Research and implementation of topology data quality improvement system in electric distribution network

Junfeng Qiao*[,a], Aihua Zhou[a], Lin Peng[a], Xiaofeng Shen[b], Sen Pan[a], Pei Yang[a], Chenhong Huang[b]
[a]State Grid Laboratory of Power cyber-Security Protection & Monitoring Technology, State Grid Smart Grid Research Institute Co.,LTD.,Nanjing, Jiangsu, 210003, China. [b]Qingpu Power Supply Company, State Grid Shanghai Electric Power Company, Shanghai, 201799, China

ABSTRACT

The demand for lean management of distribution network is growing. However, the current identification of distribution network data quality lacks a unified standard and can not achieve automatic data verification and processing, which restricts the improvement of distribution network application level. In the actual distribution network application, more experts rely on their experience to solve the quality and availability problems of local distribution network topology data as needed, lacking a global data quality evaluation and solution.With the development of smart grid, the demand for lean management of distribution network is growing. However, the current identification of distribution network data quality lacks a unified standard and can not achieve automatic data verification and processing, which restricts the improvement of distribution network application level. In actual distribution network applications, more experts rely on their experience to solve the quality and availability problems of local distribution network topology data as needed, Lack of overall data quality evaluation and solutions. This paper first combs the status quo and quality problems of distribution network topology data, collates and summarizes the current status quo of distribution network topology data, summarizes the problems existing in the application of distribution network topology data, and researches and classifies the problems, laying a foundation for the development of distribution network topology data quality evaluation system and distribution network topology data verification rules.

Keywords: Data quality evaluation, distribution electric network topology, topology verification rules, evaluation data analysis

1. INTRODUCTION

Due to the lack of unified standards for the identification of the data quality of the current distribution network, the automatic data verification and processing cannot be achieved [1], which restricts the improvement of the application level of the distribution network [2]. In the actual distribution network application, more experts rely on their experience to solve the quality and availability problems of local distribution network topology data as needed [3], and there is a lack of overall data quality evaluation and solutions [4]. In view of this, the project plans to establish a distribution network topology data quality evaluation system and standards based on the above actual conditions, to provide a basis for distribution network topology data quality governance and improvement [5]. Through the research and development of automatic verification tools for distribution network data based on graph topology, it provides automatic verification means for distribution network data governance and promotes the improvement of distribution network topology data quality [6]. Through the establishment of the standard system, the breakthrough of key technologies and the research and development of automation tools, the quality and application value of distribution network topology data will be further improved, and high-quality and highly available data will be provided for distribution network application scenarios such as fault impact range analysis, load supply and demand balance, etc [7].

*1318558905@qq.com

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277903 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2688644

Topology data quality evaluation is an important part of distribution network data quality management, undertakes the task of finding topology data problems, and is the driving force and wind vane to improve data quality [8]. At present, the evaluation of topology data quality mainly focuses on the safety and reliability of distribution network, power quality and other aspects, and the analysis of data accuracy and integrity only stays at the indicator level, without in-depth discussion on the indicator realization method [9]. At present, most of the researches on the evaluation of power data quality are too simple, and the evaluation indicators are not comprehensive enough [10]. Graph computing technology has been applied to a certain extent in business scenarios such as power grid topology analysis. Distribution network topology data mainly describes the connection relationship between distribution network equipment. Graph computing technology has strong computational advantages in topology search and relationship query, and can provide technical support such as data retrieval and data calculation for the quality check of distribution network topology data.

This paper studies the construction of the distribution network topology data quality evaluation system, proposes the intelligent distribution network multidimensional data quality evaluation method, analyzes and mines the intelligent distribution network data from multiple levels, multiple directions and multiple angles, and finally constructs the distribution network topology data check rule library to verify the rationality of the distribution network topology data quality evaluation method and the effectiveness of the distribution network topology data quality evaluation system studied.

2. RELATED WORK

Distribution network topology data is the basis of distribution network planning, and the coverage of the problem library and the accuracy of the basic data are particularly critical. It is the top priority of distribution network planning to manage and maintain the massive basic data of the distribution network and ensure the integrity and accuracy of the data. At present, the basic data of the planning project library involves multiple departments. Data from different data platforms and sources have different results. Data review mainly depends on manual work, with huge workload and prone to errors and omissions.

Figure 1. Establishing electric topology data database

Strategies for improving the quality of basic data of distribution network include:

(1)Strengthen the strength and depth of professional technical guidance, and clarify the content, standards and methods of basic data collection.

(2)Integrate the data sources of each professional platform system to ensure the uniqueness of basic data sources and the accuracy of data.

(3)Strengthen the development and application of information software, reduce the degree of manual participation in the process of data collection and maintenance, reduce the possibility of errors, and ensure the scientificity, comprehensiveness and accuracy of the basic planning data to the maximum extent.

(4)In combination with the revision of the distribution network problem library, explore the automation office scheme to reduce the manpower and material resources invested in checking, verifying and correcting the basic data of the problem library. For example, the secondary development of distribution network with VBA question bank, realizing the logic and format verification of the basic data in the question bank, automatically associating each basic data information, automatically filling in part of the data, and automatically generating the problem description, can greatly improve the quality of the basic data of the distribution network.

At the same time, it reduces the time for problem library revision, allowing planners to pay more attention to the solutions to power grid problems, rather than spending a lot of time on data processing.

3. RESEARCH ON THE METHOD AND ALGORITHM OF ELECTRIC TOPOLOGY DATA QUALITY IMPROVEMENT

3.1 Electric topology data

As the basis of distribution network planning, the integrity and accuracy of distribution network topology basic data and problem base directly affect the preparation of power network problem base. Incorrect basic data cannot truly reflect the operation of the power grid, which will affect the formulation of the planning scheme. On the one hand, it may cause waste of investment, on the other hand, it may lead to the failure to effectively solve the power grid problems.

With the development of energy Internet and information technology and the proposition of the concept of ubiquitous Internet of Things, the data of power allocation and utilization in the new environment is easier to obtain, but the data quality needs to be improved. In the intelligent distribution network, data mainly comes from hundreds of millions of smart meters, smart appliances and distributed storage devices. In addition, different power companies or organizations adopt different definitions, storage and management standards, and the multi-source data obtained is usually heterogeneous and independent. The figure below shows the data integration scenario in the intelligent distribution network, involving the data acquisition layer, information system layer and business management layer. The integration of measurement data management and other business management systems gradually realizes data sharing among automatic measurement systems, marketing management systems and production scheduling systems.

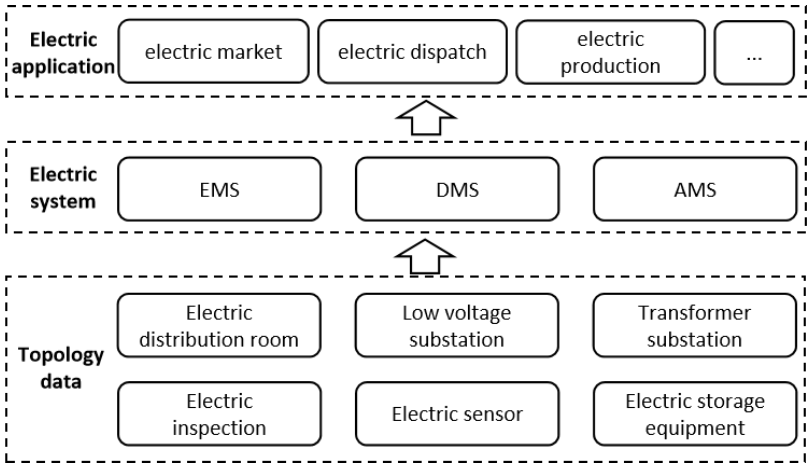

Figure 2. Electric topology data flow process

As is shown in figure.2, the integration of multi-source and heterogeneous data of distribution network in the new environment brings many challenges to data quality management, which makes information systems prone to data quality problems, mainly in the following four aspects.

(1) Power companies have established big data and cloud computing centers to process data, resulting in extremely complex data sources. The ubiquitous power interconnection construction process will lead to conflicts, inconsistencies or contradictions between data sources of different sources and structures.

(2) Due to the wide coverage of the power grid, large amount of data and insufficient maintenance level of distribution data, it is easy to make more errors in the data transmission and calculation process.

(3) The establishment of intelligent devices and systems has greatly accelerated the speed of data transmission and output. A large number of data updates will lead to the rapid generation of outdated data and inconsistent data.

(4) With the rapid development of distribution network, there are many power technology and equipment manufacturers with different standards. The data directly generated or the data standards are not perfect, and there is a greater possibility of inconsistent and conflicting data.

There are many reasons for data quality problems. Data quality problems are mainly reflected in four aspects, namely, single data source model layer, single data source instance layer, multiple data source model layer and multiple data source instance layer.

3.2 Evaluation method of electric topology data quality

The evaluation of electric topology data quality is a comprehensive and systematic project. The evaluation process of data quality needs to comprehensively consider the importance, regional, difference, real-time and other influencing factors. Through the analysis of the influencing factors of data quality, the relevant theories and methods of data quality evaluation index system are studied. In terms of the importance of data, due to the different security and reliability levels of data and different users, the requirements for the accuracy and timeliness of data are also different.

In order to make the evaluation process of topology data quality more clear and transparent, this paper designs the corresponding evaluation process based on the evaluation dimensions and evaluation model of distribution network topology data quality.

According to the designed intelligent distribution network data quality evaluation process, the specific evaluation process steps are as follows:

1) According to the needs of intelligent distribution network topology data quality evaluation, determine the data set for quality evaluation;

2) Analyze the influencing factors of data evaluation from the analysis latitude of topological data importance, region, difference, integrity, and real-time, and select five dimensions of accuracy, integrity, consistency, timeliness, and redundancy as evaluation indicators;

3) Based on each dimension of topology data quality evaluation, the corresponding evaluation rules and implementation methods are designed, and according to the corresponding evaluation rules and methods, the distribution network topology dataset is analyzed and calculated to obtain the data proportion under each evaluation dimension;

4) Based on the five dimensions of distribution network topology data quality evaluation, the weight of each evaluation dimension is determined according to the weight calculation method;

5) Calculate the data quality score according to the weight of the evaluation dimension and the proportion of the data obtained, and evaluate the level of the data quality in the "I, II, III, VI, V" level.

According to the data quality score obtained, the level of the data quality "I, II, III, VI, V" is evaluated, If the quality score is $A \in (0, a]$, the data quality level of the evaluation data object is evaluated as "V"; if $A \in (a, b]$, the data quality level of the evaluation data object is evaluated as "VI"; if $A \in (b, c]$, the data quality level of the evaluation data object is evaluated as "III"; if $A \in (c, d]$, the data quality level of the evaluation data object is evaluated as "II"; if $A \in (d, e]$, the data quality level of the evaluation data object is evaluated as "I", and the specific range can be set according to their respective requirements.

4. IMPLEMENTATION OF TOPOLOGY DATA QUALITY IMPROVEMENT SYSTEM OF ELECTRIC DISTRIBUTION NETWORK

The quality problem of electric distribution network topology data has always been the bottleneck restricting the deepening application of distribution network business. The current data verification work relies on manual sampling, and lacks efficient and standardized verification means, which seriously affects the improvement of data management. Design and develop electric distribution network topology data quality improvement system to improve the quality of electric topology data, solve the model problems caused by poor topology data quality, solve the problems of poor timeliness and inaccuracy of topology maintenance, and solve the problems of wrong analysis and calculation results caused by data.

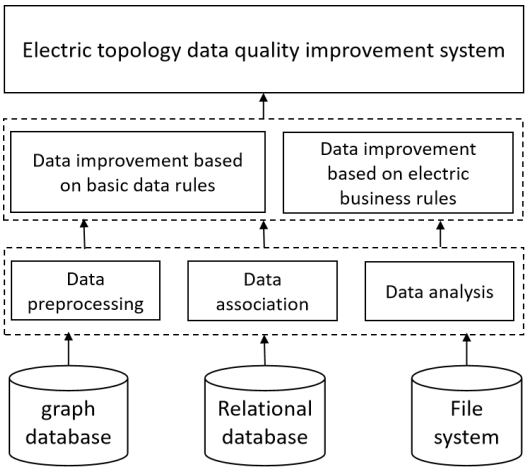

Figure 3. Design of electric topology data quality improvement system

As is shown in figure 3, the design of architecture of electric data inspection tool mainly focusing on tool development.Before the development of the tool, the operation and management mechanism of the automatic verification tool for distribution network topology data shall be defined first, and the relevant data rules and data rules shall be formulated to ensure the smooth construction of the automatic verification tool for distribution network topology data. During the construction process, it is necessary to constantly summarize and improve the data rules and business rules, and constantly improve and expand the two levels, so as to strive to achieve the goal of "construction, application and effectiveness". Guided by the concept of micro service and combined with access to multiple data sources (Mysql, Oracle, Neo4j and Tiger Graph).

4.1 Electric topology data improvement bases on basic data rules

Topology verification at the data level starts with structural processing of data. Many data models and algorithms are built on structured data. To better integrate multi-source heterogeneous data with other data sets, structural processing is an essential process. Data structure processing should first analyze the original data, extract the required information, and then further transform it into structured data. Many unstructured data and Web data exist in the form of text. Information extraction technology is needed to identify entities, attributes, relationships and other information in the text. There are also many data models that are more structured, such as the JSO format. This type of data is more flexible than relational data and requires some technical processing in the process of structural transformation. The main output form of structured processing is two-dimensional table or graph data, which requires users to determine the rules used in the data conversion process.

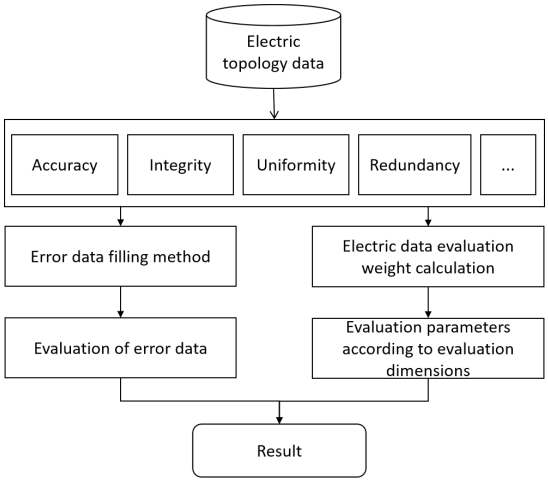

Figure 4. Electric topology data improvement bases on basic data rules

As is shown in figure 4, according to the association of electric topology and the need of topology editing, the electric topology is divided and merged, and the topology within the specified range is divided into two-dimensional grids according to the two-dimensional boundary division requirements of the grid.

According to the results of electric research, the large-scale electric topology network is divided into many small electric topology sub zones, and the analysis, monitoring and optimization of each zone can not only improve the stability of the entire electric system, but also optimize the control effect and reduce the operation cost. The division of reactive electric and voltage is of great significance to the security and stability of electric system. However, the traditional division of reactive electric area is mainly based on geographical units, with reference to operating experience, lacking of systematic theoretical basis. The characteristics of reactive electric are used to partition and identify the electric grid. This method accords with the operation characteristics of complex electric grid.

The division of the electric topology area must comply with the following three basic principles. The first is that each zone should contain at least one generator node to ensure the supply of reactive electric and the incontestability and stability of the voltage within the zone. The second is that the reasonable electric grid topology division requires the balance of supply and demand within the electric grid partitions. The partitions are connected by connecting lines, and the adjacent electric grids are standby for each other. The third is the strong coupling within each zone of the electric grid, and the weak coupling between zones, that is, the voltage change of this zone has little impact on the voltage change of adjacent zones.

4.2 Electric topology data improvement bases on electric business rules

At the electric topology business level, the causes of the problems at the business level are sorted out. According to the grid business and business rules, the problems at the business level mainly include equipment connectivity, line and single line diagram consistency, distribution transformer and substation area diagram consistency, "figure number" consistency, etc. Aiming at the above problems, an automatic business verification rule system is built to check the above several topology data business level problems.

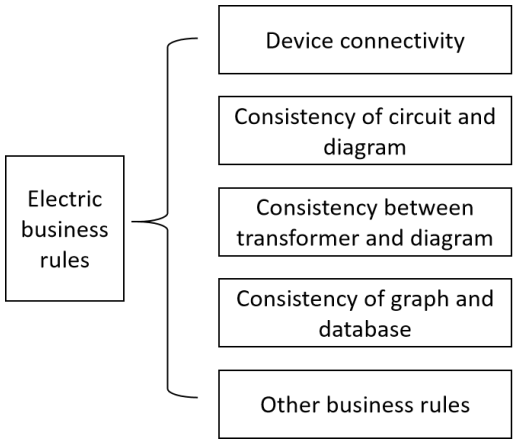

Figure 5. Electric topology data improvement bases on electric business rules

As is shown in figure3, the rule base for automatic improvement of electric topology data quality at the business level is built as follows:

1) Equipment power supply circuit detection: according to the current flow direction of the line equipment, start from the designated equipment and query according to the flow direction sequence of the power supply current to see whether there is a circuit that can return to the equipment. If there is a circuit, it needs to be identified and checked.

2) No power supply detection: trace the power supply of the equipment. Start from the current equipment and reverse search its power supply according to the current flow direction. If the end of the connecting line is found and no power supply is found, the equipment is judged to have no power supply and it is identified for verification.

3) If the equipment in the station is lack of interval, the interval attribute of the equipment in the station shall be searched. If the interval information is missing, the subsequent equipment identification and analysis will be affected. This situation shall be identified and checked.

4) The voltage of the equipment is inconsistent. In the distribution network, if the equipment connected under the same power supply does not undergo voltage transformation, but the voltage of the equipment is inconsistent with the voltage of the power supply point, the voltage of the equipment must be wrong, and it is also the object of verification and error correction. Only after the voltage transformation, the voltage of the equipment and the power supply point is inconsistent, so the above conditions are checked.

5) The line is inconsistent with the single line diagram. If the line is inconsistent with the simplified single line diagram in terms of the connection mode of the line and equipment, or the type and number of the equipment are inconsistent, the corresponding data shall be identified, checked and corrected.

6) The distribution transformer is inconsistent with the station area map.

7) The team is inconsistent with the station building.

8) Check the integrity of the account address, which mainly checks whether the authenticity of the account address and the integrity of the account address can truly reflect the real address of the equipment or line.

9) The order of the equipment drawing end is inconsistent, and the connection order of the equipment at the drawing end is inconsistent. This is because the equipment connection order is wrong due to the wrong segment information when maintaining the drawing information. This situation must be checked.

10) If the nameplate lacks an account, the account information of the nameplate shall be checked. If the account is missing or incomplete, the account information of the nameplate shall be supplemented through manual judgment or association analysis.

11) The drawing is inconsistent with the nameplate, and the equipment information on the graphic end is inconsistent with the nameplate information of the equipment. It is necessary to compare the two information, find out the differences, analyze according to manual judgment or equipment connection relationship, and correct the relevant information on the drawing or nameplate.

12) The graphic station buildings in the ledger are inconsistent, and the station buildings of the equipment in the ledger are inconsistent with the station buildings of the equipment in the graph. This is because there is a problem with the graph maintenance information. Compare the inconsistency between the station buildings of the ledger and the graph, find out the correct information of the station buildings, and update and maintain the information of the station buildings of the graph.

13) If the nameplate lacks the station building to which it belongs, the situation of the nameplate lacking the station building can be checked, and the station building to which the nameplate belongs can be supplemented by manual judgment or association analysis.

Above all, at the business level, research and establish evaluation indicators from the logic, relevance, integration and other dimensions of the business. On this basis, research and build a multi-level, multi-directional and multi-dimensional distribution network topology data quality evaluation system to help understand the true level of distribution network topology data quality and support the deepening application and decision-making of distribution network.

5. APPLICATION RESULTS

The research results of this paper establish the distribution network topology data quality check rule base system, and the results can be directly applied to the company's distribution network topology data governance and data quality improvement work. On this basis, further improve the rule base, and can be extended to the integration of the main distribution network topology data governance and data governance improvement. Through the research and application of the results of this project, we accelerated the governance of the company's distribution network topology data, improved the quality of distribution network topology data, laid a high-quality and highly available data foundation for the value discovery of topology data, and further promoted the deepening of distribution network applications.

In the next, we will continue to improve the effective of the system through the establishment of the standard system, the breakthrough of key technologies and the research and development of automation tools, further improve the quality and application value of distribution network topology data. Using the cleaned topology data of distribution network to

realize the single feeder topology simplification of medium voltage distribution network, multi-level transfer path identification and other distribution network applications.

ACKNOWLEDGMENTS

This work was supported by State Grid Corporation of China's Science and Technology Project (5400-202258431A-2-0-ZN) which is 'Research on deep data fusion and resource sharing technology of new distribution network'.

REFERENCES

[1] Wang, F., Xin, H. S., Hu, L. J., et al., Research on the development of distribution network data quality improvement and data restoration system [J]. Power Grid and Clean Energy, 35 (3): 5 (2021).
[2] Guan, Y., Guo, W. Q., Shen, T., et al., On line monitoring method and implementation of distribution network data quality based on multi data source verification [J]. Microcomputer Application, 36 (1): 3 (2020).
[3] Qin, L. W., Liang, S., Gao, L. K., etc., Application analysis of comprehensive data quality management framework in the field of distribution network [J]. Journal of Electric Power System and Automation, 32 (4): 7 (2020).
[4] Yu, Y. Z., Distribution network automation data quality analysis and data processing application development [J]. Learn Computer Easily, (002): 000 (2021).
[5] Zhang, D., Zhu, J. R., Tang, H. G., et al., One map construction of the whole network based on the unified information model of distribution network [J]. Hunan Electric Power, 40 (2): 5 (2020).
[6] Wang, J. P., Li, S., Luo, Z. Y., et al., Research on new energy integrated service platform and its application in distribution network [J]. Guizhou Electric Power Technology, 023 (001): 64-69 (2020).
[7] Yao, Y., Xu, J. Y., Cui, Q. Y., Communication topology optimization of distributed economic dispatching in active distribution network [J]. Energy engineering, (2): 5 (2020).
[8] Shan, S. R., Wang, J. L., Topology optimization method of wireless monitoring network based on GSO algorithm [J]. Smart power, 48 (6): 7 (2020).
[9] Weng, Y. X., Ma, W. Z., Shi, J., et al., Auxiliary decision-making of power grid operation mode based on semi dynamic topology optimization [J]. Power engineering technology, 40 (4): 8 (2021).
[10] Li, Z., Chang, V., Ge, J., Pan, L., Hu, H and Huang, B., "Energy-aware task offloading with deadline constraint in mobile edge computing." EURASIP Journal on Wireless Communications and Networking 2021, no. 1, 1-24 (2021).