

Application Analysis of Data Mining in Power Enterprise

Tang Jian*

Management Department, the Engineering & Technical College of Chengdu University of
Technology, Leshan City, Sichuan Province, China
email:46851059@qq.com

Keywords: power enterprise, data mining, big data

Abstract. With the progress of the society and the development of information and communication technology, information system is rapidly expanding in various industries and fields. There are more and more data that these systems collect, process, and accumulate, and the growth rate of data volume is larger and larger, and even using "mass, explosive growth" and other words cannot describe the growth rate of data. Data mining is the technology that excavates information with potential value from vast amounts of data. This information is likely to have potential value, to support decision making, so as to bring benefit for the enterprise, or seek the breach for scientific research. Using the data mining technology with powerful functions can make manufacturing power companies convert data into useful information to help decision-making, so as to gain dominant position in the market competition. The paper briefly introduces the concept and methods of data mining, and analyzes the application and the prospect of data mining in the power enterprise.

1. Introduction

In the past, it took product as oriented, and reducing the production cost determined the power enterprise's survival and development. Now, if it still takes this concept to maintain enterprise development, the result is certainly dim. Through acquisition, excavation and analysis of the huge amounts of data, power enterprises can be more economic to get more value from diverse sources. In power industry, strengthening the rapid development of the smart grid makes the information communication technology have fast convergence with power production, enterprise management in an unprecedented width and depth; information communication system has become a "nervous centralis" of the smart grid, and support a new generation of power grid production and management development. At present, the state grid corporation has set up the leading domestic and international first-class information integration platform. With the operation of centralized data centers in succession, the deployment of the scope expanding of application business, and running of structured and unstructured data centers, power grid business data have been basically formulated in terms of the total amount and species. With the gradual popularization of the subsequent smart meters, grid business data will further enrich and expand from the timeliness. The features of big data, "amount, type, time", are further outstanding in terms of classes and properties, and the power big data analysis is imminent.

At present, the grid business data is roughly divided into three categories: the first is the production data of power enterprise, such as the date of amount of power generation, voltage stability, etc; the second is the operational data of power enterprise, such as data of trading electricity price, power sale quantity, electricity customer; the third is management data of the power enterprise, such as data of ERP, integration platform, collaborative office, etc. If these data based on the actual grid situation can be fully utilized with the thorough analysis, a large number of high value-added services can provided. The value-added services will be good for grid security detection and control (including disaster early warning and processing, power supply and power consumption decision support and a more accurate power consumption prediction), customer behavior analysis and customer segmentation, power enterprise refinement operation and management, etc., to realize more scientific demand management.

2. Main methods of data mining of power enterprise

Data mining integrates database, artificial intelligence, machine learning, statistics, and other fields of theory and technology. More typical data mining methods include classification, association rules analysis, clustering analysis and isolated-point analysis, etc.

2.1 Classification.

Classification is concept description to find a category, and it represents the whole information of this kind of data, namely, the connotation description of this category; generally it is represented by rules or biomedical decision tree. Actually, based on the analysis of the sample data in the database, an accurate description is made for each category analysis or analysis model is set up or classification rules are mined, and then the classification rules are used to classify other records in the database. According to the analysis object's properties and characteristics, it establishes a different set of classes to describe things. For example, power enterprises divide customers into different categories according to the previous data, and now they can distinguish new customers who apply for a loan according to these, to take corresponding loan scheme.

2.2 Isolated-point analysis.

Data in the database have some abnormal records, and these records are called isolated point, which often includes a lot of potential knowledge, such as abnormal case in classification, special case that does not meet the rules, deviation of observation result and model prediction, value changes over time, etc. Isolated-point analysis' basic method is to look for difference between the observation result and the reference. Isolated-point detection simply is the description of a few and extremely special cases of analysis object to reveal the internal reasons. For example: in 1 million's transactions in a bank, there are 500 fraud cases, in order for the prudent management, the bank needs to find the intrinsic factors of 500 cases, to reduce the risk of operation later.

2.3 Cluster analysis.

Cluster analysis' object is a set of unclassified records, and it is unknown that these records shall be divided into several categories in advance. Clustering is the reasonable partition record collection based on the analysis of log data in the database, according to certain classification rules, to determine the type of each record. The classification rule that it uses is decided by the clustering analysis tools. Different clustering methods may have different results for the same record collection. Clustering analysis simply means to identify internal rules, and divide the objects into several categories according to these rules. For example: the applicant can be divided into high risk applicants, moderate risk applicants, and low risk applicants.

2.4 Association rules reveal the internal relations between data, and find the relation between users with the site visit of all pages.

Its description form of data mining is: set $I = \{i_1, i_2, \dots, i_m\}$ as the mining object's data set, there is an event T , if a subset of X in I has the situation that X is contained in T , I and T have association rules. For example: "among customers with the purchase of bread and butter, 90% of people also bought milk at the same time", which is expressed as: (bread, butter, (milk)). Association rule technology is used to express customers visited pages/company/product1, 90% people will visit pages/company/product2. Through constructing the correlation model, data mining on the WEB is conducted; we can better organize sites, and reduce the burden of users in filtering information. Association rule is a connection of others things when something happens. For example: people who purchase beer every day is also likely to buy cigarettes, and the proportion can be described by the associated support and credibility.

3. Necessity and feasibility of using data mining

3.1 The informatization situation of power enterprise in China makes data mining technology possible.

By Observation of the current situation of radio and television companies, it can be known that grid information has had very big progress, that is, it no longer just completes statistics with the help of computer, and management information is not a single application period, it is in the intermediate link of the informationization development, the enterprise has its own local area network (LAN), radio and television group has realized the optical fiber network covered in the whole province, to make the internal management work more efficient, such as MIS, OA, material management, financial management, and customer service centers, etc. It can gain many basic data of enterprises, and make the application platform more scientific, and when the enterprise carries on the data mining, it can take the numerous data as an effective basis.

3.2 The reformation trend of power enterprises in China makes data mining technology necessary.

Our party and government leadership group is positively changing the present situation of the industry monopoly, to promote the rationalization of competition way. In the power enterprises in our country, they has been in use of the factory network separation mode, which makes the power generation competition with scientific mode, and radio and television group also has finished this part of the work. The following step is the shift in the direction of power grid operation. For keeping advantages in the fierce competition in the future, the power enterprises must reduce the cost of production and operation as far as possible, which is conducive to better provide service for the customer, and be familiar with themselves and the actual situation of competitive enterprises. All these things need to use modern information technology to solve, and data mining technology plays an extremely important role.

4. Application of data mining in power enterprises

Business demand for data mining technology is larger, so the data mining has got a wide range of application in multiple business fields. Below, based on the characteristics of power enterprises, it will discuss the important function of the data mining technology in the power enterprise.

4.1 Guiding equipment upgrade and construction planning.

If there are two situations in the following, equipment needs to update: first of all, the power facilities suffer accidental damage, and this change will be taken timely, general electric power equipment monitoring facilities can detect this kind of failure, which can also be for maintenance in the first time. Secondly, it shall replace the aging equipment, based on experience, such as inspection the using period of equipment, etc., but this way does not have much scientific meaning, because a lot of equipment may continue to be used due to proper care, if it is changed, it will produce a great waste; also, some equipment may not be used long, but its performance has not meet the standard, if not timely replacing it, it will also have a huge waste. In general, we use a malfunction, power consuming and related electric power parameters and all kinds of data to determine the power equipment failure and aging condition, eventually determine whether to replace equipment. In addition, in many parts of our country, there are frequent power shortage situations, and the main reason is that they do not well grasp market progress trends, and have failed to meet the demand of the market in the construction of power plants and power grid, then the importance of data mining work is well reflected. The related materials of new added user, existing users, user's location, power consumption, national construction plan are seriously researched with analysis, which will be able to work out a plan for the development of the power enterprise, with this as a guide to promote the rapid development of power industry.

4.2 Reducing power consumption, improving power quality and reducing equipment consumption.

The electric power products have their own obvious characteristics, mainly reflected in the fact that it cannot be stored and only supplied on demand. However, power generation and power consumption have a big difference, in order to guarantee the quality of electric power, we must constantly improve the security of facilities, with the implementation of the scientific adjustment. Now, the main way is the construction of storage power plant, if there is excessive power, it shall be stored, when the power supply can not meet the demand, this part of the power can be used, with the arrangement of the scheduling and formulation of reasonable disorders, which can realize the flexible adjustment of power storage technology, to reduce the power waste, improve the quality of electric power, and avoid the loss of equipment.

4.3 Performance evaluation.

The power enterprises in China do not have a standard set of system to evaluate the performance of group and company. If it only evaluates its economic profit, it will have error due to the regional development, and electric power industry is related to our production life, the importance of safety and other performance is more significant than profits. But data mining technology can make the comprehensive analysis of many influencing factors, through the analysis of profit, profit growth rate, peer comparison and complaint report, production cost and data warehouse to study area or the company's operation, and use charts and other simple way to provide the basis for decision-making.

5. Summary

As the development of the power enterprises has tended to become stronger, the data mining faced by power enterprise will be a very promising field. It can guide the equipment update, guide the construction of power enterprise planning, guide power production and purchase, reduce the power loss, improve power quality, reduce loss of equipment and bring huge profits, as well as create new business growth point for the power enterprise. But there are a lot of problems in data mining faced by the power enterprise that need solving, with the development of the power enterprise and the complexity of the business, the type of data will be more and more, more and more complex, data mining will play more and more important role.

References

- [1] Ruan Wenfeng. Analysis of key performance indexes of power supply enterprises based on data mining. *Guangdong Electric Power*. 2008(11).
- [2] Niu Chenglin. Theoretical research and application of incremental data mining and its power station operation optimization. *North China Electric Power University (Beijing)*. 2010.
- [3] Chen Xingying, Zhao Xiaohua. Application of data mining in power system. *Journal of electric power science and technology*. 2007 (3).
- [4] Guo Xiaoling. Analysis of Problems and Solutions of Statistics in Power Enterprise in Information Age. *Value Engineering*. 2010 (36).