Event Extraction and Analysis for Power Media Based on Large Language Models and Knowledge Graphs: Postprint

Authors: Li Jia, Zang Yanjiao, Gu Chenlan

Date: 2025-07-09T00:00:00+00:00

Abstract

[Objective] Amidst the deepening reform of China' s electric power system and the industry' s vigorous development, electric power media serves as a critical communication bridge between the electric power sector and the public. It is currently confronted with two core challenges: the exponential growth of electric power news events and the efficient management of information, necessitating the urgent development of effective coping strategies.

[Methods] This paper aims to explore and propose an innovative solution to address these challenges. Based on comprehensive consideration of existing resources and technological advancements, we have creatively designed an event extraction and analysis methodology for electric power media that integrates large language models and knowledge graph technologies.

[Results] This methodology enables precise identification of key entities within the electric power domain and deep exploration of the intricate relationships among these entities, thereby constructing a knowledge graph for the electric power media field.

Conclusion By leveraging the network relationship structure presented by the knowledge graph and the powerful semantic understanding capabilities of large language models, this approach achieves dual improvements in both information retrieval efficiency and interactive experience intuitiveness. The methodology significantly enhances users' capabilities for information acquisition and utilization, providing robust technical support for the intelligent development of the electric power media industry.

Full Text

Event Extraction and Analysis for Electric Power Media Based on Large Language Models and Knowledge Graphs

Li Jia, Zang Yanjiao, Gu Chenlan
Yingda Media Investment Group Co., Ltd., Beijing 100052

Abstract

[**Objective**] Against the backdrop of China' s deepening power system reform and the industry' s vigorous development, electric power media, as a critical communication bridge between the power sector and the public, faces two core challenges: a surge in the volume of power news events and the need for efficient information management, urgently requiring effective coping strategies. [**Method**] This paper aims to explore and propose an innovative solution to address these challenges. Based on an in-depth consideration of current resources and technological advancements, this paper creatively designs an event extraction and analysis method for electric power media that integrates large language models and knowledge graph technologies. [**Result**] This method can accurately identify key entities in the power domain and deeply mine the intricate relationships among these entities, thereby constructing a knowledge graph for the electric power media field. **Conclusion** Through the network structure presented by the knowledge graph and the powerful semantic understanding capabilities of large language models, the method achieves dual improvements in information retrieval efficiency and intuitive interactive experience. It significantly enhances users' ability to acquire and utilize information, providing robust technical support for the intelligent development of the electric power media industry.

Keywords: Electric Power Media; Large Language Models; Knowledge Graphs; Entity Extraction; Relation Extraction
Classification Code: G224
Document Code: A
Article ID: 1671-0134(2025)03-141-05
DOI: 10.19483/j.cnki.11-4653/n.2025.03.031
Citation Format: Li Jia, Zang Yanjiao, Gu Chenlan. Event Extraction and Analysis for Electric Power Media Based on Large Language Models and Knowledge Graphs[J]. China Media Technology, 2025, 32(3): 141-145.

Introduction

In recent years, with the deepening of China's power system reform and the prosperity of the industry, electric power media has become an important bridge for communication between the power industry and the outside world, with power events forming the core of power information. Against a backdrop of information diversification and increasingly high public demands for information quality, traditional media resources, characterized by lengthy articles and difficult-to-capture key points, can no longer adapt to the fragmented reading habits of the internet environment. Electric power media urgently needs to deeply analyze and precisely manage content, particularly the identification and integration of power event data, to improve content creation efficiency and quality while optimizing user interaction experiences.

To address these challenges, this paper proposes an event extraction and analysis strategy for electric power media based on large language models and knowledge graph technologies. This strategy utilizes advanced artificial intelligence technologies to deeply mine and intelligently analyze electric power media information, providing users with intuitive and detailed information display and in-depth, rich power event information, thereby promoting the continuous progress of the electric power media industry.

1.1 Entity Extraction Technology

Entity extraction technology, as one of the core tasks in the field of natural language processing (NLP), aims to deeply mine and accurately extract key entity information from text data. This process relies on advanced semantic analysis, deep learning models, and complex natural language processing algorithms to extract meaningful entities from massive, unstructured text data such as electric power media information, including but not limited to abstract or concrete concepts like events, persons, organizations, equipment and facilities, and technical terms [1]. With the vigorous development of deep learning technology, neural network-based methods have demonstrated powerful advantages in entity extraction tasks, enabling more precise identification and extraction of entities. Meanwhile, the introduction of pre-trained language models (such as BERT, GPT, Qwen, etc.) has further enhanced model understanding capabilities and generalization ability, enabling entity extraction technology to exhibit excellent performance even when facing unknown or rare entities [2].

1.2 Large Language Model Technology

In recent years, deep learning technology has made remarkable progress in the field of natural language processing (NLP), with its wide range of applications and significant effects becoming an important force driving transformation in this field. Deep learning learns complex language patterns and rules through multi-layer nonlinear transformations, thereby demonstrating excellent learning efficiency and powerful generalization capabilities. These characteristics enable deep learning models to exhibit performance in handling natural language semantic analysis—a highly complex and uncertain task—that traditional machine learning algorithms cannot match. As deep learning technology continues to be explored and optimized in the NLP field, large language models (LLMs) have gradually emerged. As a special type of deep learning architecture, large language models take words or subwords as basic units and learn and understand the intrinsic relationships and contextual dependencies among words, phrases, sentences, and even entire documents through massive trainable text datasets. This learning process goes beyond surface-level lexical matching to delve into multiple levels including semantics, syntax, and even pragmatics, enabling the model to accurately capture and understand the diversity and complexity of human language, allowing it to navigate comprehensively when analyzing and processing the vast array of electric power media data [3-5].

1.3 Knowledge Graph Technology

Knowledge graph, as a highly structured and graphical knowledge representation method, abstracts and models entities in the real world (such as people, places, events, etc.) and their complex relationships in the form of nodes (representing entities) and edges (representing relationships). This model not only intuitively demonstrates the connections and interactions between entities but also deeply reveals their intrinsic relationships and logical structures, providing a powerful tool for understanding and analyzing knowledge within specific domains [6]. Furthermore, visualization technology provides an intuitive and rich interface for the display and application of knowledge graphs. By graphically presenting entities, relationships, and their attributes, users can easily browse and understand complex knowledge networks, discover patterns and trends within them, and thus make more informed decisions. This makes it an effective assistant for displaying the relationships between power events and their related entity elements [7].

2. Methodology

2.1 Process Overview

The general process of electric power media event extraction and knowledge graph construction encompasses several core components (see Figure 1 [Figure 1: see original paper]). (1) Data Preparation Phase: As the foundation of the entire process, this phase focuses on collecting, organizing, and cleaning the extensive data resources in the electric power media domain, including various file types and text formats, to ensure their accuracy, comprehensiveness, and timeliness, laying a solid foundation for subsequent analysis. (2) Entity Type Modeling: Based on the characteristics of the text data to be extracted, this step precisely defines the key entity types involved in electric power media, such as events, persons, organizations, equipment and facilities, and technical terms, adding descriptive explanations for these entity types. By constructing detailed entity type models, clear guidance is provided for subsequent entity recognition. (3) Retrieval Purpose Modeling: This step clarifies users' retrieval purposes for predefined entity types, aiming to tag material resource data so that retrieval results can efficiently focus on information points of interest to users, enhancing the relevance and practicality of material resource retrieval results. (4) Large Language Model Application: Leveraging advanced large language model technology and providing predefined entity types and retrieval purposes, this step conducts in-depth analysis of preprocessed data to achieve automated extraction of entities, relationships, and retrieval purposes. This process fully utilizes the model' s powerful capabilities in text understanding and semantic analysis, effectively improving extraction accuracy and efficiency. (5) Extraction Result Storage: This step systematically stores the structured information extracted by the large language model—such as entities and relationships—in a knowledge graph database, while storing the materials and extracted entity and retrieval purpose data in a relational database. This process not only achieves centralized management and efficient utilization of data but also provides convenience for subsequent data analysis and visualization display. (6) Result Evaluation and Process Optimization: By comparing extraction results with expected goals, this step comprehensively evaluates extraction effectiveness and process efficiency. Based on the evaluation results, the entity type models, retrieval target models, and application strategies of large language models are continuously iterated and optimized to ensure continuous progress and improvement in electric power media event extraction and knowledge graph construction. This paper will elaborate on these key steps in greater detail to facilitate a deeper understanding of their internal logic and operational details.

2.2 Entity Type Data Modeling

Entity type modeling plays a core role in the knowledge graph construction process. These entity types, known as ontologies within the traditional knowledge graph framework, are crucial for the efficient analysis of text content using large language models. To extract valuable entity information from massive text data, the precise meaning of entities must first be defined. This approach not only serves to limit the scope of extracted entities within specific application scenarios but, more importantly, helps large language models better understand entity type concepts through predefined detailed attributes, enabling them to more accurately parse and extract entities and their relationships [8][9]. Adopting this strategy requires a clear understanding of the text data to be processed. This means identifying the various forms in which different types of entities may appear in the text and the attributes and characteristics they possess. For example, if the content discusses a person, attributes such as name, age, occupation, and educational background can be defined across multiple dimensions; for locations or objects, attention should be paid to their physical characteristics, positions, historical uses, and other relevant attributes. This methodology ensures that model output is not only accurate but also comprehensive, providing a solid foundation for subsequent graph data application. In the context of electric power media information usage scenarios, we have further subdivided several entity types, including equipment, technology, organization, person, and event. Equipment entities encompass various hardware and software devices and facilities involved in the power system, such as transformers, circuit breakers, and Apps; technology entities include relevant technologies, standards, and protocols for power transmission and conversion; organization entities include power companies, organizational departments, or industry associations; person entities record individuals involved in the power industry, such as engineers, technicians, and project managers; event entities refer to any activities or engineering projects related to power supply, maintenance, or accident handling. Through such detailed entity classification and definition, a vast and complex knowledge network can be constructed to help users quickly locate information, solve practical problems, and even promote the development and application of new energy technologies.

2.3 Large Model Entity and Relationship Extraction

As described above, we have carefully planned a series of detailed entity modeling data categories to ensure the comprehensiveness and accuracy of data collection. For the actual entity extraction phase, we leverage cutting-edge large language model technology. These models not only possess the capability to accurately extract entity information from textual materials but also conduct in-depth parsing and fine-grained definition of the involved original materials. By providing the large language model with raw materials and well-defined entity type data, we require the model to respond with extracted entity information and attributes. This implementation enables the program to more clearly identify the entity type, name, and unique characteristic attributes of each entity during analysis, while also revealing the intricate network of relationships between them—including which entities possess specific resources, how these entities are interconnected, and under what contexts these connections arise. This approach significantly enhances the depth and breadth of data analysis, enabling a deeper understanding and mastery of entities and their interrelationships [10]. When selecting an appropriate large language model, comprehensive consideration must be given to the requirements of practical application scenarios, with priority given to models that demonstrate superior performance in Chinese processing. Since entity recognition and relationship extraction often heavily depend on the close semantic associations within the text context, special attention must be paid to whether the model supports sufficiently long context windows during selection. Models supporting longer contexts should be chosen to ensure the model can fully understand and extract the semantic information implicitly contained in the textual content. Conversely, if limited by short context support that results in overly fragmented text materials, information loss or misuse may occur, thereby affecting the quality of the entire analysis result. Through careful evaluation and selection of suitable large language models, we not only improve the efficiency and accuracy of entity modeling but also provide a solid foundation for subsequent data analysis [11].

2.4 Automatic Entity Merging

In the actual knowledge graph construction process, a common situation arises where the same entity may have multiple names that differ depending on context, domain, or language. For example, the company name "IBM" may be referred to as "International Business Machines Corporation" in technical documents, abbreviated as "IBM" in marketing materials, or even called "Big Blue" in informal contexts. This phenomenon is prevalent across all industries, as banks, hospitals, schools, and their related personnel also have their own abbreviations, aliases, or titles, which are important for understanding the essential meaning of entities [12]. To ensure that graph data can accurately capture and reflect the true relationships of things, these variant names and their connections with other entities must be unified. This involves not only the identification and recording of each entity name but also a comprehensive consideration of their descriptions, attributes, and interrelationships. By doing so, we can ultimately retain the most accurate and commonly used names while merging those that are inaccurate, making the entire dataset both rich and accurate [13]. When performing entity merging, large language models are needed to help filter out from massive amounts of information those different aliases that truly represent the same entity identity. The model can understand subtle differences in various scenarios, thereby revealing different names of the same entity hidden beneath the surface. This approach not only reduces errors and improves data quality but also makes the knowledge graph more accurate and practical, truly serving user needs and decision-making [14].

2.5 Retrieval Target Analysis and Marking

In the above process, we have successfully extracted entity names, entity attributes, and relationships between these entities from the material repository. This data is stored in the database. Additionally, we have conducted in-depth recording of the associations between entities and materials, including what types of retrieval purposes these materials support for specific entities, which is crucial for future data analysis and retrieval operations. To further improve the application effectiveness of knowledge graph data in retrieval scenarios, we have introduced the concept of retrieval target attributes. This attribute sets some basic retrieval targets for various entities, enabling the large language model to identify specific retrieval targets suitable for the entity based on the specific content of the current material when processing entities. For example, when dealing with an "organization" entity, the large language model will mark the most matching retrieval target from preset options such as "latest developments," "social activities," and "product releases" ; when a "technology" entity appears, targets such as "application cases," "research results," and "development trends" become the selection scope. This design greatly enhances the flexibility and adaptability of knowledge graph data, enabling it to more effectively serve different query requirements [15]. Through this approach, we not only improve the accuracy and efficiency of the retrieval system but also provide users with a richer and more personalized information retrieval experience. Whether searching for the latest organizational developments or looking for practical application cases within a specific technology domain, the system can quickly provide results that meet the requirements, enabling people to more conveniently access needed information and thus accelerate knowledge dissemination and utilization.

3. Application and Visualization

3.1 Large Language Model-Enhanced Knowledge Graph Retrieval

In addressing the challenges of diversity and complexity in user information retrieval, we are committed to developing an efficient and intelligent information retrieval system aimed at significantly improving user experience. When users enter query text, the system' s primary task is to promptly and accurately capture the user' s query intent. To achieve this goal, we have innovatively integrated a large language model as the core analysis engine to deeply parse the semantic connotation of user input. Specifically, the system first uses tokenization technology to segment user input retrieval text into lexical units, then efficiently matches these units with a pre-built entity database. During the matching process, we fully utilize the alias and abbreviation information annotated during entity merging to ensure that even if users input non-standard terms or abbreviations, they can be accurately mapped to target entities. Subsequently, the large language model is used to analyze the specific retrieval purpose for a particular entity in the current query, thereby achieving precise understanding of user needs [16]. Based on the above analysis, the system further retrieves relevant data from the well-constructed knowledge graph, strictly screening results according to both entity and retrieval purpose criteria. Finally, the system displays a series of manuscript lists to users, which are closely associated with the user' s query entity and purpose and are arranged according to the natural chronological order of entities and events on the timeline, forming an intuitive temporal clue [17]. This design not only greatly enhances the accuracy and efficiency of user retrieval but also significantly reduces the risk of retrieval failure due to input differences. Users no longer need to worry about the accuracy of terminology; they only need to follow their personal habits when inputting queries, and the system can intelligently guide them to relevant information, achieving a leap from "people searching for information" to "information finding people."

3.2 Intuitive Display of Knowledge Graph Network Relationship Diagrams

Based on in-depth data analysis, the knowledge graph data constructed in this paper is visualized as a network node knowledge graph chart. In this chart, each node serves as a concrete representation of an independent entity, like stars in the universe, each shining with rich information and value. To optimize the chart' s reading experience and operational convenience, we have introduced color-coding technology to distinguish different entity types with vivid colors, supplemented by detailed legends, enabling users to quickly capture the essence of information and understand core contexts while browsing. The vitality of a knowledge graph lies in its intricate network of relationships among entities, a feature vividly demonstrated in the chart through closely connected edges between points. These edges not only outline the association framework between entities but also deeply reveal the underlying logical structures and intrinsic connections by clearly labeling relationship types (such as "contains," "belongs to," "related to," etc.). This design significantly reduces the cognitive burden on users in understanding complex relationships and promotes the efficiency of information retrieval and utilization [Figure 6: see original paper]. The visualization display further incorporates dynamic interactive functions to enhance the depth and breadth of user experience. In the initial display stage, the system focuses on directly related nodes of the matching entity, i.e., the most closely related peripheral information, providing users with a concise and focused view. However, for users eager to explore in depth, they only need to simply click on any specific node, and the system will immediately respond by dynamically reconstructing and displaying an expanded knowledge graph centered on that node. This new version of the graph deeply mines and extensively presents the knowledge domain related to that node, providing users with a multi-dimensional and in-depth exploration platform that helps them comprehensively and profoundly understand complex information systems [18].

Conclusion

This paper has thoroughly explored an event extraction and analysis method for electric power media based on large language models and knowledge graph technologies, aiming to improve the production efficiency and quality of electric power media content while optimizing user experience. Through entity extraction, large language models, and knowledge graph technologies, we have constructed a comprehensive knowledge graph framework for electric power media. This framework can accurately extract various entities and their attributes from electric power media information and improve accuracy and efficiency through large language models. In application scenarios, the knowledge graph significantly enhances the convenience and accuracy of information retrieval, providing rich visual experiences and in-depth information exploration capabilities. However, facing the challenge of scarce GPU resources, this study has encountered bottlenecks when processing massive historical data, and the non-interpretability of large language models increases optimization difficulty. To overcome these pain points, future work intends to explore and optimize from several aspects: algorithm optimization and model compression, distributed computing and cloud computing resource utilization, intelligent resource scheduling and prediction, and exploration of emerging hardware technologies, to achieve deep mining and efficient presentation of data value while reducing dependence on GPU resources.

References

[1] Ma Zhonggui, Ni Runyu, Yu Kaihang. Recent advances, key technologies, and challenges of knowledge graphs[J]. Chinese Journal of Engineering, 2020(10): 1254-1266.

[2] Li Dongmei, Zhang Yang, Li Dongyuan, Lin Danqiong. A survey of entity relation extraction methods[J]. Journal of Computer Research and

Development, 2020(7): 1424-1448.
[3] Zhang Heyi, Wang Xin, Han Lifan, Li Zhao, Chen Zirui, Chen Zhe. Research on question answering system integrating large language models and knowledge graphs[J]. Computer Science and Exploration, 2023(10): 2377-2388.
[4] Huang Bo, Wu Shen' ao, Wang Wenguang, Yang Yong, Liu Jin, Zhang Zhenhua, Chen Nanxi, Yang Hongshan. Graph-model complementarity: A survey of knowledge graph and large model fusion[J]. Journal of Wuhan University (Natural Science Edition), 2024(4): 397-412.
[5] Tang Xiaosheng, Cheng Linya, Zhang Chunhong, et al. Application of large language models in the automated construction of discipline knowledge graphs[J]. Journal of Beijing University of Posts and Telecommunications (Social Science Edition), 2024(1): 125-136.
[6] Xu Zenglin, Sheng Yongpan, He Lirong, Wang Yafang. A survey of knowledge graph technology[J]. Journal of University of Electronic Science and Technology of China, 2016(4): 589-606.
[7] Liu Jin, Du Ning, Xu Jing, et al. Application and research of knowledge graphs in the power field[J]. Electric Power Information and Communication Technology, 2020(1): 60-66.
[8] Pu Tianjiao, Tan Yuanpeng, Peng Guozheng, Xu Huifang, Zhang Zhonghao. Construction and application of knowledge graphs in the power domain[J]. Power System Technology, 2021(6): 2080-2088.
[9] Yang Yuji, Xu Bin, Hu Jiawei, Tong Meihan, Zhang Peng, Zheng Li. An accurate and efficient domain knowledge graph construction method[J]. Journal of Software, 2018(10): 2887-2901.
[10] Zhang Ning, Simon Mahony. Opportunities and challenges of large language models for digital publishing[J]. Editing Friends, 2023(11): 45-51.
[11] Cai Zifan, Yu Haiyan. The evolution of AI-generated content (AIGC) and its application scenarios for smart library services[J]. Library Journal, 2023(4): 34-43, 135-136.
[12] Yang Bo, Sun Xiaohu, Dang Jiayi, Zhao Haiyan, Jin Zhi. A large language model-based named entity recognition method for medical question-answering systems[J]. Computer Science and Exploration, 2023(10): 2389-2402.
[13] Zhang Caike, Li Xiaolong, Zheng Sheng, et al. Research on knowledge graph construction and application based on large language models[J]. Computer Science and Exploration, 2024(10).
[14] Xiang Wei. A survey of event knowledge graph construction technology and applications[J]. Computer and Modernization, 2020(1): 14-20.
[15] Wang Zhiyue, Yu Qing, Wang Nan, Wang Yaoguo. A survey of intelligent question answering based on knowledge graphs[J]. Computer Engineering and Applications, 2020(23): 1-11.
[16] Zhang Erkun, Zhang Yixiao. ChatGPT implications: New issues for communication studies in the era of large language models[J]. Chinese Journal of Journalism & Communication, 2023(6): 167-176.
[17] Li Gang, Li Yinqiang, Wang Hongtao, et al. Knowledge graph for power equipment health management: Basic concepts, key technologies, and research progress[J]. Automation of Electric Power Systems, 2022(3): 1-13.

[18] Wang Yongchao, Luo Shengwen, Yang Yingbao, Zhang Hongxin. A survey of knowledge graph visualization[J]. Journal of Computer-Aided Design & Computer Graphics, 2019(10).

Author Biographies:
Li Jia (1990–), female, from Huanghua, Hebei, master' s degree, intermediate engineer, research direction: application of artificial intelligence in electric power industry media;
Zang Yanjiao (1989–), female, from Baoding, Hebei, master' s degree, intermediate engineer, research direction: integrated publishing, intelligent writing, and news big data communication;
Gu Chenlan (1975–), female, from Shanghai, master' s degree, senior engineer, research direction: application of artificial intelligence in news media.

Responsible Editor: Li Yansong

Note: Figure translations are in progress. See original paper for figures.