Research and implementation of electric distribution network data fusion method considering new energy

Junfeng Qiao*,a, Aihua Zhoua, Lin Penga, Chenhong Huangb, Sen Pana, Pei Yanga, Hua Gub
aState Grid Laboratory of Power cyber-Security Protection & Monitoring Technology, State Grid Smart Grid Research Institute Co.,LTD.,Nanjing, Jiangsu, 210003, China. bQingpu Power Supply Company, State Grid Shanghai Electric Power Company, Shanghai, 201799, China

ABSTRACT

Due to the diversity of data types and data organization methods of electric multi-source heterogeneous data, the organization and storage requirements of heterogeneous data are also different, so the difference of heterogeneous data must be considered for data integration and fusion, the integration and fusion level of power distribution and distributed new energy data resources needs to be improved.. The distribution network data resources including electrical equipment, spatial information, grid topology, power consumption information, operating conditions and other types of resources, taking into account the new energy, are obviously different in terms of quantity, scale, data model, data type, organization mode and other aspects. The data of distribution network and distributed energy comes from business systems in multiple professional fields, and there is a strong correlation between data resources at the business level. However, due to the certain independence between the data of various business systems for power distribution and the differences in field definitions and descriptions, traditional key field matching methods are difficult to achieve automatic data matching, and data fusion across business systems is faced with the problem of no uniform rules to follow.

Keywords: Electric distribution network, data fusion, data integration, multi source heterogeneous data, distributed new energy

1. INTRODUCTION

The production and operation information of electric distribution network is mainly distributed in distribution network production management system, marketing system, measurement automation system and geographic information system [1]. Distribution network production management system is a system for equipment management and resource scheduling in the production process of distribution network, which stores equipment account information and equipment defect information. The marketing system is a comprehensive management platform for distribution network customer service, electric energy metering, electric charge management, business expansion, reporting and installation, and is a comprehensive database of distribution network load information [2]. The metering automation system is a real-time data acquisition, processing and analysis system for the distribution network, which records the historical operation status of the distribution network in detail. Geographic information system is a platform for comprehensive display of distribution network geographic information and topology information [3]. It stores and manages distribution network topology data and geographic information in the form of equipment list.

This paper is oriented to the integration and fusion requirements of power distribution data resources taking into account new energy, based on artificial intelligence machine learning methods, data feature extraction technology, critical path analysis algorithm, etc., from the classification method and intelligent matching technology of power distribution and distributed new energy data resources, the integration technology and fusion algorithm of power distribution and distributed new energy data resources Three aspects of multi-source heterogeneous power distribution and distributed new energy data traceability technology are studied on the integration and fusion technology of power distribution data resources taking into account new energy to achieve the integration and deep integration of power distribution data resources.

*Email: 1318558905@qq.com

2. RELATED WORK

The development of domestic data fusion technology is mostly concentrated in cross fields [4]. A large number of researchers have applied emerging information technology to their respective professional fields, and completed a large number of data fusion and value mining work based on data orientation and fusion [5]. In the literature, the researchers used the enhanced adaptive reflectance extemporization fusion model to achieve high extemporization resolution image data fusion with the goal of extracting high-resolution time series images of rice planting area in the county, and combined with the decision tree [6]. The researchers in literature proposed a data fusion algorithm based on artificial bee colony algorithm to optimize BP neural network with the goal of reducing energy consumption in the communication process of the Internet of things [7]. By taking advantage of the low demand of neural network for the size of the original data set, the organic fusion of communication data was achieved, and the efficiency of data collection was greatly improved [8]. Literature, based on intelligent algorithms, significantly reduces data dimensions by noise filtering, clustering analysis and weighted fusion for heterogeneous data of the Internet of things, and achieves high-precision heterogeneous data fusion of the Internet of Things; For the massive Io T access nodes at the bottom, literature proposes a heterogeneous data fusion technology of that can achieve high uncertainty for optimization based on node information weighting and D-S evidence theory, which greatly reduces the amount of data transmission and ensures the smooth progress of users' subsequent work [9]. Literature proposed the information resource integration implementation strategy based on the general data exchange platform by analyzing the current situation and problems of power supply enterprise information and information resource application, as well as the demand for information resource integration and sharing of power supply enterprises, and studied the information resource integration scheme based on the data exchange platform to realize the interaction and integration sharing of heterogeneous data of power supply enterprises, so as to achieve the integrated application goal of information sharing [10]. Literature investigated the status quo of the data distribution of the internet of things in domestic communities, analyzed the community management decision-making objectives based on the Internet of data, then proposed a semantic data fusion framework, and finally proposed a data fusion scheme for domestic community management according to the framework. The research results play an important role in promoting the information construction of communities and improving the decision-making level of managers.

The effective matching of different system data provides a data basis for improving the distribution network planning level by using multi-source data, and ensures the feasibility of distribution network load rate analysis, reliability analysis algorithm, transfer power supply analysis algorithm, load forecasting algorithm, and comprehensive evaluation algorithm.

3. RESEARCH ON THE METHOD OF ELECTRIC DISTRIBUTION NETWORK DATA FUSION METHOD CONSIDERING NEW ENERGY

3.1 Electric distribution network data

Electric allocation and new energy data not only have different data types, but also have many data description dimensions. For equipment description, there are operation status timing data, basic file information, operation and maintenance records, work order working conditions, etc. Through the research of power allocation data resources taking into account new energy.

According to the above research results, the number of power allocation data resources is not only huge, but also has many business attributes and data description dimensions. Therefore, the data with a large number of dimensions must be dimension-ally reduced to describe the data resources with accurate data characteristics, laying a data foundation for subsequent data fusion and analysis.

The core of multi-source data fusion is to collect, filter, match and comprehensively apply data from different data sources. Analyzing the application scenario of multi-source data fusion in distribution network planning is the premise to improve the planning level by using multi-source data. On the basis of data demand analysis of distribution network planning algorithms, the following algorithm links are improved by multi-source data fusion.

(1) Through the integration of GIS system topology data and distribution network production management system account data, the visualization level of distribution network planning is improved, and the dynamic display of distribution network status and the interface operation of planning scheme formulation are realized.

(2) Through the integration of GIS topology data, equipment life data of distribution network production management system, defect data of distribution network production management system, and load data of metering automation system, the differential treatment of equipment failure rate in reliability analysis is realized to solve the problem of poor accuracy of conventional failure rate selection methods.

(3) The natural growth rate of load and the simultaneous coefficient of load are analyzed concretely through the integration of the marketing system industry expansion report data and the 15 min level real-time operation data of the metering automation system, so as to improve the spatial granularity of load forecasting.

(4) Comprehensive evaluation of distribution network planning scheme is realized by integrating multi-source data of GIS system, distribution network production management system, marketing system and metering automation system.

Table 1. Information of electric cable.

Item	Type	Key
Cable ID	VARCHAR2(32)	Yes
Cable name	VARCHAR2(120)	No
Voltage	VARCHAR2(32)	No
Operation date	DATE	No
Rated current	NUMBER	No

As is shown in table 1, VARCHAR2 (n) represents the string with the maximum character length of n; NUMBER (P, S) means the maximum number of digits is P, and the exact S is after the decimal point; DATE is the default date type data; IMESTAMP is the default date time data..

3.2 Electric distribution network data fusion method considering new energy

Principal component analysis is a data feature extraction method commonly used to extract data features. First, it is an analysis method to find the best feature vector from redundant data, and then convert the data from the original dimension space to a new feature space through a certain dimension transformation. For example, the original dimension space is three-dimensional (x, y, z), and x, y, z are the three bases of the original space, Use a new coordinate system (a, b, c) to represent the original data, then a, b, c are the new bases, forming a new feature space. In the new feature space, it is possible that the projection of all data on c is close to 0, that is, it can be ignored, and the data can be directly represented by (a, b). In this way, the data is reduced from three-dimensional (x, y, z) to two-dimensional (a, b). Therefore, using the principal component analysis method to extract and reduce the dimensions of power distribution data features will improve the data description efficiency. The feature extraction of power distribution data mainly includes the following steps:

①Zero average the original data. For the operation status description of a transformer equipment, including voltage, current, frequency, capacity, etc., the units and mathematical quantity levels of each parameter are inconsistent. The first step of feature extraction is to standardize the data of different units and quantity levels, integrate all kinds of data near the central store through centralization, and then further standardize the processing, so that all kinds of data have the same value range in each dimension. The original data is grouped into n rows and m columns of matrix X by column, and each row of X and the data representing an attribute field are averaged to zero, that is, the average value of this row is subtracted. For the method of centralization, see Formula 1.

$$x^{'} = x - \mu \tag{1}$$

In the above formula, x is the original sample data, μ Is the sample expectation. Through the above transformation, we get a new sample data set with the expectation of 0. For the standardized method, see Formula 2.

$$x^{'} = \frac{x - \mu}{\sigma} \tag{2}$$

Divide by the standard deviation of the original sample data on the basis of centralization to obtain a new sample data set with an expectation of 0 and a standard deviation of 1. Compared with the original data set, the original data set has a large degree of dispersion. The data sets are not merged with the origin as the center. After centralization, the data set

basically forms a cluster with the origin as the center. After standardization, The distribution range of the data set on the x and y dimensions basically falls within the [- 2,2] interval.

4. IMPLEMENTATION OF THE METHOD OF ELECTRIC DISTRIBUTION NETWORK DATA FUSION METHOD CONSIDERING NEW ENERGY

The classification of power distribution and new energy data resources is mainly carried out from two dimensions: one is from the business level, according to the business use demand of power distribution data resources, the quantitative business classification index system, and according to the business theme; the other is from the data level, combined with artificial intelligence neural network machine learning method, through data resource classification sample learning, Automatic classification of power allocation and distributed new energy data resources.

The distribution network production management system, marketing system, GIS system and measurement automation system have all established a data platform based on the Open Database Connectivity (ODBC) technology, and can be managed and operated through the Structured Query Language (SQL) sentence base. ODBC technology is an open database interconnection technology proposed by Microsoft Corporation. Its basic idea is to provide users with a simple, standard and transparent public programming interface for database connection. The developer implements the underlying driver according to the ODBC standard, which is transparent to users, and allows different technologies to be optimized according to different database management systems (DBMS). ODBC technology is widely used in various systems of distribution network, and its openness provides a technical basis for data sharing between different systems. SQL language is a standardized program statement to realize data query, extraction and update, and can realize various functions of distribution network planning database creation and management.
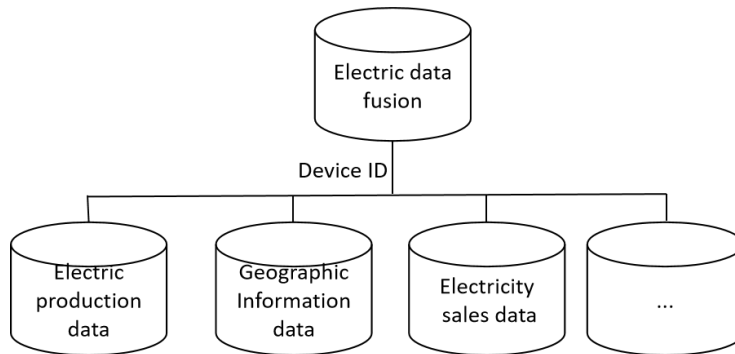

Figure1. Fusion of electric distributed data

As is shown in figure1, the device ID and naming specifications formulated by the power grid company have been widely applied in various systems of the distribution network, providing a basis for matching data items of the same equipment in different systems. On the basis of accurate matching of multi-source data information through equipment ID, as there are public data items such as equipment name and equipment dependency between systems, it can be used as an important reference for equipment matching relationship verification to ensure the accuracy of data matching.

4.1 Electric data resource classification

By sorting out the data classification rules of power distribution and distributed energy data in business themes, data indicators involved in each business are sorted out to form business theme data classification rules, and data resources are mapped to different business theme types according to business theme data classification rules.

Table 2. Electric data resource classification based on electric service logic.

Electric service item	Description
Electric business expansion	User profile data, customer service work order, distribution transformer parameters and operation data, geographic information, etc.
Electric quality monitoring	The electricity customer collects voltage, current, frozen electricity, power failure information, etc.
Operation maintenance	Equipment operation status information, maintenance plan, maintenance work order, etc.
Electric customer service	Customer service records, complaint records, customer service work orders, etc.
Distributed energy control	Distributed energy access data, distributed energy operation data, and distributed energy operation and maintenance data.

In addition to data classification based on the electric business, it is also necessary to classify data according to the basic characteristics of the data at the data level. In addition to considering the basic characteristics of the data, data classification at the data level can also be combined with manual data labels, by learning the manually labeled data labels, training in sample data to use depth learning to find the corresponding data characteristics, and then building a data classifier, Achieve intelligent classification of power distribution and distributed energy data at the data level. Data level classification is mainly based on manually labeled data category labels, through artificial intelligence machine learning algorithms, to learn the characteristics of sample data, and then solidify it into a data classifier, through which to achieve intelligent classification of data level.
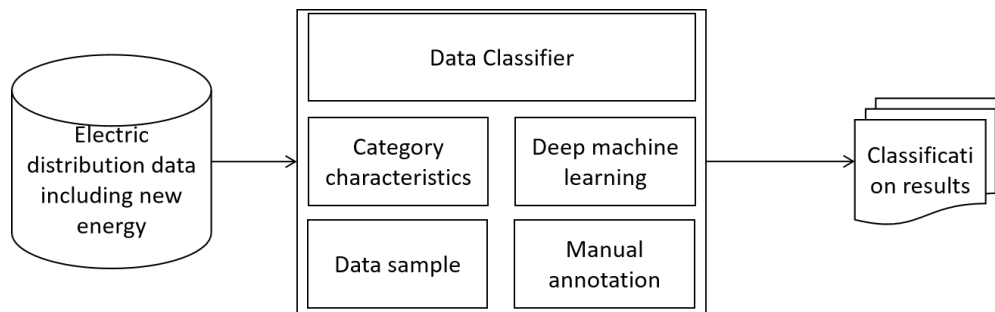

Figure 2. Classification of electric distribution data including new energy

As is shown in figure 2, the classification of electric distribution data considering new energy mainly includes category characteristics, deep machine learning, data sample, manual annotation, the classification results will be pushed to data classifier.

4.2 Intelligent matching of electric data resource

There is a strong correlation between data resources based on electric business. However, due to the certain independence between the data of each business system for power distribution and the difference in field definition and description, traditional key field matching methods are difficult to achieve automatic data matching. Therefore, it is proposed to carry out research on accurate identification methods for the business meaning of each business system data. Based on the data resources with known business meaning, a comparative sample data set is established. Through the similarity analysis between the unknown business meaning data and the sample data set, the business meaning of the unknown business meaning data is identified according to the analysis results, so as to achieve intelligent matching of data resources between the same business system and different data types. Similarity calculation is to calculate the similarity of two columns or two different data, and judge the similarity of the two data according to the similarity. This project uses the similarity calculation method of calculating Pearson correlation coefficient to achieve similarity judgment. First, standardize the data to be compared, unify the comparison data to the same dimension, and then calculate the Pearson correlation coefficient of unknown business identification data and known identification data, According to the calculation results, judge whether the Pearson correlation coefficient is within the specified confidence

interval. If it is, it is considered that the similarity between the two is very high, and they can be considered as the same type of data, and the target data can be marked with the same data identifier, otherwise it is not. By means of similarity calculation, the service identification of unknown data can be identified, so as to achieve data matching.

Find the covariance, see Formula 3.

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \tag{3}$$

Calculate the covariance of n variables to obtain the covariance matrix, as shown in Formula 4.

$$C = \left(C_{ij}\right)_{n*n} \begin{pmatrix} C_{11} & C_{12} & C_{13} & \cdots & C_{1n} \\ C_{21} & C_{22} & C_{23} & \cdots & C_{2n} \\ & & \cdots & & \\ & & \vdots & & \\ & & \cdots & & \\ & & \cdots & & \\ C_{11} & C_{12} & C_{13} & \cdots & C_{1n} \end{pmatrix} \tag{4}$$

A is a matrix of order n, if number λ And n-dimensional non-zero column vector x satisfies Ax= λ x. That number λ Eigenvalue called A, and x corresponding to Eigenvalue called A λ The eigenvector of. When the characteristic polynomial is equal to 0, it is called the characteristic equation of A, and the characteristic equation is a homogeneous linear equation group. The process of solving the eigenvalue is actually to solve the solution of the characteristic equation. For covariance matrix A, see Formula 5 for the calculation method of its eigenvalue (possibly multiple).

$$\left|A - \lambda E\right| = 0 \tag{5}$$

The eigenvalues are sorted from large to small, and the largest k eigenvectors are selected. Then the corresponding k eigenvectors are used as column vectors to form the eigenvector matrix W (n). Calculate X, that is, project the data set X onto the selected feature vector, so that we can get the reduced dimension data set X we need. The eigenvectors are arranged into a matrix from top to bottom according to the size of the corresponding eigenvalues in rows, and the first k rows are taken to form the matrix P. Y=PX is the data reduced to k dimensions. Through feature extraction of power distribution data, solve the data features that can accurately describe power distribution data resources, and provide an accurate and effective data basis for data classification and matching.
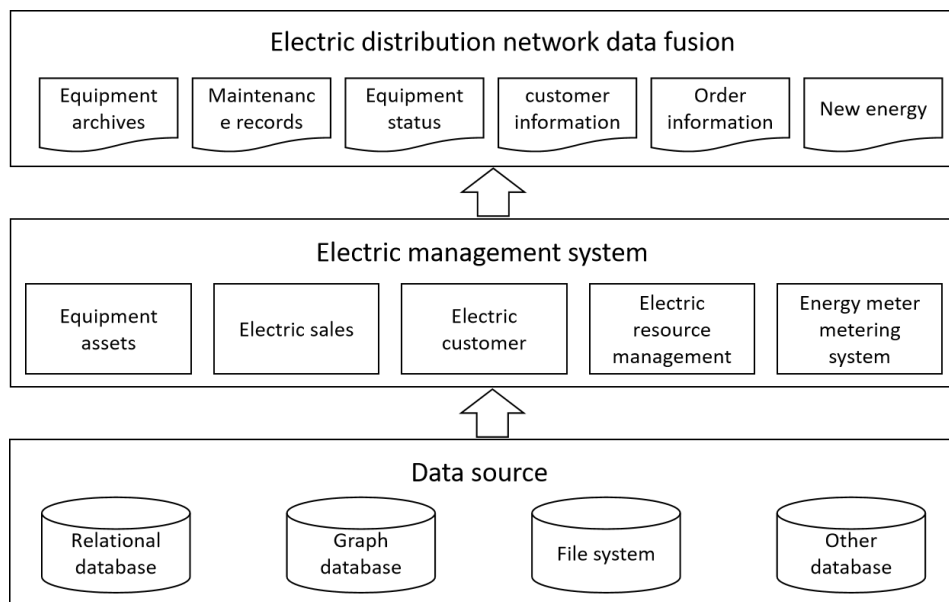

Figure 3. Electric distribution network data fusion

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 classification method of power distribution and distributed new energy data resources, propose the integration technology and fusion algorithm of power distribution and distributed new energy data resources, develop the integration and fusion tool of multi-source heterogeneous power distribution and distributed new energy data, cover resources, assets, measurement, topology, graphics and other types of data, and realize the coverage of resources, assets, measurement, topology The deep integration and fusion of multi-source heterogeneous power distribution and distributed new energy data resources such as graphics meet the traceability requirements of multi-source heterogeneous power distribution and distributed new energy data.

Based on the research results of this paper, the integration method of power distribution data resources has been pilot applied in power grid companies. More than 3PB distributed data resources in a certain region have been integrated, and the data redundancy rate has been reduced by more than 45% after the integration. Moreover, the research results of this paper are widely popularized. In the next 10 years, this result can be popularized and applied in more than three provincial power companies, Greatly improve the availability of power distribution data resources, give full play to the value of data, and better provide advanced services for power production, customer service, etc.

ACKNOWLEDGMENTS

REFERENCES

[1] Li, J. B., Research on insulator fault detection based on image recognition and intelligent algorithm [J]. China Science and Technology Information, (6): 2 (2022).
[2] He, Y., Ruan, S. W., Liu, J., et al., Research on differential operation and maintenance strategy of 10 kV distribution network based on neural network algorithm [J]. Electrotechnics, (17): 4 (2021).
[3] Kuang, H., He, X., He, M., etc., Detection of abnormal voltage data in distribution network based on bidirectional short-term memory neural network [J]. Science, Technology and Engineering, 21 (24): 7 (2021).
[4] Niu, H. P., Wang, Z. Q., Xiao, D. Y., Extraction of rice planting area at county level based on spatio-temporal data fusion [J]. Journal of Agricultural Machinery, 51 (4): 156-163 (2020).
[5] Beliakov, G., Gagolewski, M., James, S., Hierarchical data fusion processes involving the Mbius representation of capacities[J]. Fuzzy Sets and Systems, (2) (2021).
[6] Van Schyndel, R. G., Tirkel, A. Z., Osborne, C. F., A digital watermark [C]. Proc.of the IEEE Int.conf.on Image Processing 2, 86-90 (1994).
[7] Alam, F., Mehmood, R., Katib, I., Albogami, N. N., Albeshri, A. l., Data Fusion and IoT for Smart Ubiquitous Environments:A Survey[J].IEEE Access, 5, 9533-9554 (2017).
[8] Huang, M., Liu, Z., Tao, Y., et al., Mechanical fault diagnosis and prediction in IoT based on multi-source sensing data fusion.[J]. Simulation Modelling Practice &Theory, 102(0):1 (2020).
[9] Zhang, L., Gao, H. L., Wen, J., et al., A deep learning-based recognition method forde gradation monitoring of ball screw with multi-sensor data fusion[J]. Microelectronics Reliability, 75(0):215-222 (2017).
[10] Bijarbooneh, F. H., Du, W., Ngai. E. C. H., et al., Cloud-assisted data fusion and sensor selection for internet of things[J]. IEEE Internet of Things Journal, 3(3):257-268 (2016).