Data analysis method for power big data

Deshi Kong*, Xin Cao[a], Xuefeng Xiong[b]

State Grid Sichuan Marketing Service Center, ChengDu,SiChuan, 610000, China

*qubatao37050tl@163.com

[a]suquejiang4168@163.com

[b]qiaojing505975@163.com

Abstract. This paper discusses a data quality analysis and evaluation mode based on power big data. Its basic content is to use the analysis platform template to conduct acquisition control and process rule analysis, and then use the data quality management module to call out the platform database system that realizes the inspection and analysis function, and retrieve and generate inspection data from the entity base. The analysis program will classify, count, sort, and sort the data, The final generated quantitative indicator data reflecting the project implementation status and data quality will be saved in the analysis result table. The analysis result report called from the middle platform can obtain a more detailed data quality analysis and evaluation report reflecting the data quality of various quantitative projects. This paper has improved the intelligent degree of information quality management and evaluation, realized the intelligent control of information service quality, adapted to the needs of large-scale information service management, and completed the quantitative research and evaluation of information completeness, timeliness, accuracy, accuracy and other important parameters.

Keywords: big data, power, control model.

1. Introduction

With the development of China's electric power, the process of information construction is deepening, and with the gradual expansion of the total amount and type of service data information in the information system, the need for data sharing is also imminent.[1] As a resource, service data information is characterized by multiple disciplines, high data collection density, high frequency, and complex information processing.[2] It is an important subject supporting the construction and application of the national information development project.[3] At present, a large number of power data analysis results have been accumulated in some areas of the power grid.[4-7] There will inevitably be a large number of abnormal, redundant and incomplete data results here.[8] As a result, the quality problems of the analysis results of a large number of power data, such as exceptions, redundancy and omissions, will become increasingly prominent, and even cannot meet the requirements of data mining algorithms.[9-12] Therefore, it will be a great challenge to accurately find a large amount of useful information, because there are too many meaningless components in the results of massive factual data, the implementation effect of data mining algorithm will be affected.[13] At the same time, with the gradual development of practical applications, a large number of data information will be repeatedly entered, stored, and the quality of data analysis needs to be improved. The improvement of data analysis quality has become the key in the implementation of data analysis and mining system.

In the face of the problems of low data service quality caused by the characteristics of excessive power consumption data, extensive sources, numerous categories, lack of unified standards and specifications, and backward data service quality supervision system, the lack of traditional data inspection means has been unable to meet the needs of the rapid development of the current power supply industry. Therefore, it is imperative to build new data quality standards and develop new data service quality inspection means.

In order to meet the operation of power supply enterprises and improve information efficiency, it is necessary to form a complete information specification, management and evaluation process, rely on scientific and rigorous information monitoring and quality management system to continuously improve information efficiency, establish a complete information quality control framework and a complete and reasonable information value evaluation framework, restrict the deep exploration of information

resources and the all-round quality control of power supply enterprise information, and lay the foundation of information, Improve information efficiency, ensure that information is correct, timely, efficient and reliable, and provide a strong guarantee for information integration and mining. The technical problem we need to overcome is to propose a data capability classification evaluation model based on power big data analysis, which provides a strong guarantee for information integration, mining and utilization.

2. Data quality assessment system based on data mining

The data quality assessment system based on data mining technology mainly involves data quality demand classification, expert intelligence evaluation, data mining model establishment, evaluation index system setting, etc., to ensure the correctness of the evaluation, the fairness and objectivity of the conclusions. Its basic structure is shown in Figure 1.

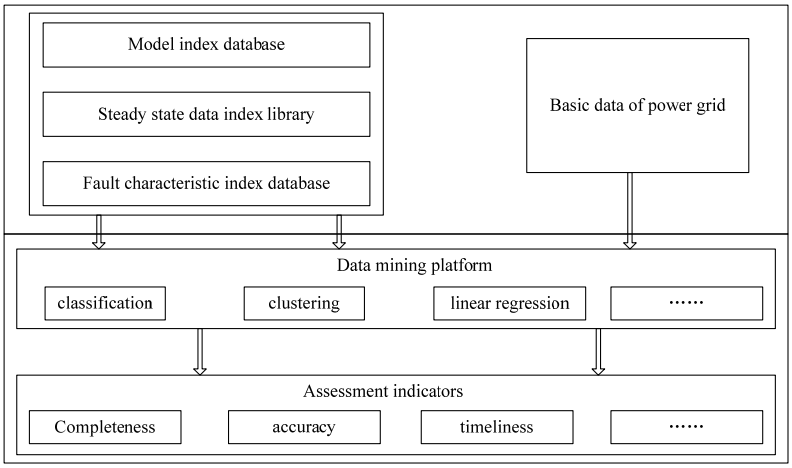

Figure 1. Data Quality Inspection Assessment Indicators

According to the requirements of the State Grid industry, special personnel were assigned to sort out the main concerns in the basic data quality management of the State Grid Corporation of China, select typical evaluation index systems, and design and establish a system library of basic data quality evaluation indicators; According to the standard process of modern data mining, it is generally conducted through data mining methods such as analysis or regression.

The comprehensive analysis and visual display platform of power grid data information quality is established. Based on the three types of data information quality evaluation index systems of modeling parameters, steady-state data information and power supply equipment fault characteristics, the massive information data are classified and refined by using data mining algorithms; Exploring the useful information and knowledge in the process, and assisting the operation and maintenance management personnel to explore the useful information and knowledge in the process, has realized the development trend prediction research and risk early warning; The development trend investigation and risk warning of power grid operation data information have been completed by using model analysis, steady-state statistics and fault characteristics to establish, train and conduct data mining mode.

From the transition of human mode evaluation to machine mode intelligence, a new information value evaluation framework is constructed from the aspects of the importance, consistency, completeness and authenticity of basic information; The quality of the operation data of the national power grid is evaluated scientifically, and the results of its quality management, assessment and inspection are comprehensively displayed by means of information visualization.

3. Data quality analysis and evaluation model

The basic components of the data quality management evaluation module include the basic management module, the digital product quality support control module, the digital product quality

determination system, the data quality management system, the data quality evaluation module, etc. The structure of this section is shown in Figure 2.

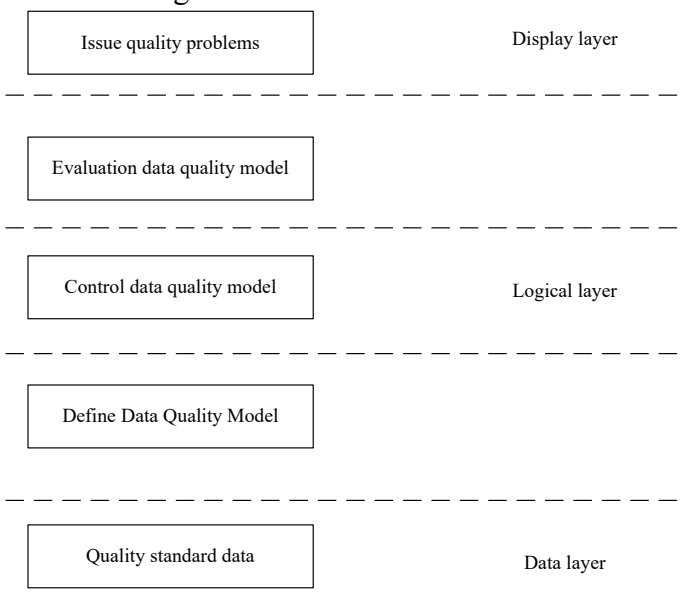

Figure 2. Functional Framework of Data Quality Analysis and Evaluation Model

(A). Basic model. The infrastructure is the foundation of the entire modeling architecture and the basic concepts and specifications of other infrastructures. It mainly includes the specifications for mapping and defining data, and introduces control criteria definition specifications, control criteria definition specifications, template definition specifications, etc.

Among them, the data specification covers the specifications that directly reflect the application field and the meaning standards of the new application library and engineering library specifications, as well as the establishment standards of new standards such as coding meaning standards, information item meaning standards, and value domain meaning standards; The constraint criterion definition standard generally describes the grammatical structure in the quality definition pattern; The standard for defining the control criteria is generally for the instructions for the use of control methods in the background operation process.

(B). Data quality definition model. As the main basis and standard of data quality research methods, the data quality feature description model is the definition of a unified standard for data quality. It is mainly expressed by quality characteristics. Its characteristics can be summarized into four important characteristics: data uniformity, data authenticity, data comprehensiveness, and data timeliness. In addition, it also involves the accuracy, availability, and legitimacy of data.

The defined data quality evaluation method based on power big data analysis can include the following contents:

Information consistency: it refers to the consistency of information contradictions and conflicts within each system, including whether the same information within the source information system is consistent, whether the source information is consistent with the extracted information, and whether the information in each management link of the data center is consistent. It is mainly used to check whether the information directly cross checked by the data is consistent.

Correctness of information: It mainly refers to whether the source of information is true, that is, the description language of information must meet the requirements of correctness and briefness, such as whether the data source is correct, whether the information domain conforms to industry norms and objective facts, whether the coding mapping relationship is correct, and whether the language logic is correct. It requires correctness and credibility to effectively reflect the real situation. Timeliness of information: it refers to the timeliness and high speed of information acquisition, transmission, management, loading and display, including timeliness of information processing, timeliness of information anomaly detection, timeliness of data timely release, etc. Information integrity: to ensure the

integrity of data information without omission, including the integrity of data source, data value, entity data type, attribute information, data, byte value, etc.

The expressed data legitimacy mainly refers to the validity of format, data type, range and business specification. Timeliness is the main standard for evaluating whether data can meet the needs. It describes the timeliness characteristics of data and its satisfaction to applications. Beneficiality mainly refers to the value of data for its own use and the degree of benefits provided by other applications. In addition, it should also include security issues, that is, the right to use data must be subject to necessary restrictions to ensure the confidentiality of data.

(C). Statistical quality model. The data quality management mode is based on the information quality definition mode. According to the determined detection range and method, the data quality is detected by self-service or manual methods, and the relevant characteristics and information of data quality are displayed, including the quality control of information detection range, information detection frequency, information detection time, information detection method and other contents.

Statistical detection objects refer to the user groups, relevant professional statistical tables and database information entities to be detected according to the collection plan. The statistical detection frequency means that the detection frequency of the stored process detection data object is set according to the collection planning and actual frequency of the statistical table. The time of data analysis and detection is based on the close moment of daily production and application, and the close moment of data analysis from generation to collection and warehousing, so as to comprehensively determine the time of the next detection. The data analysis and detection method refers to the method of executing the detection process, which can be the timed self inspection automatically controlled by the background process or the automatic detection with manual intervention.

(D). Data quality evaluation module. The data quality evaluation module is based on the definition module of data quality, and is controlled and implemented by the data quality evaluation module. The feedback quality inspection results represent the evaluation of data quality, so as to achieve the digital detection and evaluation of data quality.

(E). Data quality aided decision-making mode. The auxiliary data quality management system mainly includes report template management, authority control, database resource occupation management, etc. The core function of the data quality analysis and evaluation module is to manage the plan and constraint rules in the basic module, call the plan in the data quality management module to realize the background storage function of the inspection and analysis results, retrieve and generate the inspection results from the entity base, classify, count, analyze and sort them by the analysis program, and finally generate the data quality quantitative index results that reflect the plan implementation status, Save the results in the analysis results table, and call out the analysis results report from the hotel front desk to get a detailed data quality analysis and evaluation report that reflects the results of various quantitative indicators in the data quality problem. The program implementation flow chart is shown in Figure 3.

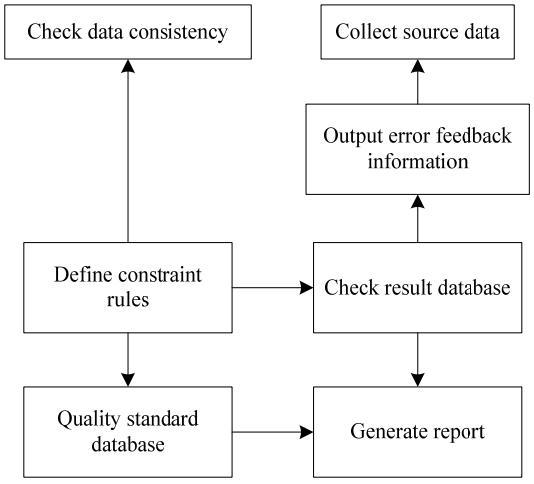

Figure 3. Process of program implementation

4. Hardware design

As shown in Figure 4, the electronic device includes a CPU, which can perform all necessary operations and processing tasks according to the microcomputer program commands stored in ROM or the microcomputer program commands added from the storage unit to the random access RAM. In RAM, it can also run all kinds of programs and information required by registers. CPU, ROM and RAM are connected to each other through the bus. The I/O port is also connected to the bus channel.

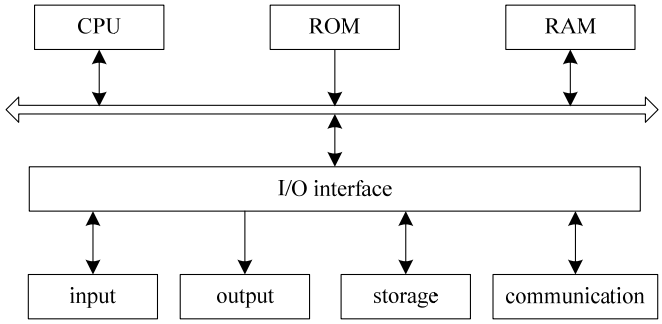

Figure 4. Block diagram of electronic equipment

Many modules on the computer are connected to the I/O port, mainly: enter the module, such as keys, mouse, etc; Input and output modules, such as various types of screens and speakers; Storage unit, such as video tape, optical disc, etc; In addition, there are communication modules, such as network card, modem, wireless communication transceiver, etc. The communication module allows computers to use a computer network containing the Internet and/or exchange data/information between other computers through various telecommunications networks.

In this article, all operations described above can be completed by at least partial use of one or more logical units. Therefore, the non limiting and practical demonstration hardware technology design logic units that can be applied include FPGA, ASIC, ASSP, SOC, CPLD, etc.

The code used to run the code in this article can be programmed in any form of one or more programming languages. These computer programs can be submitted to the information processors or controls on ordinary machines, special computer systems, and some programmable information processing devices, so as to facilitate the normal operation of the functions/actions required on the operation flow chart and/or block diagram after the computer programs are executed by the information processors or controls. Computer programs can also be run wholly or partly on a computer, or partly on a computer as a separate package of software, and partly or wholly on a remote computer or client.

5. Conclusion

The beneficial effect of this paper is that this paper mainly aims at the data analysis quality evaluation of the existing power grid big data analysis in the field of power distribution information. By analyzing the quality problems of statistical information and the main causes of their formation, and taking the integrity of data analysis, the fairness of data analysis, the integrity of data analysis, the effectiveness of statistics and other important indicators as the benchmark, this paper constructs the data analysis quality evaluation indicators, It provides a data analysis quality management and evaluation system model suitable for the national power grid big data analysis, improves the intelligent level of the statistical information quality analysis method and evaluation system, realizes the intelligent control of the data analysis quality, adapts to the needs of mass data analysis quality management, and completes the quantitative analysis and evaluation of the integrity, timeliness, accuracy, statistical consistency and other important indicators of data analysis, The service quality of power grid data analysis is effectively guaranteed, and the use value of data analysis is improved.

References

[1] Danial Hooshyar,Margus Pedaste,Katrin Saks,?li Leijen,Emanuele Bardone,Minhong Wang. Open learner models in supporting self-regulated learning in higher education: A systematic literature review [J]. Computers & Education . 2020 (prep)

[2] Mahdi Motalleb,Pierluigi Siano,Reza Ghorbani. Networked Stackelberg Competition in a Demand Response Market [J]. Applied Energy . 2019

[3] Merlinda Andoni,Valentin Robu,David Flynn,Simone Abram,Dale Geach,David Jenkins,Peter McCallum,Andrew Peacock. Blockchain technology in the energy sector: A systematic review of challenges and opportunities [J]. Renewable and Sustainable Energy Reviews . 2019

[4] Unsok Ryu,Jian Wang,Thaeyong Kim,Sonil Kwak,Juhyok U. Construction of traffic state vector using mutual information for short-term traffic flow prediction [J]. Transportation Research Part C . 2018

[5] Deshmukh Suchita,Troia Fabio Di,Stamp Mark. Vigenère scores for malware detection [J]. Journal of Computer Virology and Hacking Techniques . 2018 (2)

[6] Murilo Coutinho,Robson de Oliveira Albuquerque,Fábio Borges,Luis Javier García Villalba,Tai-Hoon Kim. Learning Perfectly Secure Cryptography to Protect Communications with Adversarial Neural Cryptography [J]. Sensors . 2018 (5)

[7] Syed Ali Raza Shah,Biju Issac. Performance comparison of intrusion detection systems and application of machine learning to Snort system [J]. Future Generation Computer Systems . 2018

[8] P. Vijayakumar,Victor Chang,L. Jegatha Deborah,Balamurugan Balusamy,P.G. Shynu. Computationally efficient privacy preserving anonymous mutual and batch authentication schemes for vehicular ad hoc networks [J]. Future Generation Computer Systems . 2018

[9] Burg Andreas,Chattopadhyay Anupam,Lam Kwok Yan. Wireless Communication and Security Issues for Cyber–Physical Systems and the Internet-of-Things [J]. Proceedings of the IEEE . 2018 (1)

[10] Nahid Gholizadeh,Hamid Saadatfar,Nooshin Hanafi. K-DBSCAN: An improved DBSCAN algorithm for big data [J]. The Journal of Supercomputing . 2020 (prep)

[11] Rudolf Scitovski,Kristian Sabo. A combination of k -means and DBSCAN algorithm for solving the multiple generalized circle detection problem [J]. Advances in Data Analysis and Classification . 2020 (prep)

[12] P. Govender,V. Sivakumar. Application of k -means and hierarchical clustering techniques for analysis of air pollution: A review （1980–2019） [J]. Atmospheric Pollution Research . 2020 (1)

[13] Ahmed Nasrallah,Akhilesh S. Thyagaturu,Ziyad Alharbi,Cuixiang Wang,Xing Shao,Martin Reisslein,Hesham ElBakoury. Ultra-Low Latency （ULL） Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research [J]. IEEE Communications Surveys & Tutorials . 2019 (1)