Research on the Construction of Hybrid Data Graph Model for Smart Grid Dispatching Based on Multidimensional Data Mining

Jiangang Lu[1*], Ruifeng Zhao[2], Huijuan Tan[3], Yue Dai[4], Chen Wang[5]

{lujiangang@gpdc.gd.csg.cn[1*], zhaoruifeng@gpdc.gd.csg.cn[2], tanhuijuan@gpdc.gd.csg.cn[3], daiyue@gpdc.gd.csg.cn[4], wangchen@gpdc.gd.csg.cn[5]}

Guangdong Power Grid Co., Ltd. Power Dispatching Control , Center,Guangzhou,Guangdong, 510000, China

Abstract. The database is more and more unable to meet the grid's demand for fast access and analysis of big data, and the large amount of data has brought greater difficulty in data analysis and processing, and also faces problems such as large grid scheduling, difficult load prediction, and fault detection. In order to ensure the smooth operation of grid scheduling, a research on the construction of a hybrid data graph model for smart grid scheduling based on multidimensional data mining is proposed. This study extracts useful features from raw data collected from various data sources, selects the graph convolutional neural network hybrid data graph model, constructs the data graph model based on the physical topological relationship of the grid and continuously optimizes and updates the model. After the simulation test, this model is able to handle a large amount of data, through the study of smart grid scheduling hybrid data graph model construction can effectively solve the problems such as the difficulty of analyzing and processing a large amount of data.

Keywords: Graph databases; Knowledge graphs; Multidimensional data mining; Smart grids;

1 Introduction

Smart grid is the latest trend in the development of power automation system in today's world, and with the advent of the big data era, human life has entered the data explosion era. It is now impossible to use traditional methods to capture and store and process data. In this case, how to store, process and use big data has become an inevitable research challenge [1]. Nowadays, many enterprises in China use real-time data analysis and Internet of Things (IoT) technology to realize comprehensive monitoring and control of power grid dispatching by means of power grid condition monitoring, power grid data storage technology and intelligent distributed system. This technology can timely discover the faults and supply abnormalities of grid scheduling, but in the complex mixed data analysis and processing, this kind of technology does not exist in regularity, and there are security problems in the operation of the grid.

Zhang Zhaoqi [2] et al. proposed the storage of insulation state text data of electric equipment based on graph database, used the knowledge graph technology to process the insulation state text data and specified the insulation state text record format, and then completed the data

PMBDA 2023, December 15-17, Nanjing, People's Republic of China

DOI 10.4108/eai.15-12-2023.2345304

storage based on the graph database. The effect of relational query was compared using MySQL relational database and Neo4j graph database. While MySQL is difficult to cope with the query of complex relational data under the large scale of data volume; Ning Wang [3] et al. proposed a hybrid relational-graph database storage system design, connecting MySQL and Neo4j through C-SQL to provide users with a unified database access interface to complete the query processing of the hybrid system; based on dynamic cost modeling The query decomposition optimization is carried out, and the experimental results show that the efficiency of the proposed hybrid storage system is significantly improved for complex queries while guaranteeing the correct query results when comparing with MySQL, Neo4j and AgensGraph.

For this reason, this paper proposes a study on the construction of a hybrid data graph model for smart grid dispatch based on multidimensional data mining, in which the study is a smart grid using multidimensional data mining to monitor a variety of data in real time, and through the smart grid dispatch operation and analysis system integrated data report, data mining and multidimensional analysis technology, extracting information from historical data, building a comprehensive information platform, realizing a data mining model based on data mining and the smart grid system for large-scale data analysis and prediction [4]. By establishing this hybrid data graph model based on multidimensional data mining, it realizes the refined management and automated decision-making of smart grid scheduling in order to meet the growing demand for electricity and improve the requirements of energy utilization efficiency.

2 Smart grid dispatching control system

Smart grid scheduling data division refers to dividing the data involved in grid scheduling into different categories or parts according to certain rules and needs in order to manage and analyze these data more effectively. For the smart grid scheduling system, there are a variety of business types and consequently many applications, mainly including the following parts: real-time monitoring, collecting and analyzing data, intelligent auxiliary decision-making, scheduling plan, etc. [5]. The architecture diagram of the smart grid dispatch control system is shown in Figure 1:

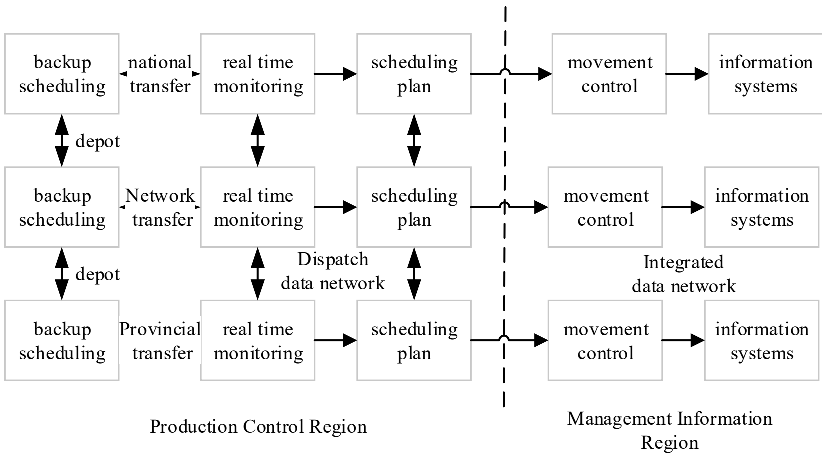

Fig. 1. Architecture diagram of smart grid dispatching control system.

Smart grid dispatch control system is a key component used to manage and monitor the operation of power systems. In it, advanced information technology, communication technology and power system control technology are combined, aiming to realize the efficient operation, reliability and sustainability of the power system [6]. On this basis, the power grid is analyzed and calculated by the formula shown below equation (1),(2):

$$\min \ or \ \max f(x) \equiv A \tag{1}$$

$$x \in W^n,\ \text{when}\ \theta_i(x) \begin{cases} \leq \\ = \\ \geq \end{cases} \lambda_i,\ i = 1,2,\ \cdots,\ m,\ x \geq 0 \tag{2}$$

where f(x) denotes the objective function, A denotes the objective value, $\boldsymbol{\theta_i(x)}$ denotes the constraints, and $\boldsymbol{\lambda_i}$ denotes the ideal constraints.

When making a decision, when the constraints are closer to the ideal constraints, their satisfaction with the decision goal is higher, and the formula for satisfaction with the decision is expressed as equation (3):

$$g(x) = \begin{cases} 0 & f(x) \leq g(x) \\ g(x) & f(x) \epsilon [g_{min},\ g_{max}] \\ 1 & f(x) \geq g(x) \end{cases} \tag{3}$$

where $\boldsymbol{g_{min}}$ denotes the target at the lowest degree of idealization and $\boldsymbol{g_{max}}$ denotes the target at the highest degree of idealization.

The above process is a claimed representation of the management and monitoring indicators of the smart grid dispatch control system, and it is expressed as the decision satisfaction of the system management and monitoring.

3 Smart Grid Hybrid Data Analysis Based on Multidimensional Data Mining

Multidimensional data mining techniques are used to discover and understand patterns, correlations, and trends in multidimensional datasets for analyzing and optimizing power system operation, planning, and decision making [7]. The multidimensional data generated during the operation of power grid is collected and predicted and analyzed using data mining techniques. By analyzing the correlations and trends between the data, it will be possible to more accurately predict aspects of grid load, reliability and energy consumption, leading to better scheduling of the grid and improving the overall efficiency of power supply. After configuring the various factors of the smart grid dispatch control system to analyze the topology of the grid, in the actual scenario of line A transformer A1, A2 will be transferred to line B, and in the system in the calculation of line loss, in accordance with the original topology of the relationship between the calculation. The topology of the grid is shown in Figure 2:

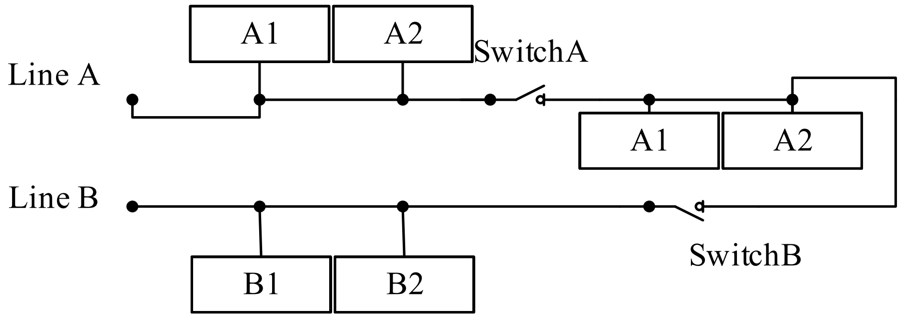

Fig. 2. Grid topology map.

In the grid topology map, the spatial dimension through multidimensional data mining is represented by equation (4):

$$\begin{cases} M\ (d) = \sum_{x=1}^{L\ (d)} f(x)|\ d|/L\ (d) \\ \mathrm{d} = \{d_{1,}\ \ d_{1,}\ \ ...d_{L\ (d)}\} \\ f(x) = x \sin \mu \end{cases} \tag{4}$$

where M (d) is denoted as the spatial correlation dimension in the multidimensional data, d is denoted as the multidimensional dataset, L (d) is denoted as the length of the multidimensional data, and μ is denoted as the angle between the dimensions.

After the computation of spatial dimension of multidimensional data mining, the formula for the computation of relational dimension of multidimensional data mining is denoted as equation (5):

$$\mathrm{C} = \frac{\sin \mu}{\ln L\ (d)} \tag{5}$$

According to the above formula for decision satisfaction and correlation dimension calculation, the object of correlation rule data mining is categorized into Boolean and numerical types, in which the rules after computational processing imply the correlation between variables. In the smart grid scheduling hybrid data mining network structure is complex and multidimensional data, the decision-making system is described through the attribute two-dimensional numerical table, where the multidimensional data information table Z is denoted as equation (6):

$$\mathrm{Z} = (F,\ \ S,\ \ V,\ \ \{F_a|a \in S_t\},\ \ \{U_a|a \in S_t\},\ \ ) \tag{6}$$

where F denotes the number of multidimensional data objects, 〖U_a denotes the multidimensional data information function, S〗_t denotes the non-empty finite set of attributes, and V denotes the logical theory defined according to S_t. The atomic formula for obtaining V according to a = v and $\boldsymbol{a} \in \boldsymbol{S_t}$、 $\boldsymbol{v} \in \boldsymbol{S_t}$is then realized by logical disambiguation, inversion and merging.

4 Constructing a Hybrid Data Graph Model for Smart Grid Scheduling

In this paper, through a number of heterogeneous databases, in the absence of a unified database access interface, the object-relational mapping is only connected to a single database, the smart grid scheduling hybrid data graph model is mainly composed of connection node class, device terminal class, and device class[8-9]. The most basic topology connection is constructed on the basis of the grid topology map as shown in Figure 3:

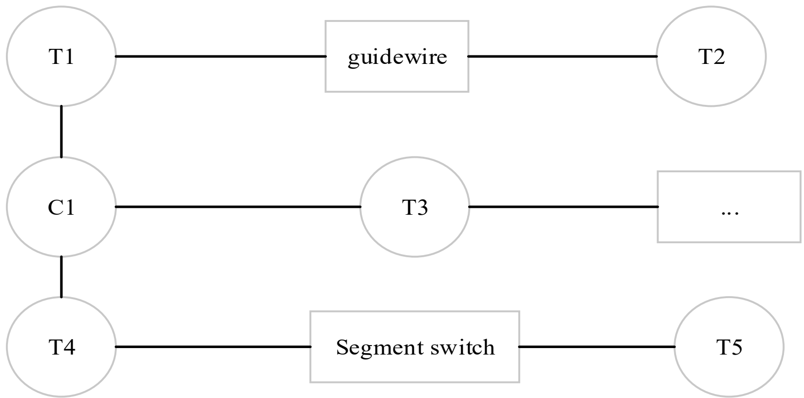

Fig. 3. Topological connection.

On the basis of topological connection, the connection node class and device sub-node class in the object class are eliminated, and only the device node class is retained[10]. Among them, in order to improve the efficiency of connectivity query for grid topology devices, Neo4j grid graph model is constructed. The model follows topological connectivity by modeling the object classes in the model as node data format in the graph database, and modeling the link relationships between objects as relational data format in the graph database[11]. Its mapping table is shown in Table 1:

Table 1. Mapping table.

Hybrid Data Graph Model for Smart Grid Dispatch	Graph models for graph databases
device class object (computing)	nodal
Topological connectivity between devices	guanxi

According to the rules of the hybrid data graph model for smart grid dispatch, only each device node is retained, excluding the terminal nodes and connection nodes of the device[12], and the graph model constructed based on Neo4j is shown in Figure 4:

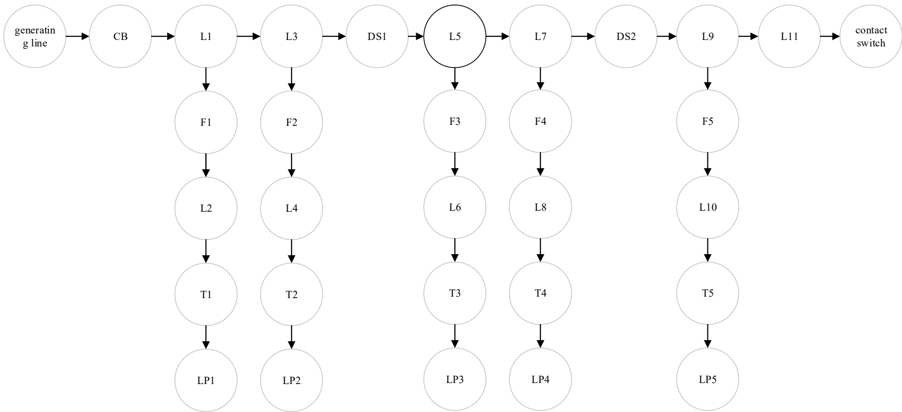

Fig. 4. Graph model constructed based on Neo4j.

The graph model consists of 31 nodes and 30 edges, in the graph model there are buses, circuit breakers, lines, fuses, distribution transformers, loads and disconnect switches included. The schema of the data is allowed to evolve flexibly with changing requirements in the graph model built on Neo4j[13]. Each node corresponds to a record in a relational database, and the attributes of nodes and edges are equivalent to the fields in the record. The content and number of attributes can be changed dynamically, and the edges between nodes can be deleted or removed freely without affecting the logic of the existing data structure [14]. This makes it very easy to add, modify or delete nodes and relations in the graph model without making large-scale changes to the entire database[15].

5 Testing and Analysis

5.1 Test preparation

The reliability calculation of the distribution system of the Neo4j graph database is carried out using the feeder of the 31-node Billington reliability test system as an example. The distribution system consists of 31 original pieces of equipment, including 30 sections of lines, 5 disconnectors, 2 circuit breakers, etc[16]. The table of equipment reliability parameters is shown in Table 2:

Table 2. Equipment reliability parameter table.

Device Type	Node Label	$\boldsymbol{\lambda}$ $/(\mathbf{substandard} \cdot \mathbf{a}^{-1})$	$\boldsymbol{\gamma}/(\mathbf{h} \cdot \mathbf{substandard}^{-1})$
Lines	L	0.046	4
Distribution Transformers	T	0.015	200
Circuit Breaker	CB	0.002	20
Disconnect Switches	DS	0.005	8
Busbar	M	0.025	10

Observe the running graph of its mixed data scrolling according to the time.

5.2 Analysis of test results

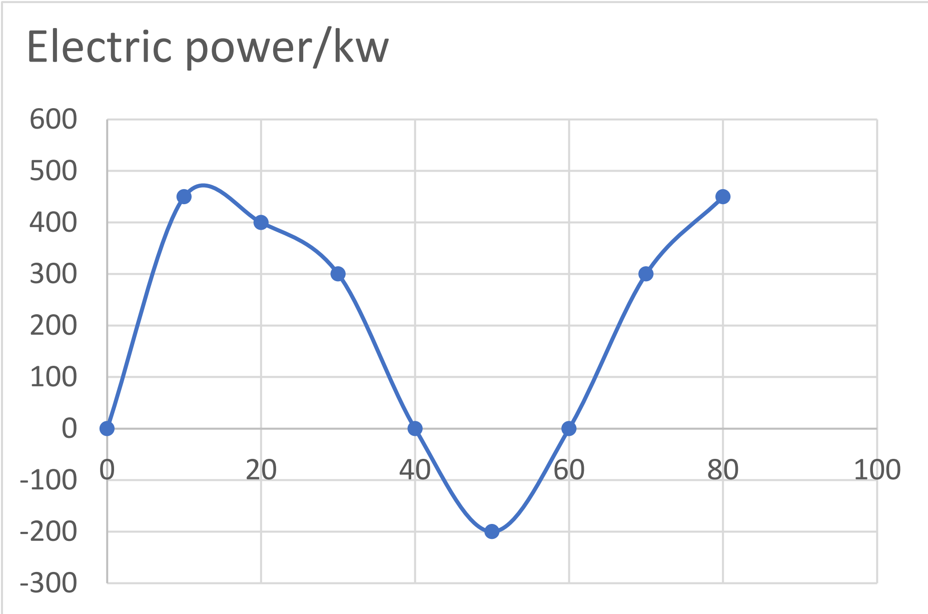

Fig. 5. Smart Grid Dispatch Data Rolling Run Chart.

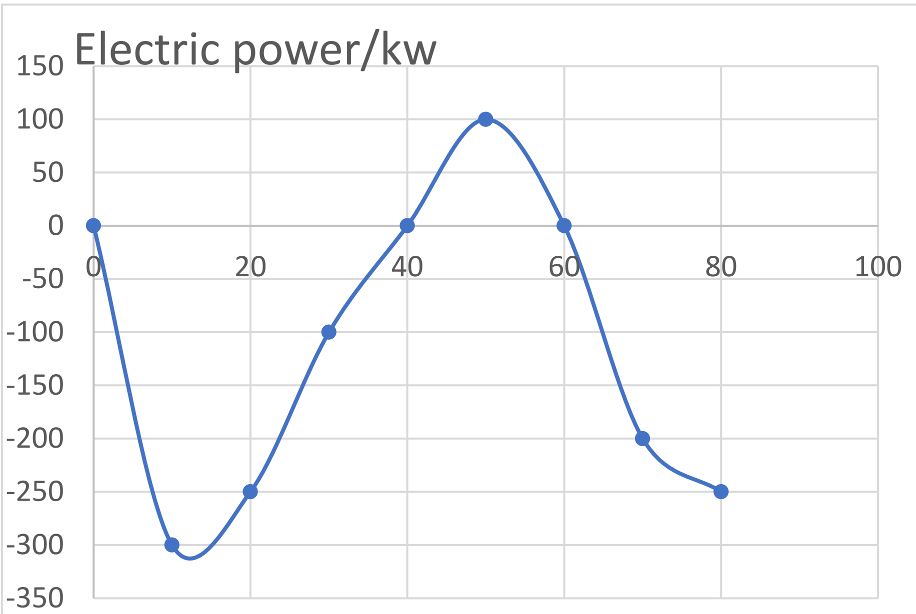

Fig. 6. Smart Grid Dispatch Hybrid Data Rolling Run Chart.

In grid scheduling, rolling runs allow continuous monitoring and adjustment of the power system operation to adapt to changing demands and conditions and to ensure its stability and efficiency under different conditions. According to the above results of the smart grid scheduling data over time, it can be seen that the smart grid scheduling data rolling run graph, is a linear curve, in the data-driven scheduling decision-making, according to the comparison of Fig. 5 and Fig. 6, it can be seen that 0 ~ 40 min is not subject to the hybrid data-driven constraints, and 40 ~ 60 min is subject to the hybrid data-driven constraints. In the smart grid scheduling hybrid data rolling operation diagram, the peak value of the peaks and peaks subject to hybrid data-driven constraints in -350 kW, not subject to hybrid data-driven constraints in the peak value of the peaks and peaks in the 480 kW. through the training to realize the mapping between the system operating state he, to the scheduling decision-making results, in the actual mixing is to carry out the dispatch decision-making at a very high speed, so as to effectively improve the accuracy of the results of the data-driven scheduling decision-making.

6 Conclusion

The article takes multidimensional data mining technology as the basis to design a hybrid data graph model for smart grid dispatching, which provides a new research idea for hybrid data graph modeling research. The research on the construction of hybrid data graph model for smart grid dispatch based on multidimensional data mining combines the grid topology map and the graph model constructed on the basis of Neo4j, and this modal data fusion can provide more accurate analysis and decision support. In future research, the experimental efforts will be increased to explore the application scenarios of hybrid data for smart grid scheduling in order to enhance the practicality of hybrid data graph model construction.

Acknowledgement.This work has been supported by a graph calculation analysis and optimiz ation software project for large-scale power grid equipment nodes(GDKJXM20222157-03600 0KK52222020).

References

[1] Pang, Hao-Lin. (2023). Research on Smart Grid Data Security and Access Control under Malicious User Environment (Master's Thesis, Dongguan Institute of Technology). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFDTEMP&filename=1023637605.nh

[2] Zhang Zhaoqi, Song Hui, Cui Guodong, Liu Guanghui, Wang Jianjun, Sheng Gehao & Jiang Xiuchen. (2020). Textual data storage of power equipment insulation status based on graph database. Electrical Automation(05),64-66.

[3] Wang, N., Zhang, W., Wang, J. H. & Ma, L. M.. (2022). A hybrid relational-graph database storage system design. Journal of Beijing University of Information Science and Technology (Natural Science Edition) (01), 58-64+70. doi:10.16508/j.cnki.11-5866/n.2022.01.010.

[4] Gong, Shuai & Song, Shan-Kun. (2022). Research on data mining methods for distributed energy systems with multidimensional association rules. Energy and Environmental Protection (10), 278-283. doi:10.19389/j.cnki.1003-0506.2022.10.047.

[5] Yu, M. L.. (2021). Research on the construction of scheduling system for smart grid based on cloud computing (Master's thesis, Southeast University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1022465848.nh.
[6] Jiahao Chen & Yuqing Jiang. (2023). Exploration of intelligent power transmission and transformation technology in smart grid. Electrical Age (09),47-49.
[7] Jia Chaoxian. (2023). A multi-dimensional data mining method for heterogeneous networks based on granularity computing. Journal of Chongqing Institute of Science and Technology (Natural Science Edition) (02),87-91. doi:10.19406/j.cnki.cqkjxyxbzkb.2023.02.012.
[8] Zhou, J., Liu, F. & Wang, D.. (2022). Research on topology modeling method for distribution networks based on graph database. Electromechanical Information (05), 13-15. doi:10.19514/j.cnki.cn32-1628/tm.2022.05.003.
[9] Sip Chunhong. (2022). Research on the construction strategy of English grammar knowledge map based on Neo4j graph database. Automation Technology and Application (07), 181-183. doi:10.20033/j.1003-7241.(2022)07-0181-03.
[10] Chenjing Han, Tianshi Wang, Kaiye Gao, Rui Peng & Jinhao Xie. (2021). Construction and realization of quality engineering technical knowledge graph based on Neo4j graph database. Quality and Reliability(02),50-55.
[11] Le JW. (2022). Intelligent positioning system for railroad data center based on Neo4j graph database and UWB technology. Railroad Computer Application(12),13-19.
[12]Wan Yanni,Qin Jiahu,Ma Qichao,Fu Weiming Wang Shi.(2022).Multi-agent DRL-based data-driven approach for PEVs charging/discharging scheduling in smart grid.Journal of the Franklin Institute(4),1747-1767.
[13]Zhao Enlai,Su Xinlei,Wang Yuling,Liu Pengyuan,Yang Jun Pan Shitong.(2022).Power Grid Dispatching Data Identification Method Based on Deep Packet Detection.Journal of Physics: Conference Series(1).
[14]Carina Taylor.(2019).Fine-grained data based on the associated large multidimensional data mining algorithms.Computer Informatization and Mechanical System(1).
[15]Pravin Varaiya,Felix F. Wu Janusz W. Bialek.(2011).Smart Operation of Smart Grid: Risk-Limiting Dispatch..Proceedings of the IEEE(1),40-57.
[16]Yang Shiping,Tan Sicong Xu Jian Xin.(2013).Consensus Based Approach for Economic Dispatch Problem in a Smart Grid.IEEE Transactions on Power Systems(4),4416-4426.