Application of Knowledge Mapping Technology in Power System

Lihua Gong¹, Kunpeng Liu¹*, Nuo Tian², Xiaoyi Wang¹, Jiajia Wang², and Jinpeng Zhu³

¹ State Grid Customer Service Center, Tianjin, 300309, China
² State Grid Customer Service Center, Nanjing, 211161, China
³ Beijing DataOcean Technology Co., Ltd., Beijing 100081, China
*E-mail: kunpeng-liu@sgcc.com.cn

Abstract. In order to improve the identification ability of power system topology, the application of knowledge mapping technology in power system is proposed. Based on the division of structured, semi-structured and unstructured, the power system data is extracted. The neo4j graph database is used to store the extracted data, and the cypher statement of NEO4J is used to complete the construction of power system knowledge map. On this basis, the power system topology is identified based on the results of semantic segmentation. The results show that in the simulation environment, the topology recognition rate can reach 100%. In the actual environment, the topology recognition rate increases with the increase of the amount of data, up to 100%

1. Introduction
The new knowledge in power system is complex and changeable: the structure of knowledge is complex and isomerization knowledge is increasing; Knowledge changes rapidly, and the update frequency is gradually shortened [1]. In the face of the rapid growth of new knowledge in the power system, the traditional knowledge organization and management methods have been unable to meet the needs of the current power system [2-3]. Knowledge map was proposed by Google in 2012, and has been developed rapidly and widely used in Internet, finance, medicine and other fields since then [4]. Since knowledge map was added to Google's search engine, its scale has tripled in seven months. By 2016, the knowledge map built by Google has more than 500 million entities and 70 billion entity relationship information [5]. In today's Internet field, knowledge mapping has become one of the basic technologies of semantic search, intelligent question answering, knowledge reasoning and other knowledge intelligent services [6]. Knowledge mapping can effectively organize, manage and utilize massive information, and realize intelligent knowledge extraction, reasoning, storage and retrieval. Its characteristics and application scenarios fit the needs of power system.

The application research of knowledge mapping in power system has been done by scholars. Among them, it is very important for higher education to analyze the correct topological relationship in power system. Considering that the actual topological structure of power system changes frequently and enormously due to the need of operation and maintenance, the data in power system information system is not updated in time, the liquidity is low and the quality is poor, which can not correctly reflect the actual topological structure of power system, the topology identification research is carried out [7]. Yu et al. Combined with the knowledge mapping technology and the specific situation of the
regulation field, proposed a knowledge mapping construction method for the intelligent regulation field \cite{8}. The application shows that the established knowledge map can automatically drive the process of line fault disposal, the accuracy of process identification is high, and the risk of manual disposal is effectively reduced, but the construction process is relatively complex. Based on this, this paper proposes the application of knowledge mapping technology in power system. Through this research, we hope to provide valuable reference for the development of power system.

2. Building power system knowledge map

2.1. Data extraction

(1) Structured data extraction

Structured data refers to data with a certain rule structure, mainly including data stored in Excel table, relational database (such as Mysql, Oracle, Microsoft Access, power system, etc.), object-oriented database (such as Db4o) \cite{9}. Usually, this type of data extraction only needs to design a simple algorithm to convert the stored data into the data form of knowledge tuple. The basic process is as follows:

1) Connect power system database;
2) Basic data initialization operation;
3) Construct SQL statement and query data;
4) Data type, structure, attribute conversion;
5) Judge whether the data exists in NEO4J database, if so, return to step 3, otherwise, store the data in step 6 (mainly judge that the information of the same node is the Labels field in NEO4J);
6) The NEO4J data storage statement is constructed, and the upper and lower relationship is determined by combining with the information extracted from the SQL statement to create the node;
7) Judge whether the query of SQL statement is completed. If the query is completed, exit the extraction program \cite{10}. If not, return to step 3 and continue to build SQL statement for data query;

(2) Unstructured data extraction

The data that can be used in the construction of knowledge mapping in vertical fields include unstructured documents and literatures within enterprises, as well as unstructured text data in the Internet. For unstructured data extraction, mainly for unstructured text document data extraction. The following main introduction is the use of natural language processing knowledge to extract data from the text. Firstly, the text is extracted by data mining technology, and then the extracted text is segmented and standardized by cleaning rules. Finally, the corresponding word segmentation, part of speech tagging and other operations are performed on the text. The pre-defined parsing rules are used to match the parsing rules. Whether the extraction is successful is judged by judging the sentence. If the extraction is successful, the corresponding knowledge is stored. Otherwise, whether the extraction is completed is judged. If the extraction is not continued. Before the extraction, the sentences are divided into separate sentences. Through the continuous extraction and retrieval of each sentence, the extraction process of unstructured data can be achieved.

2.2. Knowledge storage

After knowledge is extracted from unstructured, structured, semi-structured data, the first problem is data storage. Because after the data is written into the database, we need to consider the persistence operation, as well as the use of data retrieval scenarios, so the data must be able to effectively deal with the complex entity relationship in the constructed knowledge map. Through comprehensive consideration, this paper uses the NEO4J graph database for storage. Moreover, the storage mode of graph database has very flexible scalability, and can add and delete nodes with the change of requirements without affecting the data use of other nodes. The following summarizes the basic advantages of NEO4J as a graph database:

(1) It has high availability and supports large-scale query and data backup;
(2) The built-in graph traversal algorithms are quite rich, such as breadth first search algorithm (BFS), depth first search algorithm (DFS), minimum spanning tree routing algorithm, etc;
(3) NEO4J adopts the form of open source, free trial of community version;
(4) Enterprise Edition supports distributed graph database, and its performance is further improved;
(5) Built in Cypher, a graph data query language which is easy to learn;
(6) Without schema, it has strong flexibility and scalability.

2.3. Knowledge mapping
The nodes stored in the neo4j graph database represent the entities in the network, and the edges represent the relationships. All the data of each entity are stored and expanded through < Key, Value >, so the data storage of a node is quite rich. figure 1 is the basic structure diagram entity model of knowledge map.

![Fig.1 Basic structure of power system knowledge map](image)

3. Method flow of topology identification

3.1. Topology identification process
Based on the knowledge map, extract the building set \( B_m, B_m = \{B_{m1},...,B_{mn}\} \) of all users under the transformer \( T_m \) in a certain area, and find all user sets \( U_m, U_m = \{U_{m1},...,U_{mn}\} \) contained in \( B_m \). More than 80% of the attribute values of different power consumption address levels are the same, that is, the power user table of the same area. More than 90% of the attribute values of different power consumption address levels are the same, the power user table of the same building; More than 95% of the attribute values of different power consumption address levels are the same, that is, the power user table of the same unit in the same building; When the attribute values of power address hierarchy are exactly the same, it is the power user table of the same floor in the same unit of the same building. When it is the power user table of the same building, according to the relationship in the knowledge map, the number of transformers \( N_B \) for the building can be counted, as shown in formula (1).

\[
N_B = \begin{cases} 
1 & \text{other} \\
2 & \end{cases}
\quad (1)
\]

According to the transformer capacity \( S_{bN} \), transformer design power factor \( \cos \varphi \) and the typical load \( P_{bN} \) of power users calculated by user meter number \( N_{user} \), the number of transformers \( N_{BD} \) that meet the design specifications can be preliminarily judged, as shown in equation (2).

\[
N_{BD} = \cup \left( \frac{N_{bB} \cdot \cos \varphi}{N_{user} \cdot P_{bN}} \right) \quad (2)
\]
When $N_{RD}-N_{RBD}$=1, it means that there is an error in the household transformer relationship of the building. Comparing the other attribute values of the power user table of the same building, generally speaking, if the number of users in the middle and low rise building is small, and one transformer can meet the power demand, the power user table of the same building in the same community belongs to the same transformer power supply area; There are many users in high-rise buildings, and generally 2-4 transformers can meet the power demand, so the power user meters of the same floor in the same building in the same community belong to the same transformer. The power user table that does not conform to the design rules is selected and corrected: the transformer ID of the wrong power user table is modified to the Transformer ID with more correct number of household transformer relations in the same building or on the same floor.

4. Experiment and analysis

4.1. Basic information of the example

In order to test the effectiveness of the topology identification method based on knowledge map, this paper conducts experiments with simulated data and actual data of a power supply company. The basic situation of the example is as follows:

(1) Simulation data: a total of 3 stations, 6 transformers, 7 feeders and several power user tables are simulated, and the number of entities is about 400. Part of the user table information is set as error information to test the effect of knowledge map topology identification method.

(2) Data of power grid company: using the real data from GIS, marketing and metering information system of a power supply company, three typical stations are selected, including 432 power user tables. Because it is impossible to make field investigation on the correctness of the existing power grid data, it is considered that the existing information data is correct. The objective of topology identification method based on knowledge map is to adjust the users in one station area to other stations.

In this paper, we use the specification rate and the success rate of topology identification to evaluate the effectiveness of the topology identification method based on knowledge map. The calculation method of the two indexes is shown in formula (3) and formula (4).

\[
S_{\text{standard rate of electricity address}} = \frac{\text{Number of standardized power consumption addresses}}{\text{Total number of power address}} \times 100\% \quad (3)
\]

\[
\text{Success rate of topology identification} = \frac{\text{Number of identification error messages}}{\text{Total number of error messages}} \times 100\% \quad (4)
\]

4.2. Example analysis of simulation data

According to the above-mentioned knowledge map construction method, the knowledge map of power system topology structure is constructed, which includes 387 entity nodes, 1935 attribute nodes and 2814 relationships. Then, the power user meter address in the knowledge map is normalized, and the results are shown in table 1.

| Situation | The number of address normalization | The number of non-address normalization | Standardization rate of power consumption address |
|-----------|-----------------------------------|----------------------------------------|------------------------------------------------|
| Before address normalization | 36 | 342 | 9.52% |
| After address normalization | 378 | 0 | 100% |

The standardized address classification is stored in the knowledge map as a new attribute of the power user table. Then the topology identification analysis of the target transformer or substation area is carried out, and the identification results are shown in table 2.
Tab.2 Topology identification results (simulation)

| Number of floors | Households | Set the number of error messages | Knowledge map identification number | Identification success rate |
|------------------|------------|----------------------------------|-------------------------------------|----------------------------|
| 6                | 54         | 8                                | 8                                   | 100%                       |
| 12               | 108        | 17                               | 17                                  | 100%                       |
| 24               | 216        | 40                               | 40                                  | 100%                       |

It can be seen from table 2 that both the specification rate and the success rate of topology identification are 100%. In the case of only 9.5% of the standard power address, through the unified semantic analysis of the power users' power address in the whole station area, the standard rate of power address can be increased to 100%. Through the knowledge map method, all the wrong power user tables in different situations are identified, which has high accuracy and good identification effect.

4.3. Example analysis of actual data of power grid company

The experimental results of Section 3.2 simulation data are good. The power system characteristics of the simulation data are outstanding and complete, and there are no interference factors. The actual situation in the power system is more complex and changeable, so we use the actual data to further test the practicability and adaptability of the knowledge map topology identification method.

In fact, the power address input is random and the normalization is worse. The normalization results of power address are shown in Table 3. Invalid address means that there is no electricity address attribute in the electricity user table, which can only be supplemented according to the language relationship between the electricity addresses of the whole building.

Tab.3 Results of electricity address Standardization (actual data)

| Situation | The number of address normalization | The number of non address normalization | The number of invalid addresses | Standardization rate of power consumption address |
|-----------|-------------------------------------|----------------------------------------|--------------------------------|-----------------------------------------------|
| Before address normalization | 21 | 404 | 7 | 4.86% |
| After address normalization | 429 | 0 | 3 | 99.31% |

As can be seen from table 3, in practice, nonstandard addresses can be transformed into standard addresses by using knowledge mapping and semantic analysis, while invalid addresses can only be filled by the absence of user number in each user address of the whole building, unit or community, so address normalization cannot reach 100%.

The results of topology identification are shown in table 4.

Tab.4 Topology identification results (actual data)

| Name of station area | Households | Set the number of error messages | Knowledge map identification number | Identification success rate |
|----------------------|------------|----------------------------------|-------------------------------------|----------------------------|
| A                    | 27         | 4                                | 2                                   | 50%                        |
| B                    | 101        | 14                               | 14                                  | 100%                       |
| C                    | 128        | 17                               | 16                                  | 94.12%                     |
| D                    | 176        | 21                               | 21                                  | 100%                       |

There are four actual stations. When the number of power user tables is small, the success rate of identification is low. The reason is that the success rate of identification is relatively low due to the small amount of data and the relatively small degree of correlation between relationships; In addition, the power address cannot be 100% standardized, which is also the reason why the success rate of topology identification in station a and station C does not reach 100%. 
5. Concluding
In this paper, the application of knowledge mapping in power system is proposed. The examples of simulation data and actual data show that this method has high normalization rate in power address normalization, low cost in topology identification, small amount of data, no occupation of communication channel, and high accuracy in typical high-rise scene, which can improve the data quality in low-voltage distribution network information system. It can be widely used in most urban residents. It has high practicability for typical high-rise buildings. In the future research, we need to regularly update and optimize the latest technology of power system, the basic development specification of power system, the noun specification of power system, and so on. Aiming at the constructed knowledge map, we need to further optimize the constructed knowledge map by using the knowledge map related technologies such as entity fusion and entity link.

References
[1] Jayawardene I, Venayagamoorthy G K, Zhong X. (2020) Resilient and Sustainable Tie-Line Bias Control for a Power System in Uncertain Environments[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, PP(99):1-15.
[2] Jayantha W M, Oladinrin O T. (2019) Knowledge mapping of office workspace: a scientometric review of studies[J]. Facilities, 38(3/4):316-345.
[3] Qtaish A, Alshammari M T. (2020) A Narrative Review of Storing and Querying XML Documents Using Relational Database[J]. Journal of Information & Knowledge Management (JIKM), 18(04): 1-28.
[4] Ashajyothi P, Bhavitha V, Balakumar H, et al. (2020) Power Quality Improvement in 3-Phase Power System Using Artificial Neural Network Based Hybrid Filter[J]. Journal of Computational and Theoretical Nanoscience, 17(6):396-401.
[5] Cass S. (2019) Taking AI to the edge: Google's TPU now comes in a maker-friendly package[J]. IEEE Spectrum, 56(5):16-17.
[6] Das K, Guo F, E Nu, et al. (2020) Frequency Stability of Power System with Large Share of Wind Power Under Storm Conditions[J]. Journal of Modern Power Systems and Clean Energy, 8(2):219-228.
[7] Gao Z P, Zhao Y, Yu Y L, et al. (2020) Topology identification method of low voltage distribution network based on knowledge map [J]. Protection and control of power system, 048 (002): 34-43.
[8] Yu J M, Wang X H, Zhang Y, et al. (2020) Construction and application of knowledge mapping for intelligent regulation and control [J]. Power system protection and control, 048 (003): 29-35.
[9] Olech U P, Spytkowski M, Kwanieka H, et al. (2021) Hierarchical data generator based on tree-structured stick breaking process for benchmarking clustering methods[J]. Information Sciences, 554:99-119.
[10] S Ontaón. (2020) An Overview of Distance and Similarity Functions for Structured Data[J]. Artificial Intelligence Review, 53(11):5309-5351.