Research on the Semantic Composite Network Recognition Method of Power Distribution DTU Automation Acceptance Virtual Seat System

Hongwei Liu¹a, Yuwei He²b, Jian Jiang³c, Xingye Liu⁴d, Haoyang Cui⁵e*
¹,²,³,⁴State Grid Zhejiang Electric Power Co, Ltd. Hangzhou Power Supply Company, Hangzhou, China
⁵Shanghai University of Electric Power, School of electronic and information engineering, Shanghai, China
a(liuhw20212021@163.com, b(yvwei_he@126.com, c(jiangjianhangzhou@163.com, 
d(liuxxingye2021@163.com, e(cuihy@shiep.edu.cn

Abstract—Aiming at semantic inaccurate recognition problem caused by the lack of professional vocabulary in the existing corpus, a composite network recognition method for the semantics of the power distribution DTU acceptance virtual agent system is proposed. This method is used to extract and merge the entity, attribute and relationship information of power equipment from multiple data sources to construct a power distribution DTU acceptance knowledge graph. In addition, the graph database is the main body and the SQL database is the extension, and the corpus is stored in a distributed manner. The acceptance test results show that the overall recognition rate of the system platform for the instructions issued by the acceptance personnel in the complex environment is 92.8%, which reduces the acceptance time by more than 70%.

1. INTRODUCTION
DTU (Data Transfer Unit) is the key equipment of the distribution network, and its automated acceptance efficiency determines the delivery cycle of the distribution network upgrade [1]. At present, the automatic acceptance of distribution terminals is still generally based on the mode of communication and cooperation among field operation, maintenance personnel and the staff of the main station to test the trigger signal of electronic control. Statistics show that the acceptance of column opening and 6-interval ring network, etc., takes 20 and 90 minutes, respectively, and opening and closing the cell change acceptance takes longer [2]. In the stage of extensive construction of intelligent distribution network, the manual intervention method of the existing distribution terminal acceptance has been difficult to meet the speed requirements of rapid distribution network construction. Intelligent and efficient configuration and automatic acceptance of DTU construction has become a key technical problem that needs to be solved urgently in the development of intelligent distribution networks. In view of the current situation that the installation process of DTU is still generally used by operation and maintenance personnel, it is more important to realize the intellectualization of acceptance links that the main station acceptance is intellectualized. Through intelligent extraction of communication semantic information and automatic comparison with DTU telemetry information, the construction quality and efficiency assessment can be effectively completed [3].
The intelligent virtual agent technology, which integrates intelligent speech recognition, text recognition, and multi-round human-machine dialogue, has the advantages of repetitive tasks and automatic execution of massive transactions, and is widely used in customer service, call centers, sales transactions and other fields. If the potential of virtual agent technology can be fully tapped to form an automated acceptance mode for virtual agents as shown in Figure 1, to assist on-site personnel, it will greatly promote the application and verification of data-driven concepts in the construction of intelligent terminals in distribution networks. Scholars in domestic and foreign fields have begun to use virtual agent technology in the field of electric power customer service. For example, literature [4]–[6] proposes to build a special map for electric power communication services for power communication network management, operation inspection, risk monitoring, business scheduling, and knowledge accumulation. Reasonable semantic extraction network and corpus storage structure have great influence on the accuracy and speed of speech recognition. In reference [9], hidden Markov model is used to segment the existing power equipment defect records, and cosine similarity of discriminant vector is used to eliminate co-referential words, which improves the accuracy of semantic retrieval. In reference [10], semantic enhancement method based on topic comparison is used to reduce the redundancy of entity recognition results of LSTM network and improve the speed of speech recognition of virtual power customer service system.

The above research provides important theoretical support and reference for the construction of data transfer unit acceptance knowledge graph. However, current research on knowledge graphs in the power field mainly focuses on the construction of customer service and corpus and virtual customer service system, while the special corpus and its construction method for DTU acceptance of distribution network have rarely been reported. Based on this, this paper proposes a semantic composite network identification method for distribution DTU automatic acceptance virtual agent system according to the command joint triggering scenario and requirements of power DTU automatic acceptance business. The purpose of this paper is to identify and extract the professional words in the operation procedures, alarm information and historical operation logs of distribution DTU automatic acceptance business, construct the special corpus of distribution DTU acceptance, and form the operation mode of semantic search and joint trigger based on virtual agent.

Figure 1. Virtual agent assistant acceptance mode
2. MODEL BUILDING

In the process of implementing DTU automatic acceptance business, the first problem to be solved by virtual agent system is to build a knowledge graph of distribution terminal automatic acceptance domain (as shown in Figure 2.) which includes all acceptance professional vocabulary and has higher retrieval efficiency, so as to improve the accuracy and recognition speed of key words in acceptance speech by virtual agent system.

In this paper, the composite network and ESIM model are used to complete the recognition, extraction and fusion of professional words from multiple data sources. The distributed database model is used to store the corpus for the nodes with different entity correlation degrees. While improving the recognition rate of key words, the redundancy of nodes in graph database is reduced, and the retrieval speed is improved.

2.1. Compound network knowledge extraction model

The objects of knowledge extraction are the semi-formatted and non-formatted sentences in the distribution network regulations, historical alarms, and operation logs. The difficulty of extraction is that it is difficult to align the data by using graph mapping to obtain the relationship between power equipment from the circuit diagram topology (semi-formatted data) [11]. And the other difficulty of extraction is that when using traditional machine learning methods such as conditional random fields, support vector machines, and decision trees extracts information from texts (unformatted data) such as distribution network regulations and historical operation information, due to the high degree of discretization of differences part in co-referential vocabulary and the sparse distribution of differences part in co-referential vocabulary in the NLP data set, the accuracy and coverage of knowledge extraction are difficult to meet the requirements of engineering applications [12].

The deep neural network has powerful vector distance analysis capabilities, which can effectively identify the same vocabulary in unformatted text. However, there are obvious differences in the extraction effects of different parts of speech texts on various networks. The CNN network determines the length of the sliding extraction window according to the number of convolution kernels, which can effectively extract device attribute information with a compact input structure and a certain length. The RNN network composed of the information backward linear conduction structure has a better effect on the information extraction of the acceptance business entity with greater uncertainty in the input length. The BiLSTM network introduces a direction control factor to extract the relationship information that has a strong correlation with the context [13]. The F1 scores of these three neural network sub-modules on the NLP general data sets of three different websites are 91.7%, 99%, and 87%. The accuracy and coverage of keyword extraction from text data sources are relatively high.

The core idea of knowledge extraction is using deep neural networks to format the unformatted data in the text in order to distinguish the entities, relationships, attributes and other knowledge elements in the unformatted data [11]. This part is divided into three sub-tasks including named entity recognition, attribute extraction and relationship extraction. The three sub-task modules are implemented by CNN, RNN and BiLSTM.

The data obtained by knowledge extraction contains a lot of repetitive information. For example, 3U0, ABC-phase zero-sequence voltage, and three-phase zero-sequence voltage refer to the same objects, and distribution DTU and distribution DTU devices have the same meaning. The repetitive co-referencing text has a greater impact on the uniqueness of the keyword reference in the automatically generated acceptance report and the standardization of the report. In this paper, the relationship extens-
ion calculation is used to achieve common reference resolution. The filtered data is not stored in the map, which can be stored in a relational database and then linked with entities in the knowledge graph [14]–[17]. The following takes BiLSTM network output result processing as an example to show the process of coreference resolution. The processing methods of CNN and RNN network output results are similar.

Using ESIM co-referential resolution model

\[
\alpha_i = \omega_i \cdot FFNN_p(x^*)
\]

\[
\alpha_{ij} = \frac{\exp(\alpha_i)}{\sum_{k=\text{START}}^{\text{END}}(i) \exp(\alpha_k)}
\]

\[
\hat{x}_i = \sum_{k=\text{START}}^{\text{END}}(i) \alpha_{ij} \cdot x_j
\]

\(x^*=[h_{1,1}, h_{1,-1}]\) is the output of the BiLSTM network. 1 and -1 represent the direction in which the LSTM recognizes the context. \(\hat{x}_i\) represents the attention coefficient, which is the accumulation of the weights of the co-referential words in a sequence span \(i\). The ESIM co-referential resolution model is shown in Figure 3.

After the weighted summation of the output \(x^*\) of the BiLSTM network, the sequence integrated into the attention mechanism can be obtained, denoted by \(\hat{x}\). The sequence is comprehensively evaluated on its accuracy, recall, and harmonic F value through a three-layer scoring structure. The sequence with the highest core is considered to be the target sequence that needs to be stored in the map. The data source information is classified and modularized according to the length of the input text and the contextual relevance. The repeated information in the extraction results is resolved by the ESIM model, which improves the accuracy and coverage of keyword extraction, and eliminates the repetitive common reference. The influence of the text on the standardization of the acceptance report saves the time cost of node traversal in the graph database and improves the keyword retrieval speed. The data source information is classified and mod-
ularized according to the length of the input text and the contextual relevance.

2.2. Distributed storage model of joint related database and relational database

The power equipment of the distribution network is deeply coupled and has a high degree of data correlation. When the SQL-type database is used to store the target sequence obtained from knowledge extraction, it is necessary to establish a complex relationship between different rows of tables and multiple tables to characterize the coupling relationship between multiple power devices. The index relationship of the Entity Relations is difficult to construct, and the retrieval time is long, which cannot meet the requirements of real-time extraction of keywords in the on-site acceptance speech.

Graph database is a kind of NoSQL database, which is different from the storage form of two-dimensional tables in relational database. The data in graph database is stored in the form of node-relation (N-R) graph [18]–[20]. The topological relationship among the power equipment participating in the joint debugging is accustomed to be represented by circuit diagrams in engineering. It has a natural similarity with the data storage form of graph database. This can effectively reduce the complexity of relationship representation. An instruction in the automatic acceptance of a power distribution DTU usually contains multiple interrelated entities. Compared with relational databases that store data in tabular form, native graph databases represented by Neo4j have built-in more efficient traversal algorithms. When retrieving a group of highly related entity nodes, there is no need to establish links between multiple data tables. Just specify the starting node and the traversal direction. The time and complexity of data retrieval operations are greatly reduced. Therefore, it is more convenient and quicker to "add, delete, modify, and check" data.

However, after repeated access tests on the data with different attribute richness in the two databases, it is found that for entity classes with rich entity attribute information and some frequently updated data (such as events, logs, fault records, etc.), use Relational database storage can update the attribute information of an entity multiple times without traversing all associated entities in the database. For example, a group of entities has a relevance of 10, and each entity contains data of 3 attributes as experimental samples. One attribute of any entity is updated 100 times. The result shows that the data is updated when the same memory is occupied. The speed is increased by about 25%.

In order to ensure the real-time performance of keyword retrieval during on-site speech recognition, and take into account the data update speed when the corpus is used for case recording, reference, reasoning and other in-depth applications, a distributed storage technology of joint graph database and relational database is proposed. The pivot nodes selected by the extended calculation of entities and relationships with an entity relevance greater than 10 can be found in the knowledge graph by creating a formatted index label in the attribute column of the pivot node of the graph database to point it to the corresponding table in the relational database. Link to data. The distributed storage structure with entity association as the extension calculation trigger condition simplifies the relationship construction process. While retaining all the results of knowledge extraction, it solves the problem of low voice recognition efficiency caused by high coupling of equipment and improves the data update speed.
2.3. Semantic search model of aggregation engine based on sentence pattern matching and knowledge graph
The situation of the distribution DTU automation acceptance site is complex, and the question and answer sentences are changeable. Therefore, this paper proposes a dual-engine semantic search technology based on sentence matching and the distribution DTU automation acceptance knowledge graph. First, through text analysis, clarify the relationship between the node to be identified and the object to be searched. For "Hello, I am XXX", "Dial XXX", "Correct, please start", the more fixed command responses are based on ESIM. The model is matched with the sentence patterns in the predefined corpus, and the predetermined operation in the template library is executed when the results are consistent. For a long sentence scenario that contains multiple equipment operating parameters, the knowledge graph is directly called to form the answer. In the process of virtual agent-assisted acceptance of DTU, the search method of aggregation engine reduces the return time of speech recognition results and improves the efficiency of acceptance.

3. EXPERIMENTAL RESULTS AND ANALYSIS
Based on the Distributed DTU automation acceptance business scenario, with the help of the Open5200 distribution automation management platform of a State Grid electric power company, the IVCA system is used to test the accuracy of the semantic recognition of the distribution DTU acceptance corpus constructed by this method and the rapidity of the virtual agent-assisted acceptance mode.

3.1. Accuracy and analysis
In an environment of 40dB background noise, using a sampling frequency of 11.025KHz, 16-bit quantized numbers, and mono wav format audio, the IVCA system was subjected to an acceptance rule increase test. The acceptance rule is a data set composed of four voice data sets: Mandarin data, accent Mandarin data, Mandarin natural conversation data and real network customer service voice data. The result is shown in Figure 4.

![Figure 4. Multi-language acceptance rules plus test results](image)

When only using Mandarin speaking data for training, the word error rate of the speech recognition model is 28.42%, after superimposing the accent in Putonghua, Putonghua natural conversation, and live online customer service voice in the data hall, the word error rate WER was reduced to 7.2%. The overall system recognition accuracy rate ACR can be expressed as

\[ ACR = 1 - WER\% \] (4)

The word error rate WER is defined as follows

\[ WER = 100 \cdot \frac{S + D + I}{N} \% \] (5)

Among them, \( S, D, I \) are the number of words added, replaced, and deleted in the recognition result, and \( N \) is the total number of words. Applying aggregate AI scheduling engine technology driven by knowledge, the overall recognition rate ACR (calculated according to the word error rate) of the IVCA system is 92.8%. In the multi-bay ring network DTU acceptance service, the working hours of the main station operation and maintenance personnel are reduced from h-level to m-level, which can improve the
problem of low efficiency of traditional mode acceptance under the background of massive power DTU access in the process of distribution network upgrade and transformation.

3.2. Rapidity and analysis

3.2.1 Rapidity when the acceptance result is normal: On September 15, 2020, the acceptance personnel used the IVCA system to conduct an automated acceptance testing on the distribution DTUs of 10 switch stations (the green part in Figure 5) between the two main power distribution stations of Fuqian and Xixing. Take the acceptance process of DTU (equipment name new1Tq1) as an example, The on-site acceptance personnel measured the opening and closing status of interval 1 switch (equipment name 1Tqkg1) and interval 2 switch (equipment name 1Tqkg2). After the three-phase current of the network bus is distributed, the IVCA is input by voice. After confirming that the voice recognition result is correct, the system will report it to the power distribution master station. The system automatically generates an acceptance report as shown in Figure 6. The aggregate AI confirms that the return value in the acceptance report is consistent with the acceptance rules (the remote signal is correct, and the remote measurement accuracy is within the operating range). After that, make a prompt for the DTU acceptance of the next switch station.

After the system is put into operation, under the condition that the on-site collection and telemetry results are consistent, the work time of the main station operation and maintenance personnel the acceptance of a single switch station DTU is reduced from the original 20 minutes to 0 minutes, which greatly improves the automatic acceptance of the power distribution DTU efficiency, significantly eases the work pressure of the main station operation and maintenance personnel.
the voltage of the DTU battery of the switch station and other information with the on-site acceptance personnel. Then modify the terminal debugging plan and execute it in accordance with the processing flow of the DTU terminal data abnormality. During this process, the operation and maintenance staff of the main station reduced the working time for acceptance from the original 20 minutes to 5 minutes. Alleviate the problem of a large gap in the joint commissioning of distribution network automation effectively. In addition, the handling of power distribution DTU acceptance abnormalities is more refined and standardized.

4. CONCLUSION
This paper proposes a composite network recognition method based on the semantics of the virtual agent system for automatic acceptance of power distribution DTU. For professional vocabulary information in unformatted texts such as distribution network regulations, a composite network extraction and distributed storage method is adopted to construct a special corpus for the field of power distribution DTU acceptance. The virtual agent assists the field staff in the power distribution DTU acceptance test, and the recognition rate of the field staff's acceptance speech in the complex system environment is extremely high, reducing the main station operation and maintenance staff's acceptance work time by more than 70%. The joint debugging test results show that the method can significantly improve the efficiency of the automatic acceptance of distribution DTUs. During the upgrading and transforming of the distribution network, the method also can effectively cope with the problems of the access of a large number of newly added power distribution terminals, the cumbersome maintenance of old terminals which leads to insufficient use of the traditional acceptance mode, and the increase of the shortage of personnel in the automation and joint debugging of the distribution network.

The research on the construction method of special knowledge graph for distribution network automation management is still in its infancy, and effective mining methods for historical operation logs, alarm information, fault records and other internal laws have not yet been formed. Manual intervention is still necessary to change the acceptance plan after the acceptance report is abnormal. Therefore, how to fully mine the inherent laws contained in the data source and guide the adjustment of the acceptance plan will be the future research direction.

ACKNOWLEDGMENT
This study was supported by the Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd(5211HZ19014R). The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

REFERENCES
[1] R. Zeng, Y. Xu and D. Wu, "Normalization Performance Evaluation Method for Distribution Automation Terminal," 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada, 2019, pp. 1833-1838, doi: 10.1109/ISIE.2019.8781296.
[2] Loken R S J, Apostolov A. Testing of Protection in a Digital Substation Protection Automation and Control System (SPACS)[J]. 2020.
[3] Song X, Yuan H, Xiaoxuan Z, et al. Evaluation Model for World-Class Distribution Network Management[C]//2018 China International Conference on Electricity Distribution (CICED). IEEE, 2018: 59-62.
[4] Yang H, Luo L, Chueng L P, et al. Deep Learning and Its Applications to Natural Language Processing.
[5] Shen G , Wang W , Mu Q , et al. Data-Driven Cybersecurity Knowledge Graph Construction for Industrial Control System Security[J]. Wireless Communications and Mobile Computing, 2020, 2020(6):1-13.
[6] ZHANG C, Zhiyong Y U, Zhen Z, et al. User image construction framework based on big data fusion background for power grid business travel[J]. Telecommunications Science, 35(12): 148.
[7] Tang, Xisheng (Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing; 100190, China) Source: Dianli Xitong Zidonghua/Automation of Electric Power Systems, v44, n1, p23-28, January 10, 2020

[8] Chen, Zhengping (State Grid Fujian Electric Power Dispatching and Control Center, Fuzhou; 350001, China); Mi, Weimin; Lin, Jinghuai; Wang, Heng; Wang, Hao; Dong, Genyuan Source: Dianli Xitong Zidonghua/Automation of Electric Power Systems, v43, n22, p173-178 and 186, November 25, 2019

[9] Liu, Ziquan (College of Electrical Engineering, Zhejiang University, Hangzhou; 310027, China); Wang, Huifang Source: Dianli Xitong Zidonghua/Automation of Electric Power Systems, v42, n14, p158-164, July 25, 2018

[10] Huang H, Hong Z, Zhou H, et al. Knowledge Graph Construction and Application of Power Grid Equipment[J]. Mathematical Problems in Engineering, 2020.

[11] Wang J, Wang X, Ma C, et al. A survey on the development status and application prospects of knowledge graph in smart grids[J]. IET Generation, Transmission & Distribution, 2021, 15(3): 383-407.

[12] Qiao, Ji (China Electric Power Research Institute, Haidian District, Beijing; 100192, China); Wang, Xinying; Min, Rui; Bai, Shuhua; Yao, Dong; Pu, Tianjiao Source: Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering, v40, n18, p5837-5848, September 20, 2020

[13] Li Z , Li Q , Zou X , et al. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings[J]. Neurocomputing, 2021, 423:207-219.

[14] Li Z , Ding X , Liu T . Constructing Narrative Event Evolutionary Graph for Script Event Prediction[J]. 2018.

[15] Sun, Hongbin (Department of Electrical Engineering, Tsinghua University, Haidian District, Beijing; 100084, China); Huang, Tianen; Guo, Qinglai; Zhang, Boming; Guo, Wenxin; Liu, Wentao; Xu, Taishan; Xu, Tao Source: Dianwang Jishu/Power System Technology, v44, n1, p1-8, January 5, 2020

[16] V Yawahare H R , Karde P P , Thakare V M . A Hybrid Database Approach Using Graph and Relational Database[C]/ 2018:1-4.

[17] Zhao, M.; Wang, H.; Guo, J.; Liu, D.; Xie, C.; Liu, Q.; Cheng, Z. Construction of an Industrial Knowledge Graph for Unstructured Chinese Text Learning. Appl. Sci. 2019, 9, 2720.

[18] Gao Y , Li Y F , Lin Y , et al. Deep Learning on Knowledge Graph for Recommender System: A Survey[J]. arXiv, 2020.

[19] Wang Y, Zhang K, Dai Q, et al. Construction and Application of Knowledge Graph in Full-service Unified Data Center of Electric Power System[J]. 2018.

[20] Wikidata as a knowledge graph for the life sciences[J]. eLife Sciences, 2020, 9.