Typical Case Study and Enlightenment of Data Mining at Home and Abroad

Guang Chen[1, a *], SuWei Liu[1,b], Wei Liu [1,c],Di Wang[1,d], YuKun Song[1,e] and YuBing Kang[1,f]

[1]State Grid Energy Research Institute Co. Ltd., Beijing 102209, China

[a]chenguang@sgeri.sgcc.com.cn, [b]liusuwei@sgeri.sgcc.com.cn, [c]liuwei@sgeri.sgcc.com.cn, [d]wangdi1@sgeri.sgcc.com.cn, [e]songyukun@sgeri.sgcc.com.cn, [f]kangyubing@sgeri.sgcc.com.cn

[*]corresponding author

Keywords: Electric Power System; Jarvis; Balanced Scorecard

Abstract: In this paper, the typical cases of data mining at home and abroad are studied in depth, and three aspects are mainly studied: first, the application of data mining technology in foreign power systems; second, Baidu Jarvis data analysis and mining platform; third, ExxonMobil Balanced Scorecard and strategic map examples. Through the study of typical cases, this paper draws some enlightenment from the application of data mining in enterprises.

1. Application of Data Mining Technology in Foreign Power Systems

In recent years, foreign academia and enterprises have carried out in-depth research and practical application of data mining technology in power system. The application fields mainly focus on dynamic security assessment, load forecasting, fault diagnosis, engineering cost, marketing support and so on.

1.1. Dynamic Security Assessment of Power System

K. R. Niazi and other scholars proposed the application of hybrid method based on artificial neural network and decision tree in online security assessment of power system. They proposed how to test the dynamic security of each generation dispatch by using decision tree technology in online dynamic security prevention of power system, and provided correct guidance through generation redistribution. Optimizing the power generation capacity and saving the cost of power generation.

1.2. Power System Load Forecasting

Electric load forecasting is a very important work of electric power dispatching, which is related to the operation plan of each unit in power plant. Some foreign scholars have designed a data mining model in the form of decision tree and applied it to daily load forecasting combined with regional power grid meteorological load database. In addition, some scholars use the time series model to forecast the load using the existing data series, and use the adaptive decision tree to cluster the relevant attributes such as user power consumption records, season, climate stored in the database, so as to formulate the appropriate tariff table.

1.3. Fault Diagnosis of Power System

Fault diagnosis of power system is to identify the location, type and misoperation device of fault elements by using information about power system and its protection devices, and the identification of fault elements is the key problem. The fault diagnosis model based on time series can be used in the field of fault diagnosis of high voltage transmission line system. According to the correlation of fault time series in time and space characteristics, the relativity and knowledge between events are mined by the principle of optimizing similarity of dynamic programming algorithm, and the solution of diagnosis problem is transformed into seeking and real-time fault. The standard fault sequence pattern with the most similar fault sequence pattern and the least operation cost can correct the abnormal events in the abnormal event sequence pattern and ensure the high fault

tolerance of the fault diagnosis system. In addition, some scholars have designed the algorithm of fault diagnosis decision-making based on rough set theory, so as to establish a fault comprehensive knowledge base for power grid fault diagnosis.

4. Project Cost

Engineering cost is a complex process of multi-variable and non-linear. Some foreign scholars use the historical cost data of the built projects, combine data mining with artificial neural network, and construct the cost model of the project. By using this model in the cost prediction and budget review of the power grid project, the subjective and one-sided effects of the existing manual review are effectively overcome, and the accuracy of the cost prediction is improved. Sex.

5. Marketing Support

Based on the actual data of electric power marketing, foreign electric power enterprises mining association rules and clustering from the aspects of quantity (electricity), price (electricity price), cost (electricity fee), loss (line loss). For example, according to the attributes of customers, clustering and grouping electricity customers, developing customer credit evaluation, preventing malicious arrears, detecting abnormal electricity use and preventing electricity theft.

2. Baidu Jarvis Big Data Analysis and Mining Platform

In the era of big data, enterprises not only want to improve their work efficiency through big data analysis and mining technology, but also are often hindered by the "threshold" brought by machine learning and other related technologies, so they urgently need a data analysis and mining product to cross this "gap". In this context, Baidu has developed a product, Jarvis, which can provide a package of data analysis and mining solutions for enterprises. Jarvis is a tool and platform to support the application development of big data analysis and mining. It is positioned between enterprise developers and big data analysis and mining technology. It can provide visual interaction support for enterprises, so that big data analysis and data mining technology can quickly be transformed into specific products to meet the application scenarios of enterprises.

Jarvis uses the principle of vertical hierarchy and horizontal hierarchy to design technology stack, which ensures that the whole process can solve data processing, computing resources, operator algorithm support, environmental deployment and other issues in the process of data analysis and mining. At the same time, Jarvis classifies functions for users and maintains extensibility to the greatest extent. It locates itself in data analysis products that can be used by data scientists, business developers, data analysts, product managers, decision analysts and other users at the same time.

Jarvis uses visual management to control the whole process of data analysis. In the link of data connection, Jarvis can support structured, unstructured and other types of data access, and support reading and flexible mounting of private data, Bos, distributed HDFS, relational database and other types of data sources. In data preparation, Jarvis can provide cleaning and preprocessing tools to support text, image and other types of data, which is convenient for efficient data preparation. In data analysis, Jarvis uses Spark to process data, supports PB-level SQL interactive query and analysis, and provides rich visual data exploration tools to facilitate developers to obtain high-value and effective samples. In the process of mining and modeling, Jarvis has built-in abundant basic operator algorithms for developers to model and develop efficiently. At the same time, it presupposes classical pituitary industry solutions, which can be implemented efficiently in matching scenarios at low cost. In the process of model deployment, the model generated by Jarvis can be released and deployed directly to the outside world to support free expansion. In process monitoring, Jarvis can monitor the whole workflow of data analysis, and new data can trigger automatic retraining process.

Jarvis uses cloud native service architecture to manage platform development resources. In the whole process and service of data analysis and mining, different scenarios, different data, different processing stages, different developers' demands for environment and resources are various. This requires that the resource management of data analysis and mining platform (including development

environment resources) can be flexibly accessed and flexible. Extension and expansion are convenient to ensure stability and efficient utilization of resources. To this end, Jarvis uses cloud native service architecture to achieve the above goals.

Jarvis has built-in generic algorithms, models and industry solutions. Different enterprises in the same industry often have common data analysis and mining scenarios, such as power consumption forecasting, equipment fault detection, fault prediction and so on. These similar scenarios solve similar problems and analyze similar data. Jarvis has built-in a large number of general basic algorithms, general models, vertical solutions, etc., which greatly improves the efficiency of data analysts.

3. Examples of ExxonMobil Balanced Scorecard and Strategic Map

ExxonMobil is the world's leading oil and petrochemical company, ranking ninth in the Fortune 500 rankings in 2019. However, in 1992, Mobil was a very weak company, with profitability at the end of the industry, serious bureaucratization of the management system and low organizational efficiency.

In order to resist the fierce competition from outside, Mobil introduced two management tools, Balanced Scorecard and Strategic Map, to implement a new enterprise development strategy in 1994, and carried out a series of organizational changes. By 1995, Mobil's profitability had risen to the top of the industry. In 1999, Mobil and Exxon merged into Exxon Mobil. Since then, ExxonMobil has maintained its competitive advantage in the highly competitive industry.

Balanced Scorecard and strategic map are the core of Mobil's strategic management. By using the Balanced Scorecard and strategic map, Mobil has created a performance-oriented culture, which effectively improves management efficiency, reduces costs, improves the productivity of the whole value chain, increases the marketing volume of high-priced and high-quality products and services, and improves the return on capital of enterprises.

Mobil's Balanced Scorecard is divided into four levels, including nine strategic themes, 17 strategic objectives and 26 measurement indicators. Its greatest feature is that it is designed concisely, accurately, concisely and easily understandable, easy to communicate, and can quickly and accurately disseminate the key information of enterprise development strategy to the outside world.

Based on the balanced scorecard, Mobil's strategic map further establishes the interrelationship between themes and objectives, and clarifies the causal relationship needed to achieve themes and objectives.

4. Inspiration

4.1. Data mining technology has been widely used in power system security assessment, load forecasting, fault diagnosis, engineering cost and other fields, which has played an important role in reducing costs, improving efficiency and efficiency of power grid enterprises.

In recent years, foreign academia and enterprises have carried out in-depth analysis and Research on the application of data mining technology in power system. It can be concluded that the data mining technologies often used in power systems include (but are not limited to): artificial neural network, decision tree, time series, clustering, rough set theory, association rules and other algorithms. The application of data mining technology in power system mainly involves dynamic security assessment, load forecasting, fault diagnosis, engineering cost, marketing support and so on.

4.2. In the era of big data, there is an urgent need for enterprises to have a data analysis and mining product that can provide a package of solutions to meet the demands of various types of users.

In the era of big data, enterprises not only want to apply data analysis and data mining technology to improve work efficiency, but also face the problem of too high technical threshold. They urgently need a data analysis and mining product that can provide a package of solutions,

while meeting the needs of data scientists, business developers, data analysts and product managers. A variety of users, such as decision analysts, put forward diverse demands. This product should be equipped with a rich algorithm library, can access a variety of data sources, can achieve data connection, data preparation, data analysis, data modeling, model deployment, process monitoring and other data processing processes, with functional scalability, is both a tool and a platform. Baidu's Jarvis is such a product.

4.3. Balanced Scorecard and Strategic Map Theory can be used as an important reference for companies to build decision-making scenario database.

Balanced Scorecard and strategic map are effective tools for strategic analysis and performance management of business operations. Balanced scorecard includes four levels: finance, customers, internal processes and learning and growth. Each level also includes many themes and goals. On the basis of the balanced scorecard, the strategic map further establishes the relationship between the various themes and objectives, and clarifies the causal relationship needed to achieve the themes and objectives.

Balanced Scorecard and Strategic Map can be used in the whole process of the company's decision-making scenario database construction, and provide an important reference for the company to divide the different categories of scenario database and determine the specific scope of the analysis theme.

Acknowledgments

This work was supported by the science and technology project of State Grid Corporation of China, called “Research and Application of Data Mining Implementation Library on Company Operation Monitoring”.

Reference

[1] J.Q. Li, J.Z. Liu, C.L. Niu and L.Y. Zhang. Proc. International Conference on Machine Learning and Cybernetics (Guangzhou, China, August 18-21, 2005), Vol. 3(2005), p.1642.

[2] J. Lou, H. Lu, J. Xu, and Z. Qu: A data-mining approach for wind turbine power generation performance monitoring based on power curve, Vol. 10 (2016) No.2, p.137.

[3] K. Mollazade, M. Omid, F. A. Tab, Y. R. Kalaj and S. S.Mohtasebi: Data mining-based wavelength selection for monitoring quality of tomato fruit by backscattering and multispectral imaging, Vol. 18 (2015) No.4, p.880.

[4] Yue Jinyan, Jin Shuiying. Construction of enterprise performance evaluation system based on Balanced Scorecard [J]. Heilongjiang Foreign Trade and Economics, 2011 (09): 112-115.

[5] Zhang Yuemei, Luan Qingwei. Research on Strategic Implementation Framework Based on Balanced Scorecard [J]. China Soft Science, 2003 (02): 86-90.