INTELLIGENT QUESTION-ANSWERING SYSTEM FOR GRID SCHEDULING USING KNOWLEDGE GRAPH AND LARGE LANGUAGE MODEL

Wei LI[1,*], Bo ZHOU[2], Siqi SHEN[3], Zhangguo CHEN[4]

With the increasing complexity of power system operations, scheduling tasks require more advanced knowledge acquisition and question-answering systems. Existing power grid scheduling question-answering systems have issues such as complex operations and incomplete management frameworks. To address these problems, this study proposes an intelligent question-answering system for power grid scheduling based on a Knowledge Graph and a Large Language Model. By constructing a Knowledge Graph specific to power dispatching and integrating multimodal data extraction and graph fusion technologies, the system enables comprehensive modeling and dynamic updates of scheduling knowledge. A multi-turn question-answering and context memory mechanism is designed based on Large Language Model technology to improve the accuracy and naturalness of responses. Experimental results show that the proposed system maintains an accuracy rate of over 90% across different test batches, with a recall rate reaching up to 88.62%. In comparison, the highest accuracy rates of the public information model combined with the Knowledge Graph question-answering system and the deep learning question-answering system are 86.37% and 75.49%, while the Knowledge Graph question-answering system alone achieves only 73.84%. These results indicate that the proposed system significantly enhances the efficiency of scheduling knowledge retrieval and improves the professionalism of intelligent question-answering, providing reliable decision-making support for power grid regulation.

Keywords: Knowledge graph; Large language model; Power grid scheduling; Intelligent question-answering system; Multi-turn question-answering

1. Introduction

Power grid security is the lifeline of the power industry, and the intelligent question-answering system for power dispatching plays a crucial role in ensuring grid security. It has brought revolutionary changes to the field of power dispatching

[1] Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan, Guangdong, 528000, China, e-mail: wei2508000@163.com

[2] Nanjing NARI Information and Communication Technology Co., Ltd., Nanjing, Jiangsu, 211100, e-mail: China

[3] Nanjing NARI Information and Communication Technology Co., Ltd., Nanjing, Jiangsu, 211100, China

[4] Nanjing NARI Information and Communication Technology Co., Ltd., Nanjing, Jiangsu, 211100, China

[1]. With the development of big data and artificial intelligence, the application of cloud computing, the Internet of Things, and big data has enabled power grid scheduling systems to achieve deep automation and extensive intelligence [2]. Current power grid scheduling intelligent question-answering systems can monitor grid conditions in real time and promptly identify potential security risks. However, they may still encounter comprehension errors when handling complex, ambiguous, or multi-meaning queries [3]. Moreover, the performance of an intelligent question-answering system heavily depends on the quantity and quality of training data, making it susceptible to inaccuracies when data reliability is compromised [4]. The Knowledge Graph (KG) can integrate structured and unstructured data from different sources into a unified knowledge system, modeling real-world complex relationships through nodes and edges [5]. Additionally, the Large Language Model (LLM) leverages extensive computational and hardware resources to process complex and large-scale datasets, demonstrating strong feature extraction and representation capabilities. Therefore, this study integrates KG and LLM while incorporating a multi-turn question-answering and context memory mechanism to develop a novel power grid scheduling intelligent question-answering system. The proposed system aims to improve power dispatching efficiency and enhance the intelligent development of the power industry while ensuring the stable and secure operation of the power grid.

The research has achieved three key innovations, as follows. Technological integration innovation: In previous studies, LLM and KG were applied separately to different sub-tasks of the power system, without forming a collaborative mechanism. The research deeply integrated the two, constructing a "KG structured knowledge support + LLM semantic understanding and reasoning" dual-core architecture, addressing the shortcomings of insufficient knowledge accuracy and weak semantic understanding of a single technology in the question-answering system. Function optimization innovation: The previously proposed application paradigm of large models did not involve multi-round conversations and context memory functions and could not adapt to the continuous interaction requirements in the scheduling scenario. The research designed a "fixed window + intelligent truncation + incremental update" context memory mechanism, combined with multi-round question-answering logic, enabling the system to have continuous and coherent interaction capabilities. Scenario focus innovation: Previous studies focused on evaluation, planning, data integration, etc., and did not form a specialized solution for the core scenario of "intelligent question-answering for power grid scheduling". The research focused on the query characteristics of the scheduling scenario, optimized the modeling of KG entity relationships and the domain fine-tuning strategy of LLM, making the system more in line with the actual usage needs of scheduling personnel. The main contribution of this research lies in proposing an intelligent question-answering framework for power grid scheduling that integrates Knowledge Graph (KG) and Language Model (LLM). This framework addresses the shortcomings of existing systems in complex semantic understanding and dynamic knowledge update, enabling comprehensive modeling and flexible application of scheduling knowledge. Additionally, a multi-round question-answering mechanism based on LLM and context memory mechanism has been designed, which improves the accuracy, naturalness, and logical coherence of system responses, and enhances the human-computer interaction experience.

This study consists of four sections. The first section reviews the domestic and international research on KG, LLM, and power grid scheduling question-answering systems. The second section focuses on optimizing KG using LLM and constructing the intelligent question-answering system for power grid scheduling. The third section presents the performance evaluation of the proposed system. Finally, the fourth section provides the overall conclusions of this study.

2. Related works

As a popular technology in the field of artificial intelligence, KG continuously refines knowledge logic and models through learning functions such as reasoning, annotation, and error correction, enhancing system intelligence. Many researchers have conducted extensive studies on its advantages. Mao et al. proposed a novel fault diagnosis model based on KG for traditional machine fault root cause analysis. They used a network stacking method to capture semantic information, and the experimental results showed that this model achieved superior fault diagnosis performance [6]. Shen et al. introduced a cohort network approach to analyze the impact of environmental exposure on disease risk. They visualized the connections between exposure and outcomes using KG, and simulation experiments confirmed the effectiveness of this method [7]. Xiang et al. aimed to integrate platforms in the Internet of Things and proposed an improved fusion technique to analyze its features. They extracted specific vocabulary based on KG, and the results indicated that this approach provided valuable data references for translation system development [8]. Moreover, LLM have demonstrated remarkable potential not only in natural language processing tasks but also in various interdisciplinary applications. Hu et al. proposed an intelligent auxiliary evaluation method based on LLM for problems such as low evaluation efficiency of distribution network projects. The results show that this method can quickly and accurately identify common problems in the technical and economic dimensions of distribution network projects [9] Yao et al. proposed a large model application method in response to the intelligent demands of power systems. The results show that this method significantly enhances the efficiency and accuracy of perception, planning, and control tasks in power systems through cross-domain pre-training and feast-shot transfer, providing a new paradigm for the development of smart grids [10].

On the other hand, to ensure the secure and stable operation of power grids, power grid dispatching has gradually become a crucial guarantee for high-quality and cost-effective grid operation. Some researchers have explored this topic. Zhang et al. proposed a quantitative representation method to address the shortcomings of existing power grid dispatching models. They validated the approach by identifying fault equipment and types, and simulation experiments confirmed its applicability [11]. Fei et al. introduced a novel two-layer optimization strategy to address deficiencies in microgrid dispatching, using randomly generated initial populations. The experimental results demonstrated the excellent optimization efficiency of this method [12]. Rong et al developed an innovative optimization dispatching model to enhance the power system's resilience against extreme weather conditions. They transformed the model into a linear problem for solving, and the results verified the effectiveness of their approach [13]. Li et al. proposed a novel distributed coordination method to overcome the limitations of traditional unit regulation capabilities. They optimized power dispatching strategies to enhance frequency regulation, and simulation experiments validated the superiority of this method [14]. Kadir et al. focused on operational decision-making for transmission systems under extreme events. They applied an integrated testing platform to solve the problem, and the results showed that this approach effectively reduced power flow on disconnected lines and mitigated load interruptions [15]. Lin C et al. proposed a hybrid model combining pre-trained language models and structured models to ensure the safe and stable operation of the power system. Experimental results showed that this model performed well in diagnosing defects of power grid equipment [16]. Zhou X's team addressed the challenges of on-site safety management and information complexity in power grid construction and proposed a method for constructing multimodal knowledge graphs based on large language models. The results indicated that this method could quickly query safety incidents and provide efficient tools for managers [17].

From the research conducted by scholars worldwide, it is evident that while power grid dispatching has achieved significant progress, challenges such as high integration complexity and operational management difficulties remain. To address these issues, this study proposes an intelligent power grid dispatching question-answering system that integrates KG and LLM. By incorporating multi-turn question-answering and contextual memory mechanisms, the system is further optimized. This study aims to enhance the professionalism and coordination capabilities of power grid dispatching question-answering systems, providing reliable data support for power system operations.

3. Construction of the KG-LLM-based power grid dispatching question-answering system

3.1. Design of the KG-based power grid dispatching question-answering system

KG integrates and unifies data from different sources, preventing data loss between different systems or departments within an organization. Regardless of the original format and source of the data, KG ensures that data is always transformed and standardized into a common framework that is understandable and usable for all users [18]. This study applies KG to the power grid dispatching question-answering system, providing structured knowledge in the power grid dispatching domain, which enhances the system's interpretability during queries. Relation extraction in KG-based question answering mainly includes three parts: entity recognition, relation classification, and relation extraction. Entity recognition, as the foundation of relation extraction, identifies entities with specific meanings from text. These entities form relational triples, typically including names of people, places, and organizations. The calculation formula for the accident response knowledge base is shown in Equation (1).

$$O_{KG} = \{T(E), T(F), T(R)\} \tag{1}$$

In Equation (1), $T(E)$ represents the set of entities with consistent attributes and similar functions in the field of power grid dispatching, $T(F)$ represents a set of parameters that share the same attribute characteristics, $T(R)$ represents a set of logically consistent semantic-related relationships. By integrating KG, the power grid dispatching question-answering system can analyze the power grid operation status in greater depth, enabling real-time monitoring, prediction, and optimization. The system framework is shown in Fig. 1.

As shown in Fig. 1, the user interface serves as a bridge for interaction between users and the system, receiving users' text and voice inputs and converting them into a processable format. The frontend question-answering generation module quickly generates answers based on user input, keyword extraction, retrieval, and question answering. After data enters the backend offline data processing module, it undergoes parsing, structuring, and indexing to support fast responses from the frontend. Among these, the frontend question-answering generation module plays a key role in user interaction. The power grid dispatching question-answering system dynamically processes user queries through the frontend and backend processing modules and utilizes the existing knowledge base to provide accurate answers. The calculation formula for the range characteristics in the power grid is shown in Equation (2).

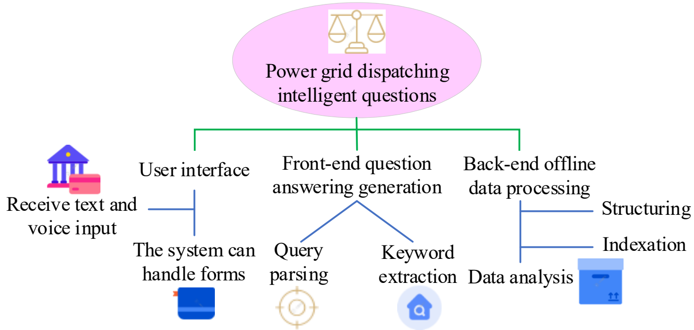


Fig. 1. Intelligent question-answering framework for power grid dispatching

$$Feature1 = \max\left[x(t)\right] - \min\left[x(t)\right] \tag{2}$$

In Equation (2), $x(t)$ represents the time series, $\max(\bullet)$ indicates the maximum value, and $\min(\bullet)$ represents the minimum value.The calculation for node current is shown in Equation (3).

$$\dot{I}_N = \frac{\dot{S}_i}{\dot{V}_i} \tag{3}$$

In Equation (3), S_i represents the injected power at the node, and V_i represents the voltage of the node. KG optimizes the fault-handling process, ensuring that users receive quick and accurate responses when encountering problems. The KG-based power grid dispatching question-answering system is shown in Fig. 2.

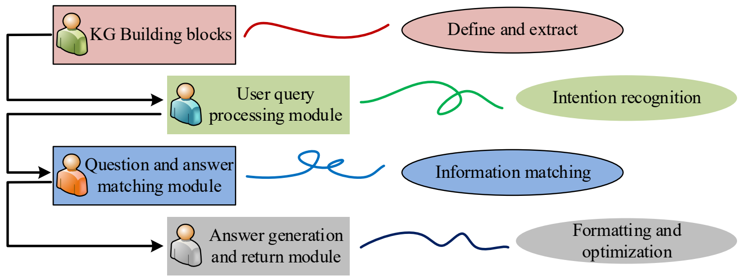


Fig. 2. KG-based intelligent question-answering system for power grid dispatching

As shown in Fig. 2, the KG-based power grid dispatching question-answering system integrates specialized knowledge of power grid dispatching into a structured knowledge graph and uses natural language processing to understand and answer user queries intelligently. The KG construction module defines and extracts entities, relations, and attributes, while the user query processing module handles user queries, including preprocessing and intent recognition. The system matches the queries with the information in the power grid dispatching knowledge graph to obtain the most relevant answers, which are then formatted and optimized before being returned to the user. The node calculation formula in the power grid dispatching KG is shown in Equation (4).

$$P = \left\{ \mathrm{Pr}\,oject, Trans_f, Teach_{HV} \right\} \tag{4}$$

In Equation (4), $\mathrm{Pr}\,oject$ represents a specific power grid project, $Teach_{HV}$ refers to the transmission technology used in the power grid application, and $Trans_f$ indicates the equipment involved in the power grid construction. The calculation formula for the parsed KG data is shown in Equation (5).

$$C_j = f(\sum b_s W_{sj} + W_j) \tag{5}$$

In Equation (5), b_s represents the original data of KG, W_{sj} is the bias weight parameter, W_j represents the weight parameter in the knowledge graph, and $f(*)$ represents the parsing function in the knowledge graph. On one hand, the retrieval-based KG uses the structured KG as the core retrieval foundation. The graph stores structured knowledge such as entities, relationships, and attributes in the field of power grid dispatching, supporting precise entity association queries and logical reasoning. On the other hand, non-structured text data is integrated as a supplementary retrieval source, including the original text of power grid dispatching regulations, historical fault handling reports, operation logs, etc. These data are preprocessed and then mapped and associated with the structured knowledge in the graph. When receiving user queries, the system first parses the query intent through LLM and extracts keywords, and triggers dual retrieval. It first matches directly relevant entities and relationships from the structured knowledge graph to quickly obtain accurate answers. If there is no direct matching result in the graph, it retrieves the associated non-structured text data and extracts key information from the text using LLM. It then combines the existing knowledge in the graph for logical integration to generate a complete answer. The deployment of the knowledge graph follows a four-step process: "data preprocessing - graph construction - storage deployment - dynamic update". The data sources include grid dispatching regulations texts, equipment parameter manuals, historical dispatching logs, and fault handling reports. In the preprocessing stage, the Python NLTK toolkit is used for text segmentation and stop word removal, and regular expressions are used to extract core entities such as equipment names and fault types. Structured data is standardized using SQL statements for field formatting. The graph is constructed using Neo4j as the graph database, adopting a "entity - relationship - attribute" tripartite modeling scheme. Entities are batch-labeled using the spaCy named entity recognition model, and relationships are extracted based on the domain rule library. Finally, the graph data in RDF format is generated. The Neo4j database is deployed on a Windows Server 2019 server, with 16GB of memory and 512GB of SSD storage. The APOC plugin is enabled to support batch data import and optimization of graph query. Common query templates are written in Cypher language and encapsulated as API interfaces for the front-end question-answering module to call. The dynamic update mechanism synchronizes newly added dispatching data through scheduled tasks. The Diff algorithm is used to compare the differences between the new and old data, and the graph's entity attributes and relationships are automatically updated. For manually corrected knowledge, a web-based visual editing interface is provided, allowing manual addition or deletion of triples.

3.2. Development of the LLM-integrated power grid dispatching question-answering system

Although applying KG to the power grid dispatching question-answering system supports complex queries and data analysis, the effectiveness of KG highly depends on data quality and accuracy. Meanwhile, LLM learns more generalized knowledge during training and has stronger feature representation and pattern recognition capabilities across large datasets [19]. Especially when applied to the power grid dispatching question-answering system, LLM quickly extracts key information from vast amounts of data and provides timely answers [20]. The comparison of core technical requirements for LLM is shown in Fig. 3.

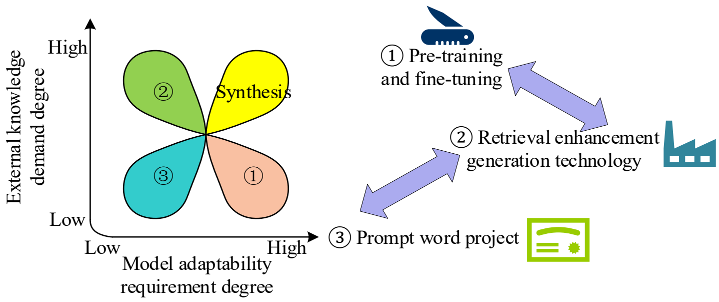


Fig. 3. Comparison of requirements for three technologies in LLM

As shown in Fig. 3, when using pretraining and fine-tuning techniques, the model is trained on large-scale text data to learn language knowledge and semantic information, while fine-tuning with a small amount of task-specific data. When the model has low adaptation capability requirements but high external knowledge requirements, retrieval-augmented generation enhances the model’s generation capability by leveraging relevant data sources. When both model adaptation capability and external knowledge requirements are high, a combination of these three techniques is required. LLM excels in logical reasoning by learning complex relationships between entities and performing logical inference. Based on this, the improvements of the KG-based power grid dispatching question-answering system with LLM integration are shown in Fig. 4. As shown in Fig. 4, the integration of LLM with KG enhances KG’s limitations in dynamic information updates and complex semantic understanding, transforming static information into dynamic and interactive knowledge to improve system intelligence.

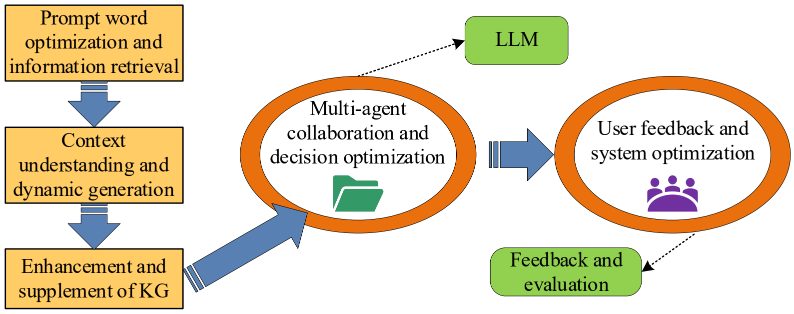


Fig. 4. Schematic diagram of LLM improvement mechanism

In multi-agent collaborative systems, LLM functions as the central control unit, coordinating decision-making and scheduling among agents. By optimizing resource allocation and context switching, it ensures efficient collaboration between agents while continuously improving system performance based on user feedback. The calculation formula for the association strength between different KG nodes is shown in Equation (6)[21].

$$Y = \left\{ X_i \middle| g_i^{hp}(C_j) \right\} \tag{6}$$

In Equation (6), X_i represents the KG feature parameters, and $g_i^{hp}(C_j)$ is the direction feature of the relationships between different KG nodes. Context memory is stored in a conversation history stack, with the maximum capacity set to 5 rounds. The text length limit for a single round of conversation is 512 tokens. If the text length of a single round of query exceeds 512 tokens, a "tail truncation + core semantic retention" strategy is adopted. The core information of the query is identified through a keyword extraction algorithm. If the total length of the 5 rounds of conversation history exceeds the maximum input limit of the model, the earliest round is truncated according to the first-in-first-out principle. The update strategy adopts a "real-time incremental update + timeout failure" mechanism. After each new interaction is completed, the current round of user query and system response are automatically appended to the conversation history stack, and the earliest round exceeding the 5-round limit is deleted. The conversation history storage uses key-value pair format. The session timeout time is set to 30 minutes. If there is no new interaction within 30 minutes, the context memory of the session will be automatically cleared. If the user explicitly mentions instructions such as "reset query" or "ignore previous content" in the conversation, the system will immediately clear the current conversation history stack and start a new context memory cycle. Each time a new query is made, the processed historical conversation is concatenated with the current query in the format "[historical conversation 1]→[historical conversation 2]→…→[current query]", and the attention mask mechanism is used to highlight the weight of the current query. This study integrates LLM and KG to construct a power grid dispatching question-answering system and incorporates multi-turn dialogue and contextual memory mechanisms. The structure is shown in Fig. 5.

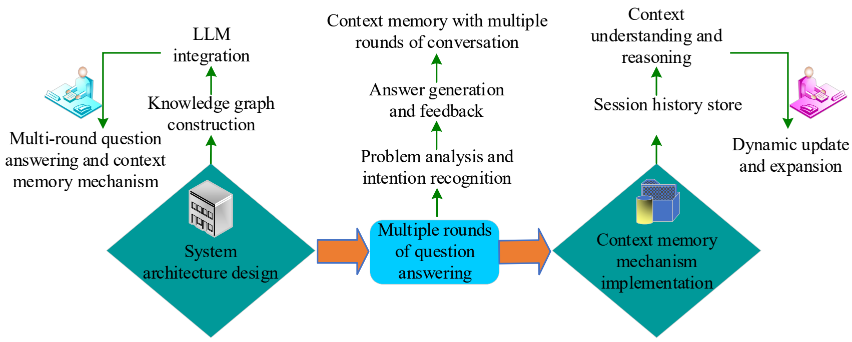


Fig. 5. The KG-LLM-based power grid dispatching question-answering system

As shown in Fig. 5, the system is designed with multi-turn dialogue processes and a contextual memory mechanism, allowing users to engage in continuous and logical conversations while recording users' past queries and responses to provide context for future answers. During multi-turn dialogues, the system optimizes responses based on conversation history. The implementation of the contextual mechanism relies on a well-structured storage design, while LLM's contextual understanding ability analyzes relationships within conversation history to generate more relevant responses. The calculation formula for the document-word joint probability distribution in the question-answering process is shown in Equation (7)[22].

$$P(d_i, w_j) = P(d_i)\sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{7}$$

In Equation (7), $\sum_{k=1}^{K} P(w_j|z_k)$ and $P(z_k|d_i)$ represent the probability of word w_j appearing in topic z_k and the probability of topic z_k appearing in document d_i, respectively. The calculation formula for word vectors is shown in Equation (8).

$$\mathrm{d}' = \begin{bmatrix} b_{00} & b_{01} & \cdots & b_{0n} \\ b_{11} & b_{12} & \cdots & b_{1n} \\ \vdots & & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \tag{8}$$

In Equation (8), m, b, and n represent the text vector length, bias term, and embedding vector length, respectively. The calculation formula for entity power knowledge weights is shown in Equation (9).

$$M = \frac{1}{\lg(1+E|\varepsilon|)} \tag{9}$$

In Equation (9), $E|\varepsilon|$ represents the number of entity power knowledge entries in the power grid dispatching question-answering system. The calculation formula for the similarity between power knowledge entities is shown in Equation (10).

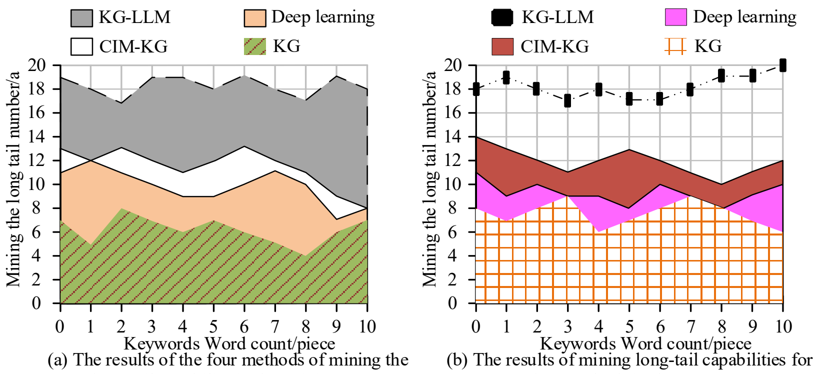
$$sim(a,b) = \frac{\sum_{\varepsilon \in (\phi(a) \cap \phi(b))} \frac{1}{\lg(1+E|\varepsilon|)}}{\sqrt{|\phi(a)||\phi(b)|}} \tag{10}$$

In Equation (10), a , b , and $\phi(a)$ represent different power knowledge entities and the knowledge similarity function, while $\phi(b)$ and E represent transformation constraint functions and the entity set of similarity-based multi-source power grid data, respectively. The deployment of the large language model focuses on "model selection - fine-tuning training - deployment optimization - context mechanism implementation". The open-source LLaMA-2-7B model is selected as the base model. The deployment environment is Ubuntu 22.04 operating system, with hardware configuration of 2 NVIDIA A100 GPUs (80GB memory per GPU) and 64GB memory. It relies on the PyTorch 2.0 framework and uses the Transformers library to load the model. Through FlashAttention, the efficiency of attention calculation is optimized. The training data uses the question-answer dataset in the power grid dispatching field, with the data format uniformly being (question, context, answer). The fine-tuning strategy adopts LoRA (Low-Rank Adaptation), freezing the parameters of the main model, and only training the low-rank matrix (rank 8). The learning rate is set to 2e-4, the batch size is 8, and the training rounds are 3. The model inference interface is encapsulated by FastAPI, supporting HTTP request calls. The response time for a single query is controlled within 500ms. Redis is used to cache frequently accessed query results, with a cache validity period of 1 hour, improving the concurrent processing capability. For long text queries, a sliding window truncation strategy is adopted to retain the core semantic information. Context memory is stored through a "dialogue history stack", with a stack capacity of 5 rounds. Each new query is concatenated with the historical dialogue and the current query in the format of "[historical dialogue 1]→[historical dialogue 2]→[current query]" and input to the model. At the same time, the attention masking mechanism highlights the weight of the current query. The dialogue history storage is in key-value format, with a session timeout of 30 minutes. The study selected the open-source LLaMA-2-7B as the base model. During deployment, the majority of parameters of the main network of the model were frozen. Only the low-rank matrix of the model was fine-tuned for training. The rank of the low-rank matrix was set to 8. Moreover, the low-rank matrix was fine-tuned using a domain-specific dataset from the power grid dispatching field, enabling the model to learn knowledge in areas such as power system terminology, dispatching rules, and fault handling logic, while retaining its original natural language understanding and logical reasoning capabilities.

4. Performance of the KG-LLM power grid dispatching question-answering system

4.1. Performance comparison of the KG-LLM question-answering system

To validate the performance of the proposed intelligent power grid scheduling question-answering system, the study compared the KG-LLM-based power grid intelligent scheduling question-answering system with the KG-based system, the Common Information Model and Knowledge Graph (CIM-KG), and the deep learning-based power grid dispatch question-answering system. The experiment was conducted in a Windows 10 Professional (64-bit) operating system, with an AMD Ryzen 7 5800H processor equipped with Radeon Graphics, 16GB of memory, a 512GB SSD hard drive, and an RTX 3060 graphics card. The database and development languages used were Neo4j and Python, respectively. The experimental dataset consisted of the power grid historical data and the power grid operation and scheduling technical question-answering dataset provided by the L2RPN competition hosted by a French grid company. The historical data from the L2RPN competition served as the training dataset, which included voltage fluctuations, topology changes, and load power variations during grid operation. The official dataset acquisition website for the L2RPN competition is: https://l2rpn.chalearn.org/. The power grid operation and scheduling technical question-answering dataset served as the test set, providing answers to questions about grid knowledge and AC resonance. The study first compared and analyzed the ability of the four question-answering systems to mine long-tail problems in both the training and test datasets, as shown in Fig. 6.



(a) The results of the four methods of mining the long-tail capability in the training set
(b) The results of mining long-tail capabilities for four methods in the test set

Fig. 6. Results of mining long-tail capabilities of four question-answering systems

As shown in Fig. 6(a), when the number of keywords is three, the KG-LLM question-answering system mined the most long-tail problems, with a total of 19. The CIM-KG, deep learning, and KG question-answering systems mined 12, 10, and 7 long-tail problems, respectively. In Fig. 6(b), the long-tail problems mined by the research question-answering system ranged from 17 to 20, which was similar to the results in the training set. Among the comparison systems, the CIM-KG question-answering system mined a maximum of 14 long-tail problems, while the deep learning and KG question-answering systems mined a maximum of 11 and 9, respectively. This shows that the research question-answering system was able to mine more low-frequency but important long-tail problems from the vast amount of data. The study then analyzed the F1 value of the four question-answering systems in the training and test sets, as shown in Fig. 7.

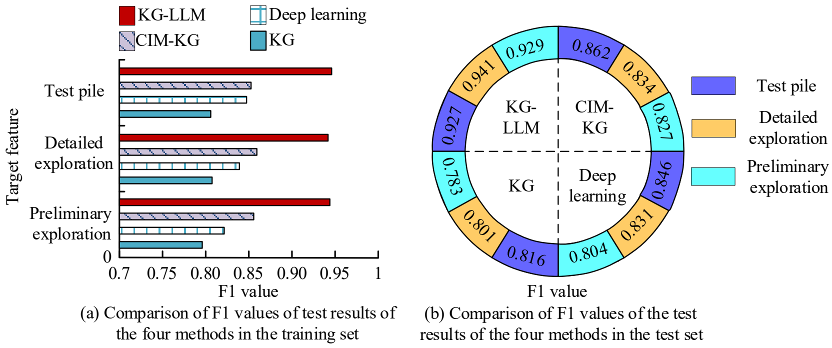
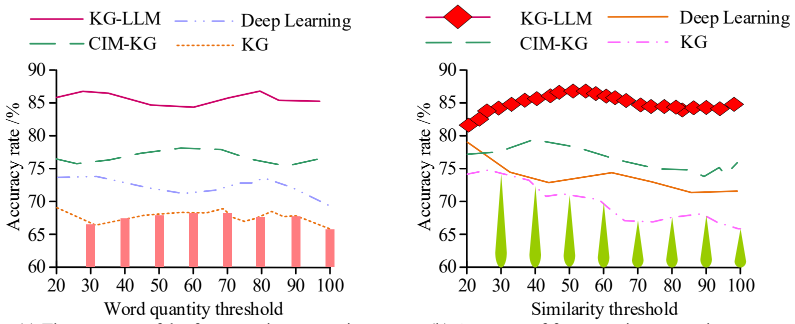


Fig. 7. F1 value test results of four question answering systems

As shown in Fig. 7(a), in the training set, when the target feature was the test pile, the research question-answering system had an F1 value of 0.941. The CIM-KG and deep learning question-answering systems had F1 values of 0.856 and 0.842, respectively, while the KG question-answering system had the lowest F1 value, at only 0.809. When the target feature was preliminary investigation, the F1 value of the research question-answering system ranged from 0.90 to 0.95, which was higher than the F1 values of the comparison systems. In Fig. 7(b), when the target feature was detailed investigation, the research question-answering system had an F1 value of 0.941, while the KG question-answering system had the lowest F1 value, at 0.801. When the target feature changed to the test pile, the KG-LLM and deep learning question-answering systems had F1 values of 0.927 and 0.846, respectively. These results indicate that the research question-answering system demonstrated high stability in its F1 values and excellent feature extraction capabilities.

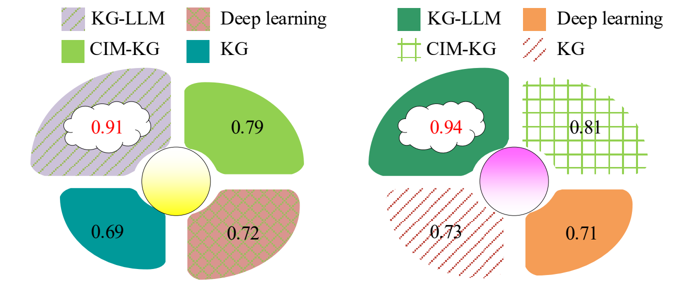
4.2. Practical application of the KG-LLM power grid dispatching question-answering system

To further validate the effectiveness of the proposed power grid scheduling intelligent question-answering system, the study selected the power infrastructure project corpus from Changzhou, Jiangsu Province, for the power grid scheduling question-answering task. After experimental preprocessing, the study created a training dataset of 3022 entries. First, the study compared the accuracy of the four question-answering systems under different word quantity thresholds and similarity thresholds in the power grid scheduling question-answering task, as shown in Fig. 8.

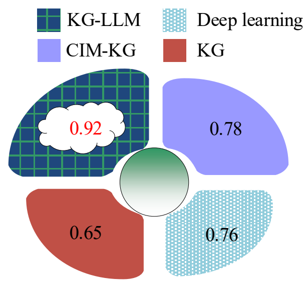


(a) The accuracy of the four question answering systems under different word quantity thresholds
(b) Accuracy of four question answering systems under different similarity thresholds

Fig. 8. Accuracy results under different word quantity thresholds and similarity thresholds



(a) Grade precision of different methods in question 1
(b) Grade precision of different methods in question 2



(c) Grade precision of different methods in question 3

Fig. 9. Grade precision results in different problems

As shown in Fig. 8(a), the KG-LLM question-answering system consistently maintained an accuracy above 80% under different word quantity thresholds, while the comparison systems had lower accuracy than the research question-answering system. In Fig. 8(b), when the similarity threshold was set to 50, the research question-answering system and the CIM-KG question-answering system achieved accuracy rates of 87.64% and 78.13%, respectively. The accuracy of the deep learning and KG question-answering systems was lower compared to the KG-LLM question-answering system. Based on the above results, it can be concluded that the proposed question-answering system had a high accuracy rate in power grid scheduling, and its accuracy remained stable when compared to the other systems. Next, the study expanded and classified different power grid scheduling problems, setting equipment overload, frequency deviation, and load variation as Problem 1, Problem 2, and Problem 3, respectively. The accuracy and reliability of the question-answering systems when answering different types of questions were evaluated based on grade precision, where higher grade precision indicated better accuracy and reliability of the system. The grade precision of the four question-answering systems for different problems is shown in Fig. 9.

From Fig. 9, it can be seen that the KG-LLM question-answering system achieved the highest grade precision in all problems. For Problem 1, the grade precision of the KG-LLM question-answering system reached 0.91, while the KG question-answering system had a grade precision below 0.70. For Problem 2, the grade precision of the CIM-KG and KG question-answering systems increased to 0.81 and 0.73, respectively. The grade precision of the KG-LLM question-answering system remained above 0.90 for Problem 3. These results show that the KG-LLM question-answering system was able to retrieve the most relevant information from a wider knowledge base. Finally, the study selected 300 question-answering data entries as test data and conducted three rounds of testing. The first column of the table represents three independent test datasets or test scenarios. Each batch uses the same experimental environment, dataset size, and evaluation metrics. The purpose is to verify the performance stability of the system under different data distributions or test scenarios. The settings of batches 1-3 eliminate the randomness of a single experiment and ensure the reliability and universality of the results through multiple repeated tests. Obtaining the system test results for the different question-answering systems as shown in Table 1.

Table 1

System test results of different question and answer systems

Experimental batch	Question-answering system	Accuracy (%)	Recall (%)	F1 value
1	KG-LLM	**90.51**	**86.24**	**89.35**
	CIM-KG	80.67	75.31	78.36
	Deep learning	77.35	72.16	75.91
	KG	70.96	71.35	69.58
2	KG-LLM	**91.26**	**87.54**	**88.63**

	CIM-KG	76.58	76.82	76.42
	Deep learning	76.93	73.41	74.63
	KG	71.44	70.09	70.59
3	KG-LLM	**90.03**	**88.62**	**87.54**
	CIM-KG	77.19	74.96	75.98
	Deep learning	77.05	71.68	75.14
	KG	71.96	70.28	70.08

From Table 1, it can be seen that the KG-LLM question-answering system maintained an accuracy of over 90% in different experimental batches. Its recall rate reached a maximum of 88.62%, and the F1 value reached a maximum of 89.35. The accuracy of the CIM-KG question-answering system was lower than that of the KG-LLM system, with a maximum recall rate of 76.82% and a maximum F1 value of 78.36. The accuracy of the deep learning and KG question-answering systems in the second experimental batch was 76.93% and 71.44%, respectively, much lower than that of the KG-LLM system. These results show that the KG-LLM question-answering system provided more accurate answers, was capable of handling complex problem scenarios, and offered more comprehensive solutions. It achieved a good balance between accuracy and knowledge coverage, meeting both precision and recall requirements.

5. Conclusions

The study addressed the current limitations in real-time monitoring and data analysis in power grid dispatch question-answering systems by proposing a method that combines KG and LLM to build an intelligent question-answering system. This system retrieves relevant information from the KG based on the user's intent and uses the LLM to reason with the retrieved knowledge. Experimental results showed that when the word count threshold was set to 40, the accuracy of the KG-LLM question-answering system was 85.78%. In comparison, the CIM-KG question-answering system and deep learning-based question-answering system achieved accuracies of 77.59% and 72.83%, respectively, while the KG question-answering system only reached 67.74%. When the target feature changed to trial pile testing, the F1 values of the KG-LLM and deep learning question-answering systems were 0.927 and 0.846, respectively. The KG-LLM question-answering system maintained a classification accuracy above 0.90 for problem 3, while the KG question-answering system showed a classification accuracy below 0.70, performing significantly worse than the proposed research system. Overall, the intelligent question-answering system proposed by the research demonstrated superior accuracy in understanding user query intentions and addressing more long-tail issues. The system was also highly efficient in retrieving and matching information from large-scale knowledge bases, effectively improving the efficiency of power grid dispatch. The research focuses on the core demands of intelligent grid dispatching, knowledge services, and human-machine collaboration. The research content aligns with the core themes of "Promoting Digital and Intelligent Technological Innovation and Engineering Applications in Power Systems" in mainstream journals such as "IEEE Transactions on Power Systems" and "Journal of Smart Grid", and the proposed technical solutions like the KG-LLM integration architecture and context-aware question-answering mechanism can provide new technical references for intelligent grid dispatching and human-machine interaction decision-making in the power system field. They have clear journal theme adaptability and engineering application value. However, the F1 values of the research question-answering system gradually decreased across different experimental batches. Therefore, further optimization of the proposed system will be carried out in future work.

Acknowledgments

This work was supported by the Research on Knowledge Service and Human-Machine Efficient Collaboration Technology for Main Distribution Network Scheduling Driven by Knowledge Graph and Large Model Integration, with the project number GDKJXM20231034 (030600KC23090015).

REFERENCES

[1] Li J, Li Y, Liu H. Short-term electric power and energy balance optimization scheduling based on low-carbon bilateral demand response mechanism from multiple perspectives. IET Generation, Transmission & Distribution, 2024, 18(24): 4245-4258. https://doi.org/10.1049/gtd2.13231.

[2] Ruan J, Liang G, Zhao H, Liu G, Sun X, Qiu J. Applying large language models to power systems: Potential security threats. IEEE transactions on smart grid, 2024, 15(3): 3333-3336. https://doi.org/10.1109/TSG.2024.3373256

[3] Li Y, Chen J, Luo X, Zheng H. Intelligent human-computer interactive training assistant system for rail systems. High-speed Railway, 2025, 3(1): 64-77. https://doi.org/10.1016/j.hspr.2025.02.001

[4] Shen Y, Shao J, Zhang X, Lin Z, Pan H, Li D. Large language models empowered autonomous edge AI for connected intelligence. IEEE Communications Magazine, 2024, 62(10): 140-146. https://doi.org/10.1109/MCOM.001.2300550

[5] Zhou X J, Xu D Z, Huang Y R. Harmonic Interaction Analysis Between Multi-Inverter System and Power Grid Based on Equivalent Single-Inverter Model. Journal of Electrical Engineering & Technology, 2023, 18(1): 193-203. https://doi.org/10.1007/s42835-022-01203-2

[6] Mao Z, Wang H, Jiang B, Xu J, Guo H. Graph Convolutional Neural Network for Intelligent Fault Diagnosis of Machines via Knowledge Graph. IEEE transactions on industrial informatics, 2024, 20(5): 7862-7870. https://doi.org/10.1109/TII.2024.3367010

[7] Shen Y, Kioumourtzoglou M A, Wu H, Vokonas P, Spiro III A, NavasAcien A, Baccarelli A A, Gao F. Cohort Network: A Knowledge Graph toward Data Dissemination and Knowledge-Driven Discovery for Cohort Studies. Environmental Science And Technology, 2023, 57(22): 8236-8244. https://doi.org/10.1021/acs.est.2c08174

[8] Xiang Y, Chen Y, Ye F H. Enhancing computer-aided translation system with BiLSTM and convolutional neural network using a knowledge graph approach. Journal of supercomputing,

2024, 80(5): 5847-5869. https://doi.org/10.1007/s11227-023-05686-2
[9] Hu W, Zheng T, Du P, Chen Z. An Intelligent Assisted Assessment Method for Distribution Grid Engineering Quality Based on Large Language Modelling and Ensemble Learning. Protection and Control of Modern Power Systems, 2025, 10(3): 83-97. https://doi.org/10.23919/PCMP.2024.000347
[10] Yao Q, Fang F, Chen Y, Liu J, Mo H, Ao Y. AI Large Models for Power System: A Survey and Outlook. IET Smart Energy Systems, 2025, 1(1): 3-21. https://doi.org/10.1049/ses2.70000
[11] Zhang X, Du M, Wang Y, Zhang H, Guo Y. Research on Power Grid Fault Diagnosis Based on a Quantitative Representation of Alarm Information. IEEE Transactions on Industrial Electronics, 2023, 70(9): 9582-9592. https://doi.org/10.1109/TIE.2022.3213893
[12] Fei L, Guangsen G, Jianhua Z, Lu W, Hengdao G. Double-layer optimal microgrid dispatching with price response using multi-point improved gray wolf intelligent algorithm. Electrical engineering, 2024, 106(3): 2923-2935. https://doi.org/10.1007/s00202-023-02108-7
[13] Rong J R, Zhou M, Zhang Z, Li G Y. Coordination of preventive and emergency dispatch in renewable energy integrated power systems under extreme weather. IET renewable power generation, 2024, 18(7): 1164-1176. https://doi.org/10.1049/rpg2.12893
[14] Li Z, Bai N, Cheng Z, Wang Y, Si J. Adaptive Chaotic Gray Wolf Optimizer-Based Optimization of Decentralized AGC and Power Dispatching Controllers for Integrated Energy System with Heterogeneous Power Sources. Journal of Electrical Engineering & Technology, 2024, 19(2): 1097-1111. https://doi.org/10.1007/s42835-023-01668-9
[15] Kadir S U, Majumder S, Anurag K. SrivastavaAjay Dev ChhokraHimanshu NeemaAbhishek DubeyAron Laszka. Reinforcement-Learning-Based Proactive Control for Enabling Power Grid Resilience to Wildfire. IEEE transactions on industrial informatics, 2024, 20(1): 795-805. https://doi.org/10.1109/TII.2023.3263500
[16] Lin C, Zheng Z, Cai S, Fu L, Xie W, Ma T, Zhang Z. Knowledge graph completion for power grid main equipment using pretrained language models[C]//International Conference on Intelligent Computing. Singapore: Springer Nature Singapore, 2023: 828-838. https://doi.org/10.1007/978-981-99-4752-2_68
[17] Zhou X, Shi J, Dong L, Zhang Y, Pan J, Huang H. Construction of a Multimodal Knowledge Graph for Power Grid Construction Safety Based on Large Language Models[C]//2024 International Conference on New Power System and Power Electronics (NPSPE). IEEE, 2024: 21-28. https://doi.org/10.1109/npspe62515.2024.00013
[18] Pathak G, Singh B, Panigrahi B K. Utility Interactive Renewable Microgrid with Backup Power Functionality. IETE journal of research, 2024, 70(4): 4333-4344. https://doi.org/10.1080/03772063.2023.2194282
[19] Wang S, Zou X, Zhu W, Zeng L. Insulator Defects Detection for Aerial Photography of the Power Grid Using You Only Look Once Algorithm. Journal of Electrical Engineering & Technology, 2023, 18(4): 3287-3300. https://doi.org/10.1007/s42835-023-01385-3
[20] Shi H, Fang L, Chen X, Gu C, Ma K, Zhang X, Zhang Z, Gu J, Lim E G. Review of the opportunities and challenges to accelerate mass-scale application of smart grids with large-language models. IET Smart Grid, 2024, 7(6): 737-759. https://doi.org/10.1049/stg2.12191
[21] Liyanage C, Nutkani I, Meegahapola L, Jalili M. Stability Enhancement of Power Synchronisation Control Based Grid-Forming Inverter Under Varying Network Characteristics. IEEE transactions on sustainable energy, 2024, 15(3): 1799-1813. https://doi.org/10.1109/TSTE.2024.3379548
[22] Zhang Z, Qin B, Gao X, Ding T. CNN-LSTM based power grid voltage stability emergency control coordination strategy. IET generation, transmission & distribution, 2023, 17(16): 3559-3570. https://doi.org/10.1049/gtd2.12890