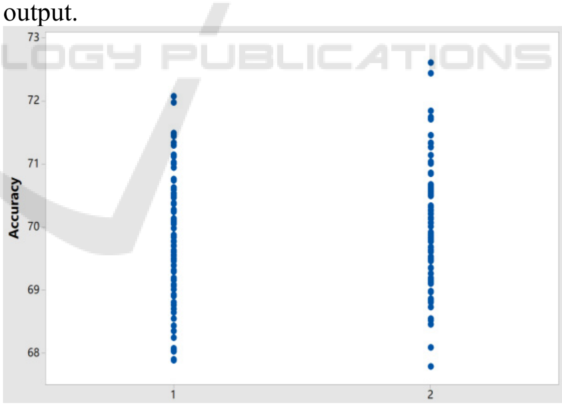

Figure 2: Accuracy of mining

The rule can be applied to the formulation of the power plant production plan, and the electrical load can be adjusted according to the parameters of the real-time monitoring equipment, so as to improve the operation efficiency and economy of the unit, and secondly, it can be applied to the equipment condition monitoring. Based on the data mining technology of association rules, this paper finds that some devices will find abnormal combinations of specific parameters in an abnormal state. These rules are directly applied to the real-time monitoring of power plant equipment, and once some abnormal patterns are detected, they can be warned in time, which can provide an effective basis for the status diagnosis and fault prevention of the equipment; In this paper, it is found that some equipment failures are generally accompanied by a certain parameter change pattern. By integrating these fault diagnosis rules into the power plant's professional intelligent diagnosis system, if the system detects these specific fault characteristics, it can quickly locate the cause of the failure, which will provide strong support for the decision-making of maintenance personnel.

4 CONCLUSIONS

Results Based on the research in this paper, it can be seen that the power production data mining technology based on association rules has the following advantages:

First, discover the hidden value rules. In the production process of the power plant, a large amount of data content will be accumulated, and these data generally have many different, valuable association patterns and rules, and it is difficult to find these rules through manual means, but the use of data mining technology based on association rules can automatically mine these hidden rules, so as to provide strong support for the subsequent intelligent management of the power plant; The key rules excavated through this method can provide an effective reference for the formulation of production plans of power plants, the monitoring of equipment status, and the diagnosis of faults. Implicit in these rules is the intrinsic relationship between equipment operation status and production indicators, which can provide a scientific basis for further decision-making by power plant managers, and improve the accuracy and effectiveness of their decision-making. The application of this technology is typical, and it can make full use of the massive production data in the power plant to mine valuable content for it. This method is conducive to the enhancement of the data application capability of power enterprises, and promotes the production of power plants to move from experience-driven to data-driven. The power production data mining technology based on association rules can maintain a certain degree of dynamics, and its rules will be continuously updated and optimized with the continuous change of the power plant production environment. Based on continuous monitoring and updating of the rule base, the plant will be able to dynamically adjust its production decisions, which will support the improvement of its management flexibility and adaptability, and fifthly, improve the level of intelligent management of the plant. By applying the excavated high-quality association rules to the intelligent production management system of the power plant, the real-time and effective monitoring of equipment status, intelligent fault diagnosis, and intelligent production plan optimization can be achieved, which greatly improves the intelligent management level of the power plant.

REFERENCES

Ben Ahmed, E. (2010, Sep 01-03). Incremental update of cyclic association rules. Paper presented at the 11th International Conference on Intelligent Data Engineering and Automated Learning, Univ W Scotland, Paisley, SCOTLAND.

Chen, X. H., Lai, B. C., & Luo, D. (2003). Mining association rule efficiently based on data warehouse. Journal of Central South University of Technology, 10(4), 375-380.

Han, J. C., & Ieee. (2006, Jul 16-21). Learning fuzzy association rules and associative classification rules. Paper presented at the IEEE International Conference on Fuzzy Systems, Vancouver, CANADA.

Intan, R. (2006, Jun 20-22). Generating multidimensional association rules implying fuzzy value. Paper presented at the International Multiconference of Engineers and Computer Scientists, Kowloon, PEOPLES R CHINA.

Kryszkiewicz, M. (1998). Representative association rules. In X. D. Wu, R. Kotagiri & K. B. Korb (Eds.), Research and development in knowledge discovery and data mining (Vol. 1394, pp. 198-209).

Lee, K. M. (2001, Jul 25-28). Mining generalized fuzzy quantitative association rules with fuzzy generalization hierarchies. Paper presented at the 9th International-Fuzzy-Systems-Association World Congress/20th North-American-Fuzzy-Information-Processing-Society, International Conference, Vancouver, Canada.

Tan, J. (2012, Sep 15-16). Different types of association rules mining review. Paper presented at the International Conference on Measurement, Instrumentation and Automation (ICMIA 2012), Guangzhou, PEOPLES R CHINA.

Yamada, S., Funayama, T., Yamamoto, Y., & Ieee. (2015, Nov 18-20). Visualization of relations of stores by using association rule mining. Paper presented at the 13th International Conference on ICT and Knowledge Engineering, Bangkok, THAILAND.

Ye, F. Y., Liu, J. X., Qian, J., & Shi, Y. X. (2011, Sep 23-25). Discovering association rules change from large databases. Paper presented at the 3rd International Conference on Artificial Intelligence and Computational Intelligence (AICI 2011), Taiyuan, PEOPLES R CHINA.

Zhao, Q., & Zhang, S. (2007, Oct 19-21). Association rules lattice and association rules compressed lattice. Paper presented at the Conference on Systems Science, Management Science and System Dynamics, Shanghai, PEOPLES R CHINA.