

Comprehensive Evaluation and Prospect of Data Mining Tool Base for Company Operation Monitoring

Guang Chen^{1,a*}, ZhanJie Liu^{1,b}, XinDa LI^{1,c}, WeiPing Cui^{1,d}, YuKun Song^{1,e}, YuBing Kang^{1,f}

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

^achenguang@sgeri.sgcc.com.cn, ^bliuzhanjie@sgeri.sgcc.com.cn, ^clixinda@sgeri.sgcc.com.cn, ^dcuiweiping@sgeri.sgcc.com.cn, ^esongyukun@sgeri.sgcc.com.cn;

^fkangyubing@sgeri.sgcc.com.cn

*The corresponding author

Key words: Comprehensive Evaluation; Professional and technical personnel; Train

Abstract: In order to better use the research results to support the company's related business development, this topic intends to make a comprehensive evaluation and outlook from three dimensions (personnel dimension, time dimension, R&D dimension) on the application status, promotion prospects, use methods of the research results, and put forward targeted optimization proposals, and clear. The technology roadmap of the project will be further popularized and applied in order to give full play to the potential and practical value of the project research results.

I. Personnel Dimension

The data mining tool library of company operation monitoring contains a series of tools, which is a collection of tools. There are not only lightweight tools (such as visualization tools for power grid maps, monitoring and analysis tools for users' electricity theft) but also comprehensive tools with high coding ability to skillfully and completely use various functions in the tool library, and general tools (such as transportation) for multi-task. There are also special tools for specific links and fields (such as data analysis and verification evaluation tools, data cleaning tools, data desensitization tools, etc.). In addition, by providing external interfaces, the company's operational monitoring data mining tool library also includes some common and mature open source tools and commercial tools, which can basically meet most of the current and future data mining and data analysis needs of operational monitoring business.

1. General Operational Personnel

General business personnel lack the necessary professional programming ability, data analysis ability is not strong. Short-term operation training will enable them to master lightweight tools in the company's operation monitoring data mining toolbox (such as power grid map visualization tools, user stealing behavior monitoring and analysis tools), comprehensive data mining tools with "drag, pull, pull" characteristics, simple and easy-to-operate EXCEL, etc. Tableau and other commercial tools make it possible to use the above tools to carry out some routine data mining and data analysis work.

2. Professional and technical personnel

Professional technicians have high coding ability and strong data analysis ability. This part of the staff can skillfully use all the tools in the company operation monitoring data mining toolbox. Nevertheless, in order to give full play to their professional advantages, it is suggested that they mainly use special tools such as data analysis and verification assessment tools, data cleaning tools,

data desensitization tools to do data preprocessing, which can not be done independently by ordinary business personnel, and use high-level programming such as SQL, R, Python. The open source tools and commercial tools required by capabilities focus on solving some complex data mining problems and data modeling.

3. Decision-making and management personnel

As the leadership of an enterprise, decision-makers and managers are often busy and do not have professional coding skills. The advantages of decision-makers and managers are that they have a wide range of knowledge, strong business and business management capabilities, are familiar with the overall situation of the enterprise, and are often "unique" in the main problems and key bottlenecks faced by the enterprise. Therefore, decision-makers and managers should not directly use the tools in the company operation monitoring data mining toolbox to carry out data analysis, which is more suitable for data mining requirements and business scenario design. The specific operation of the tools can be handed over to general business personnel and professional and technical personnel.

II. Time Dimension

At present, the comprehensive data mining tools in the data mining tool library of company operation monitoring have been applied in some provincial companies' customer service, operation and maintenance business areas, and special tools such as data cleaning tools, data desensitization tools, data analysis and verification and evaluation tools have also been carried out in the data preprocessing link. Some application attempts preliminarily meet the data mining and data analysis needs of the operation supervision, and support the operation supervision to carry out high-efficiency and high-quality modeling analysis and results display business. However, there are also some problems in this process:

Firstly, the comprehensive tools of data mining and special tools such as data cleaning and data desensitization, which are developed on this topic, have not yet become the main tools for operation supervision to carry out business.

At present, there are many data mining software on the market, including data mining tools in specific fields and general data mining tools, enterprise-level data mining software and small, lightweight data mining software for specific application problems, among which there are many excellent commercial tools and flexible open source tools.

At this stage, Operations Supervisor has introduced some commercial tools and open source tools according to business needs, and at the same time, it uses several data mining and analysis software to carry out its work. For example, in thematic analysis, Operations Supervision introduced Oracle, EXCEL, SPSS, R and other mainstream data processing software; in visualization applications, Operations Supervision introduced Tableau, VBI and other visualization tools to configure the monitoring screen. However, the integrated tool of data mining for operation supervision based on this topic has not yet become the main tool for operation supervision to carry out business.

Secondly, operational supervisors generally lack professional skills in data analysis, and the use of self-developed tools and imported tools is affected to a certain extent.

Operations supervision practitioners mostly come from marketing, scheduling and other professional fields. Their code programming ability and data analysis ability are not high, and their professional knowledge and skills in data processing, data modeling and so on are lacking.

For the lightweight tools in the tool library and the comprehensive tools with the characteristics of "drag, pull, drag", simple and easy-to-use EXCEL, Tableau and other commercial tools, operators can normally use these tools to carry out routine data mining and data analysis after simple training; for Python, Sklearn and other complex. Miscellaneous open source and commercial tools, as well as specialized tools such as data analysis, verification and evaluation, data cleaning, data desensitization, etc., can be mastered by operators only after long-term and systematic training.

Thirdly, the introduction and post-evaluation of data mining tools are lack of institutional norms. Repeated and invalid imports occur occasionally, resulting in waste of resources.

At present, the company's big data application is mainly based on the independent demand and research of various professions and units. Decentralized management results in repeated introduction and development of data mining tools with the same or similar functions, which wastes a lot of internal resources. In addition, the addition, deletion and updating of data mining tools should be balanced between the professional flexibility of software and the universality of software, taking into account the practicability and cost-effectiveness of tools, which requires professional scientific calculation.

However, at present, the company lacks a clear system and process for this link, which makes it impossible to hold accountable even if the tool is invalid. At the company level, it is urgent to manage the process of introducing data mining tools as a whole.

1, recent

Firstly, we positioned the comprehensive tool of data mining in operation supervision as a universal and basic tool to actively promote the operation supervision business. Fully implement the deployment and application of data mining integrated tools in operation supervision business, take comprehensive tools as a general and basic analysis and exploration tool to carry out operation supervision business, carry out sub-thematic and sub-scenario analysis and modeling, and carry out technology exchange, mining competition organization and sharing of results based on comprehensive tools.

The second is to train the operational personnel systematically in their professional skills. It is suggested that systematic training of mature tools such as RapidMiner and Tableau should be strengthened to enhance the tool application ability of operators in operation supervision, and packaging and training of common software packages such as python, sklearn and matplotlib should be organized so as to reduce the difficulty of personnel's use and fully release the value of existing tools.

2. Medium and long term

Promote the company to establish an open and transparent data mining tool introduction and post-evaluation system. Promote the company to establish the introduction and evaluation system of data mining tools, clarify the process of adding and deleting tools, updating tools and implementation plans, and stipulate that special departments and personnel should complete tool evaluation, tool library maintenance and tool training.

Fully implement the application and landing of the data mining tool library, timely obtain feedback from operators on the use of tools, periodically evaluate the use of data mining tools and tool library, and dynamically adjust the function and composition of tools in the tool library according to the operational requirements of the operation monitoring. Keep track of the development trend of open source tools and commercial tools, regularly and optimally incorporate them into the operation supervision data mining tool library, and promote the continuous improvement and update of the tool library.

III. R&D Dimension

1. Outcome support

Under the background of the company's vigorous promotion of the construction of "ubiquitous power Internet of things", the research results of this topic can support the state-owned network Internet department, large data centers, provincial companies and other institutions to carry out data desks, data models, data applications and other key tasks. The details are as follows:

First, support the company to carry out data desk business.

According to the requirement of the company's ubiquitous task of building the Internet of Things, the data station is oriented by the application demand, supporting the access and storage of a large number of terminal data collected by pilot construction units, providing open data sharing services for all professions, grass-roots units and external partners, and continuously precipitating common data service capabilities. The data station includes many technical components such as data access, data storage and data calculation, which involve many manufacturers, poor technical system, complex technology integration among components, lack of friendly tools, high requirements for professional ability and difficult application.

The tools in the database of data mining tools developed by this project cover all the key links in the basic process of data mining, and realize the connection with the external tools through integration technology, which can effectively support the users of various technical abilities to carry out data processing, analysis, exploration, modeling, display and sharing. Service-oriented approach promotes data value mining and decision support.

The second is to support the Internet Department and large data centers to carry out normal monitoring, thematic monitoring and large data mining analysis.

According to the work arrangement of Internet Department and big data center, data application mainly includes normal monitoring and analysis, thematic monitoring and analysis, cross-professional large data mining and analysis, etc.

The company decision scenario database designed and constructed in this project can support the Internet Department and large data centers to carry out normal monitoring and analysis; carry out thematic monitoring and Analysis on business environment, large customers and high-voltage users'electricity cost recovery risk; and carry out cross-professional large data mining and analysis.

Third, support large data centers to develop data value-added services.

National Network Big Data Center is the leading unit of the company's data value-added services. The algorithms, tools, methods and scenarios developed in this paper can support large data centers to compile business plans and work plans for data value-added services, and on this basis, form a number of replicable and popularizable business models and data products of data value-added services for power grid enterprises.

2. Deepening Suggestions

On the one hand, from the point of view of technology and tools, the tools in the data mining tool library of company operations supervisor studied and constructed in this topic are enough to support the current and future data mining needs of operations supervisor. On the other hand, with the rise of data processing technologies such as knowledge mapping and semantic retrieval, the importance of "business scenarios" for data analysis is unprecedented. The key to the effectiveness of data processing technologies such as knowledge mapping and semantic retrieval lies in the judgment of "business scenarios".

However, the lack of "business scenarios" with high value, strong specificity and decision-making level is one of the main shortcomings faced by operational supervision in data analysis. The inadequate development and utilization of "business scenarios" by Operations Supervisors makes it difficult to produce influential research results.

It is suggested that the Ministry of Internet should strengthen the design and development of "business model" on the basis of existing self-research tools and data mining tools purchased from abroad, focusing on developing a number of high-value, exclusive and decision-making-oriented "business" centered on the need of the company to build a "three-type, two-network, world-class" energy Internet enterprise. Scene.

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] cao hanyu. Analysis on the application of big data in post-legislative evaluation [J]. Communications of nature dialectics,2018,40(11):11-19.
- [2] Huang Jiejun, Pan Ping, Wan Youchuan. Applied Research of Data Mining Technology [J]. Computer Engineering and Application, 2003 (02): 45-48.

- [3] Liao Zhiwei, Sun Yaming. Data mining technology and its application in power system [J]. Power system automation, 2001 (11): 62-66.
- [4] Hao Xianchen, Zhang Degan, Gao Guanglai, Zhao Hai. Problems in Data Mining Tools and Applications [J]. Journal of Northeast University, 2001 (02): 183-187.
- [5] Jiao yafei. Information quality assessment of power system based on big data [D]. North China electric power university,2017.