

## Research Article

# How to Construct a Power Knowledge Graph with Dispatching Data?

Shixiong Fan,<sup>1</sup> Xingwei Liu,<sup>1</sup> Ying Chen ,<sup>2</sup> Zhifang Liao ,<sup>2</sup> Yiqi Zhao,<sup>2</sup> Huimin Luo,<sup>2</sup> and Haiwei Fan<sup>3</sup>

<sup>1</sup>Beijing Key Laboratory of Research and System Evaluation of Power Dispatching Automation Technology, China Electric Power Research Institute, Haidian District, Beijing 100192, China

<sup>2</sup>School of Computer Science and Engineering, Central South University, Hunan 410000, China

<sup>3</sup>State Grid Fujian Electric Power Co., Ltd., Fuzhou 350003, China

Correspondence should be addressed to Ying Chen; 1074647728@qq.com and Zhifang Liao; zfliao@csu.edu.cn

Received 24 April 2020; Revised 7 May 2020; Accepted 26 May 2020; Published 14 July 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Shixiong Fan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Knowledge graph is a kind of semantic network for information retrieval. How to construct a knowledge graph that can serve the power system based on the behavior data of dispatchers is a hot research topic in the area of electric power artificial intelligence. In this paper, we propose a method to construct the dispatch knowledge graph for the power grid. By leveraging on dispatch data from the power domain, this method first extracts entities and then identifies dispatching behavior relationship patterns. More specifically, the method includes three steps. First, we construct a corpus of power dispatching behaviors by semi-automated labeling. And then, we propose a model, called the BiLSTM-CRF model, to extract entities and identify the dispatching behavior relationship patterns. Finally, we construct a knowledge graph of power dispatching data. The knowledge graph provides an underlying knowledge model for automated power dispatching and related services and helps dispatchers perform better power dispatch knowledge retrieval and other operations during the dispatch process.

## 1. Introduction

Smart grids have made important progress in the research and integration of dispatch automation systems [1]. According to the relevant dispatch documents such as the power dispatching control rules and the experience of the dispatcher, together with the dispatching system data and the operation state of the power grid, the dispatcher judges whether the current operation state of the power system needs dispatching and what kind of dispatching behavior is to be executed. In the actual power dispatching scenarios, power dispatching tasks are still highly dependent on the dispatcher's business knowledge and dispatching experience. Most dispatchers only understand local business knowledge [2] and cannot effectively respond to other dispatching business or global business.

With the continuous integration of multiple dispatching services, the expertise of experts or dispatchers also needs to be integrated to meet the needs of simultaneously handling complex multiservice power dispatching problems. To provide dispatcher

reference for dispatchers, digital power experts have written the texts of power dispatch, which summarize and contain all aspects of the global power dispatch business. Hence, studying knowledge organization methods for global dispatching texts, building knowledge models based on multiple dispatching behaviors, and implementing knowledge expressions that flexibly and clearly express business logic will help to improve the degree of automation of dispatching systems and provide global knowledge support for intelligent grid dispatching.

The text of power dispatching is characterized by knowledge-intensive and abundant knowledge types, and it is a kind of unstructured data. Compared with structured data with strict format and specification, the expression mode of power dispatching text is more flexible and is more difficult to read and understand. So, it is necessary to explore a natural language processing method for the dispatching text and a behavioral knowledge organization method that is suitable for the characteristics of the power dispatching behavior.

To solve these problems, we use the knowledge graph to organize knowledge of power dispatching behavior. Traditional relational databases face the problems of repeated data, weak data relationships, and difficult updates. Comparatively, the knowledge graph organizes knowledge in a graph topology, which is more in line with the structure of the power system. It regards relationships as an important knowledge element, which can better describe knowledge entities and their relationships such as power environment, dispatching roles, and power dispatching behavior. Its knowledge storage and retrieval are more flexible. First, we analyze the text of power dispatching behavior, combine natural language processing technology to build a power domain dictionary, mine power domain phrases, and identify entities based on domain phrases and domain dictionaries to achieve coreference resolution. Then, we analyze the text characteristics of dispatching behavior, define and organize the relationship of power dispatching behavior, and use the graph structure to store entities and relationships, thereby constructing a knowledge graph of power dispatching behavior.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 describes the constructing method of knowledge graph of power dispatching behavior. Section 4 introduces the experiment and evaluation results, data set, experimental design, experimental details, and experimental results. Section 5 introduces an experimental summary and future work outlook.

## 2. Related Work

Google first proposed the concept of the knowledge graph in 2012 [3], which can formally describe things and their related relationships in the real world. The knowledge graph uses  $\langle \text{entity}, \text{relationship}, \text{entity} \rangle$  triples to store knowledge, and it uses entities as nodes, relationships as edges to build a knowledge network, which conforms to the behavior rules of general subjects, actions, and action objects, and uses graph structures to describe the relationship. At present, many well-known knowledge graph projects organize a large amount of data, extract knowledge from them for organization and management, and provide users with high-quality intelligent services, such as understanding the semantics of search and providing more accurate search answers. In recent years, due to the development of crowdsourcing [4] and open-source ecosystems [5–7], the related research of constructing knowledge graphs by crowdsourcing and knowledge graphs in software has become a new research topic in the field of knowledge graphs, which also shows that knowledge graphs are flexible to organize domain knowledge.

Entities are the basic units in the knowledge graph, including attributes, attribute values, and the correspondence between related entities. Wang [8] designed a named entity recognition system based on text structure features. This method needs to design the characteristics of text in different fields separately, which is not universal. As the scale of data grows, the study on multilayer architecture and deep learning is extraordinarily important and necessary [9]. In order to reduce manual rules and improve the generalization

capabilities of the model, many methods based on deep learning have been used for named entity recognition in recent years. For example, Lample et al. [10] proposed a neural network structure based on bidirectional LSTM and CRF for entity recognition. This method does not rely on artificial features and domain-specific knowledge and has excellent versatility. In the same period, Chiu and Nichols [11] proposed a bidirectional LSTM and CNN hybrid model for automatic detection of the word and character-level features, eliminating the need for most feature engineering. To overcome the problem of missing Chinese information when using the English relationship extraction method, Han et al. [12] generated a large-scale Chinese relationship extraction data set based on the Chinese Encyclopedia and proposed an attention model based on the entity character features in the Chinese relationship extraction method. Leng and Jiang [13] proposed an improved SDAE model for entity-relationship extraction. This method eliminates the need to annotate the relationship manually and can extract the relationship between the entities automatically through contextual features.

There are many types of power grid equipment, and the relationship between the types of equipment is complicated. There is little research on knowledge graphs in the field of power grids. Tang [14] merged existing multisource heterogeneous power equipment related data to construct a power equipment knowledge graph to improve the data storage efficiency and extraction process. Given the problem of flattening and efficient utilization of power asset information, Yang [15] proposed the general process of constructing professional knowledge graphs in the power field and proposed a multisource heterogeneous power asset information fusion method based on knowledge fusion. Li et al. [2], based on the underlying data and business logic data of the smart grid dispatching control D5000 system, used a top-down and bottom-up method to construct a dispatch knowledge graph. These power knowledge graph construction methods based on power system data such as dispatch management systems face several challenges: huge data size, complex data types, diverse knowledge content, low-value density, and low data quality. Therefore, the research on text data with higher value density has attracted much attention in recent years. Wang et al. [16] proposed a method of entity recognition, coreference resolution, and relationship extraction based on the records of defects in electrical equipment and automatically constructed the knowledge graph of defects in electrical equipment to improve the retrieval quality of the records of defects in electrical equipment. Existing research usually processes unstructured text according to different text types and modes in their application scenarios, and most of them are not universal. Hence, it is necessary to study and design a knowledge graph construction method for power dispatching behavior based on power dispatching text.

## 3. Methodology

In this section, this paper will introduce the knowledge graph constructing method of power dispatching behavior, and its technical roadmap is shown in Figure 1. First, a

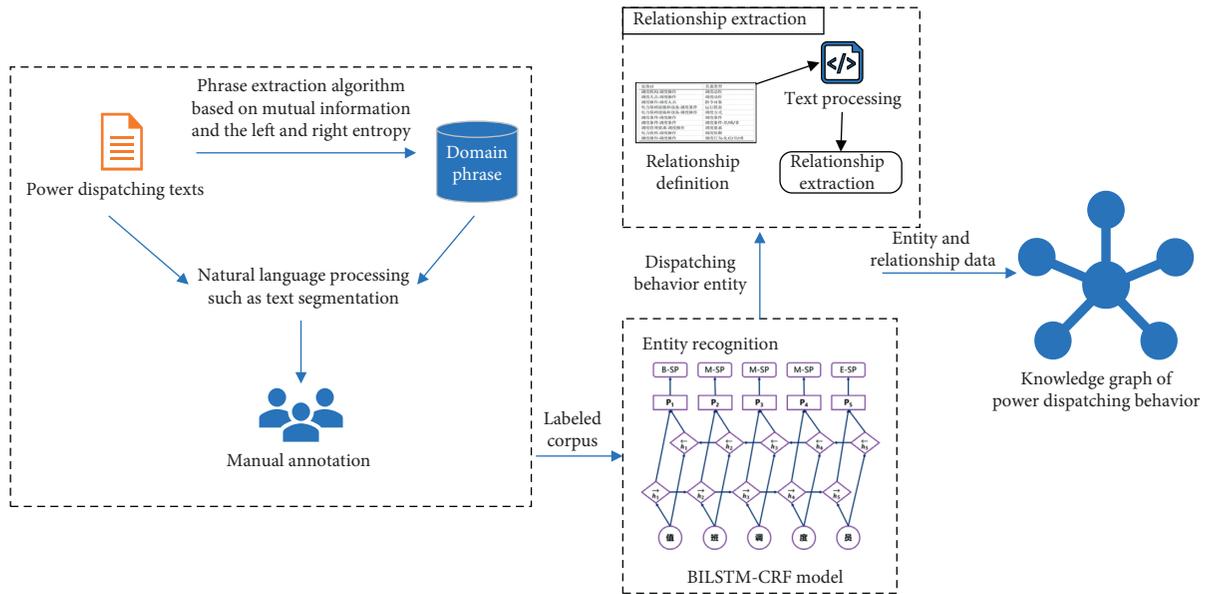


FIGURE 1: Technical roadmap for constructing a knowledge graph of power dispatching behavior.

phrase extraction algorithm based on mutual information and left and right entropy is used to extract power domain phrases, construct a professional dictionary for power dispatch, and prepare a corpus of power dispatch behavior. Then, the BiLSTM-CRF model is built to train labeled data and identify and extract entities in the power dispatching domain. Finally, by analyzing and summarizing the entity relationships, the power dispatching behavior relationships are extracted, and a graph database is used to store and construct a knowledge graph structure.

**3.1. Corpus Construction.** Constructing a high-quality domain text corpus is a prerequisite for acquiring knowledge entities of power dispatching behavior. However, there is a lack of labeled data in the field of power dispatching, and manual labeling consumes much energy. It is affected by the complexity of the domain entity category and the professionalism of the labeling personnel. Therefore, this paper uses a phrase extraction algorithm based on mutual information and left and right entropy to get candidate phrases and selects and annotates them manually to get high-quality annotation data.

**3.1.1. Phrase Extraction Algorithm Based on Mutual Information and Left and Right Entropy.** Most of the grid dispatching entities are nested combinations of multiple words. Therefore, in the traditional corpus labeling process, the original corpus must first be segmented to clarify the boundary of the words, which is convenient for manual labeling later. Existing word segmentation toolkits, such as jieba word segmentation tools, mostly use dictionary-based word segmentation methods. The dictionaries used are cross-domain general dictionaries, most of which commonly used vocabularies and lack professional vocabulary in the power field. For example, the word “operation instruction ticket” will be divided

into three words “operation,” “instruction,” and “ticket” when using the general dictionary. If we use the original corpus in the field of power dispatching to directly segment words, the effect of this method is not satisfactory. Therefore, in the stage of the cold start of corpus labeling, in order to obtain labeled corpus, this paper uses a novel unsupervised word discovery algorithm, which is a phrase extraction algorithm based on mutual information and left and right entropy.

The algorithm first calculates the mutual information between the words in the corpus. The formula is as follows:

$$PMI_{x,y} = \log_2 \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

In Formula 1,  $p(x,y)$  is the probability of two words appearing, and  $p(x)$  is the probability of a single word appearing. We use specific examples to explain this. There are three dispatching behavior words: “Provincial Dispatching,” “Dispatcher on duty,” and “Provincial Dispatcher on duty.” If the word frequency of “Provincial Dispatching” is 10, the word frequency of “Dispatcher on duty” is 20, and the word frequency of “Provincial Dispatcher on duty” is 5, the total number of words is  $N$ , and the total number of double words is  $M$ , then we have the following formula:

$$PMI_{\text{Provincial Dispatching, Dispatcher on duty}} = \log_2 \frac{5/M}{(10/N) * (20/N)} \quad (2)$$

The mutual information can reflect the relationship between two words well. The higher the mutual information value is, the higher the correlation between  $X$  and  $Y$  is, the more likely  $X$  and  $y$  are to form phrases. On the contrary, the lower the mutual information value is, the lower the correlation between  $X$  and  $Y$  is, the more likely there is a phrase boundary between  $X$  and  $y$ .

Mutual information indicates the relevance of the two words. Also, we need to calculate the degree of freedom of the word. The degree of freedom refers to the degree of diversity of adjacent words that appear on the left and right sides of the word. If the left and right sides of a candidate word are different words in different sentences, the smaller the connection between the word and other words, the greater the internal connection between the candidate words, that is, the greater the possibility that the candidate words have boundaries and are a single word.

We use the left and right entropy to measure the degree of freedom. Entropy can describe information uncertainty. In information theoretic learning, correntropy has been a widely used nonlinear similarity measure method due to its robustness [17]. The larger the left entropy and right entropy of a candidate, the more uncertain the words that may appear on the left and right sides of the candidate and the higher the degree of freedom. The formula for calculating the left and right entropy is as follows:

$$\begin{aligned} E_L(W) &= - \sum_{\forall a \in A} P(aW | W) \cdot \log_2 P(aW | W), \\ E_R(W) &= - \sum_{\forall b \in B} P(Wb | W) \cdot \log_2 P(Wb | W). \end{aligned} \quad (3)$$

Taking the left entropy as an example, suppose that the ‘‘Dispatcher on duty’’ has several kinds of collocations: ‘‘National Dispatcher on duty,’’ ‘‘Province Dispatcher on duty,’’ and ‘‘Temporary Dispatcher on duty,’’ then the left entropy of the word ‘‘Dispatcher on duty’’ is as follows:

$$\begin{aligned} & - E_L(\text{Dispatcher on duty}) \\ &= P(ND | \text{Dispatcher on duty}) \\ & \quad \cdot \log_2 P(ND | \text{Dispatcher on duty}) \\ & + P(PD | \text{Dispatcher on duty}) \\ & \quad \cdot \log_2 P(PD | \text{Duty Officer}) \\ & + P(TD | \text{Dispatcher on duty}) \\ & \quad \cdot \log_2 P(TD | \text{Dispatcher on duty}). \end{aligned} \quad (4)$$

The final input result is the score of a series of words. The calculation formula of the score is as follows:

$$\text{score} = \text{PMI} + \min(\text{left entropy}, \text{right entropy}). \quad (5)$$

These scores are sorted from high to low. We add the top 100 words to the jieba word segmentation dictionary and then perform word segmentation processing on the original corpus text to facilitate the manual labeling of the role of words in the later period.

**3.1.2. Manual Annotation.** Due to the fuzzy boundary of Chinese words and a large number of cross-nesting structures in the grid dispatching entity, the complexity of the identification task increases. Furthermore, the data set in this paper contains multiple categories of entities. Consequently, according to the word segmentation results obtained by the unsupervised phrase extraction method, the word segmentation results need to be returned to the original corpus

after manual inspection. Then, we use the BMESO labeling mechanism to convert it to the input format required by the model and finally get a labeled Training data set. The definition of the BMESO annotation model is shown in Table 1.

The labeling tool we use is YEDDA. For named entities in the field of power dispatching, we have summarized many dispatch documents and dispatch glossary classification methods. Then, a number of collaborators form a team to collaborate to review and determine the dispatching behavior entities and finally they are classified as follows:

- (1) Scheduling mechanism (SM): including China’s five major power generation groups, regional power generation groups, State Grid Corporation of China, Regional Power Grid Corporation, management organizations, and departments at all levels
- (2) Scheduling personnel (SP): including leaders of various organizations, technical personnel at all levels, and dispatching personnel on duty at all levels
- (3) Scheduling operation (SO): including but not limited to dispatching operation related to the protection device
- (4) Facilities (Fac): such as transformer, bus, line, circuit breaker, switch, knife gate, protection device, primary equipment, secondary equipment, electrical equipment, boiler equipment, steam (water, gas) turbine equipment, power transmission equipment, transmission equipment, converter equipment, power system, chemical treatment, and fuel transportation
- (5) Management requirements (MR): including scheduling management scope (equipment name), scheduling management mode, and scheduling instructions
- (6) Electric power data (EPD): such as power-related documents, systems, and operation tickets
- (7) Scheduling condition (SC): the objective conditions for certain dispatching under the power performance, such as the conditions for the power outage and power stations or substations on both sides of the line
- (8) Equipment state (ES): such as operation, maintenance, standby, charging, power transmission, power failure, and other equipment states

**3.2. Entity Extraction.** In the past few years, the rapid development of machine learning has attracted the attention of many researchers [18]. In order to identify and extract knowledge entities, this paper uses a Bidirectional Long Short-Term Memory (BiLSTM) model and Conditional Random Fields (CRF) model as a named entity recognition model. We use the annotated data of the annotated corpus above for model training and extraction of knowledge entities in the field of power dispatching behavior. The BiLSTM model is composed of forward LSTM and backward LSTM. The LSTM model can memorize the long-term dependence of sentences from front to back, but it cannot encode information from back to front. Compared with a single LSTM model, the BiLSTM model can obtain bidirectional semantic

TABLE 1: The definition of the BMESO annotation model.

Annotation	Meaning
B	The first word of the entity
M	The internal words of the
E	The suffix of the entity
S	Single entity words
O	Nonphysical constituent words

dependence and obtain more comprehensive text information. However, the BiLSTM model does not guarantee that the prediction results obtained at each output layer are correct, and some prediction results that do not meet the constraints of the training set may appear. Therefore, the CRF model can be introduced to learn the constraining rules, thereby reducing the output of the model the probability of an illegal sequence. The annotated corpus constructed above prepares for the building of an entity recognition model. At the same time, annotated data is used for model training to identify entities with domain knowledge of power dispatching behavior.

The BiLSTM + CRF model is mainly composed of three layers, and the schematic diagram of the model is shown in Figure 2. The first layer is the embedding layer. The word vector is trained by inputting the pretrained character vector and word vector, and the dictionary obtained in the previous corpus labeling process is added to make the generated word vector more capable of expressing semantics.

The second layer in the middle is the forward and backward LSTM layer. In order to make full use of word meaning and word order information, the input sequence of the character vector and the word vector of the matching dictionary are subjected to feature fusion through network calculation.

The BiLSTM layer automatically extracts sentence features, uses the char embedding sequence  $(x_1, x_2, x_3, \dots, x_n)$  of each word in a sentence as the input of each time step of Bi-LSTM, and then uses the hidden state sequence  $(\vec{h}_1, \vec{h}_2, \vec{h}_3, \dots, \vec{h}_n)$  output by the forward LSTM and the reverse LSTM  $(\overleftarrow{h}_1, \overleftarrow{h}_2, \overleftarrow{h}_3, \dots, \overleftarrow{h}_n)$ . The hidden state output at each position is stitched by a position to obtain a complete hidden state sequence:  $(h_1, h_2, h_3, \dots, h_n) \in \mathbb{R}^{n \times m}$ .

The output of this layer is the score of each label of a word by selecting the highest label score as the label of the word.

Finally, the CRF layer is introduced for sentence-level sequence annotation. The parameter of the CRF layer is a  $(k+2) \times (k+2)$  matrix  $A$ ,  $k$  is the number of labels in the label set, and  $A_{ij}$  represents the transfer score from the  $i$ -th label to the  $j$ -th label. When labeling a location, you can use the label that has been labeled before. The reason for adding 2 is to add a start state to the beginning of the sentence and an end state to the end of the sentence. Adding the CRF layer can consider the order between the labels of the output words of the Bi-LSTM layer, adding some constraints to the last predicted label to ensure that the predicted label is legal.

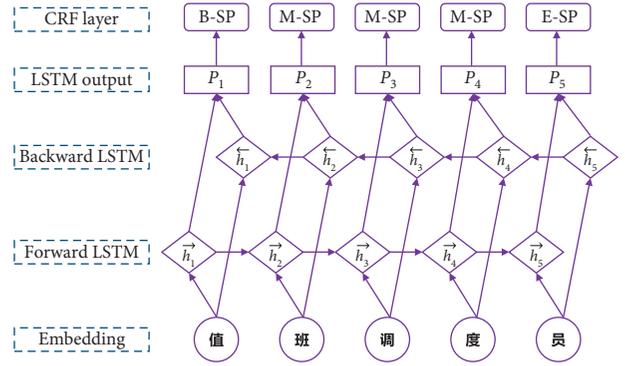


FIGURE 2: Schematic diagram of the BiLSTM-CRF model.

This paper introduces the Dropout mechanism to prevent overfitting. The Dropout mechanism prevents overfitting by randomly deleting hidden neurons in the network with a certain probability. The neurons in the input layer and the output layer of the network remain the same. In this way, the hidden neurons deleted in each iteration cycle are different, which increases the randomness of the network and improves the generalization ability of the network. The model code is shown in Algorithm 1.

**3.3. Relationship Extraction.** In order to mine the relationship of power dispatching behavior, this paper needs to analyze the language characteristics of the relationship description of power dispatching text. Since the power dispatching text is an unstructured natural language text written in Chinese, it has the characteristics of the Chinese language grammar and the power field. Its specific characteristics are as follows.

The sentence contains a large number of power domain entities. In a sentence related to scheduling behavior, there may be three or more behavior subjects and objects at the same time. The relationship network formed by the relationship between any two entities in the sentence is complicated. However, the entity-relationship category is relatively straightforward, and the entity relationships between the restricted entity categories mostly belong to one category.

Each sentence in the dispatching text corresponds to a dispatch behavior, and each segment corresponds to a type of dispatch scenario with various types. Understanding the dispatch statement requires professional knowledge of electricity, and it is difficult for nonprofessionals to learn. The Chinese grammatical structure is more flexible and sophisticated than English, with many grammatical phenomena such as condition, sequence, causality, and passive. Different writers have different language habits, and different scheduling behaviors will also use different expressions. At present, there is a lack of available syntactic knowledge rule base in the field of electric power.

The dispatching text which is the basis of dispatching behavior is based on the summary of real-world dispatching behaviors. The content is refined, the data volume is

Input: self.  
 Output: Trained model.  
 (1) Initialize the model.  
 (2) Define the Embedding layer.  
 (3) Add the Embedding layer to the model.  
 (4) Add forward LSTM to the model//units = 128, return\_sequences = True.  
 (5) Add Dropout.  
 (6) Add backward LSTM to the model//units = 64, return\_sequences = True.  
 (7) Add Dropout.  
 (8) Add TimeDistributed layer to the model.  
 (9) Define the CRF layer and Add the CRF layer to the model.  
 (10) Parameter status of each layer of the output model.  
 (11) Return model.

ALGORITHM 1: How to build BiLSTM + CRF named entity recognition model.

inadequate, and there is a lack of labeled data. The characteristics of multiple entities in the sentence make it challenging to label entity-relationship data. Machine learning algorithms commonly used in the general field often require large amounts of labeled data and cannot be directly applied to power dispatch texts.

Based on the above characteristics, we define the types of power dispatching behavior relations, as shown in Table 2. In the knowledge graph, the edges representing the relationship have directions, and the relationship edges in different directions may have different relationship types.

According to the above definition, most of the two entities have only one type of relationship. If two entities appear in a general sentence and their entity type meets the predefined relationship, it can be considered that there is a predefined relationship between the two entities. To extract the entity relationship, if there are multiple entities of the same type in a sentence, there may be a special relationship between these entities, such as a union. When analyzing power dispatching behavior sentences, words such as “common,” “parallel,” “and,” and “or” are often used in the sentence to express the order, parallel, and other relationships. If there are related words in the sentence that represent particular sentence patterns such as juxtaposition, negation, and time, it can be determined that the sentence has a special relationship and a particular relationship type. For the dispatcher and dispatch operation entity, the relationship type between the two types of entities is judged according to the position characteristics of the entity in the sentence. If the dispatcher entity is before the dispatch operation entity, the relationship arrow is directed by the dispatcher to the dispatch operation. Otherwise, the relationship arrow is determined by the dispatch. The operation is directed to the dispatcher. Therefore, this paper sorts out and extracts the entity relations of power dispatching behavior.

**3.4. Knowledge Graph Construction and Retrieval.** After extracting power dispatching behavior entities and relationships, we use a graph database to store entity and attribute information and rely on entity relationships to

connect directed edges between entity nodes, thereby constructing a knowledge graph structure. We use a graph database query language to provide a retrieval method based on knowledge graphs. Neo4j database is one of the more popular graph databases, with good performance and a friendly user interface. We use the Neo4j database as a storage database to construct a knowledge graph for power dispatching and use the declarative graph query language Cypher provided by the Neo4j database for knowledge graph retrieval.

## 4. Experiment

Based on the knowledge graph construction method proposed above, this paper presents the experimental work of labeling corpus construction, knowledge entity extraction, and knowledge graph construction of power dispatching behaviors with the power dispatch text data set. In this section, we will detail the experimental design, experimental details, and experimental results.

**4.1. Data Sets and Data Preprocessing.** In this paper, we crawled 29 documents related to power dispatching behavior such as power grid dispatching procedures, basic knowledge of dispatching, and disposal plans of dispatching failure. These documents were written by professional power dispatchers, and these documents fully describe the power dispatching business process, dispatching requirements, and dispatching behavior of dispatchers in the dispatching process. In this paper, the above documents are used as the original corpus for entity extraction and knowledge graph construction experiments. In order to facilitate the follow-up work, we unify the document format, remove the spaces and numbers in the document, and leave only character-type data.

### 4.2. Experiments and Result Analysis

**4.2.1. Construction of Power Dispatching Behavior Annotated Corpus and Entity Extraction.** There are a large number of unlabeled entity vocabularies in the field of power grid

TABLE 2: Predefined types of power dispatching behavior relations.

Entity pairs	Relational type
Scheduling mechanism—scheduling operation	Scheduling action
Scheduling personnel—scheduling operation	Scheduling action
Scheduling operation—scheduling	Instruction object
Personnel facilities—scheduling condition	Running state
Facilities—scheduling operation	Scheduling mode
Scheduling condition—scheduling operation	Scheduling condition
Scheduling condition—scheduling condition	Scheduling condition—and/or/not
Management requirements—scheduling operation	Scheduling requirements
Electric power data—scheduling operation	Scheduling basis
Scheduling operation—scheduling operation	Scheduling behavior—order/and/or

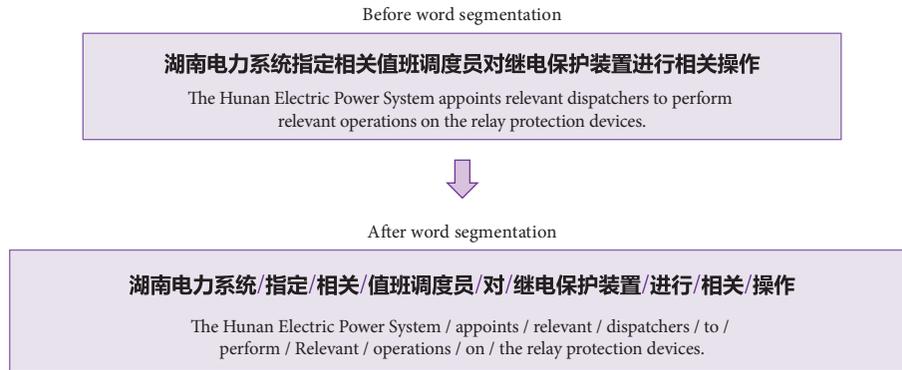


FIGURE 3: Partial word segmentation results of power dispatch text.

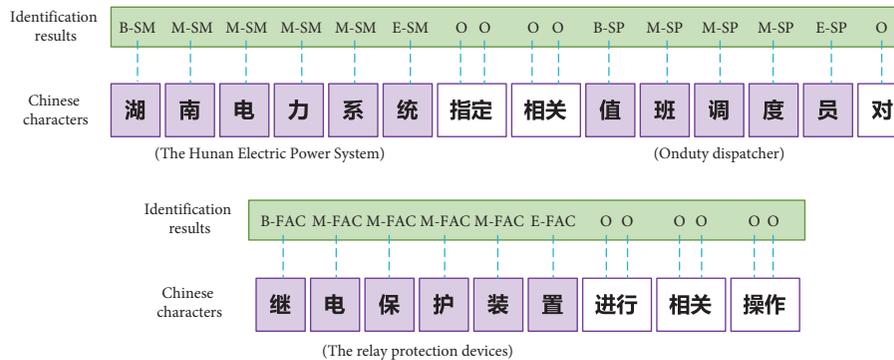


FIGURE 4: Partial recognition results of power dispatching behavior entities.

dispatching in the obtained power dispatch text data set. Due to professional domain issues, these documents have no distinct word boundaries. Then, we use a phrase extraction algorithm based on mutual information and left and right entropy to extract domain words and use the extracted domain words as a custom dictionary of Chinese words segmentation tool named “jieba” to assist in document segmentation. As can be seen from the word segmentation results in Figure 3, the use of the phrase extraction algorithm can improve the quality of word segmentation and separate the professional vocabulary in the power field such as Hunan Power System and Relay Protection.

According to the entity category of power dispatching behavior defined in this paper, we complete the construction of the labeled corpus of power dispatching behavior by manually labeling the corpus after word segmentation. Shown in Figure 3, we use the code to build the BiLSTM-CRF model, using an annotated corpus as the training set, to realize the entity recognition of text for power dispatching behavior. The recognition effect of the final model is shown in Figure 4. The entity extraction method in this paper can extract the entity vocabulary of power dispatching behavior from the power dispatching sentence and classify the entities. It can be seen that the entity extraction method in this

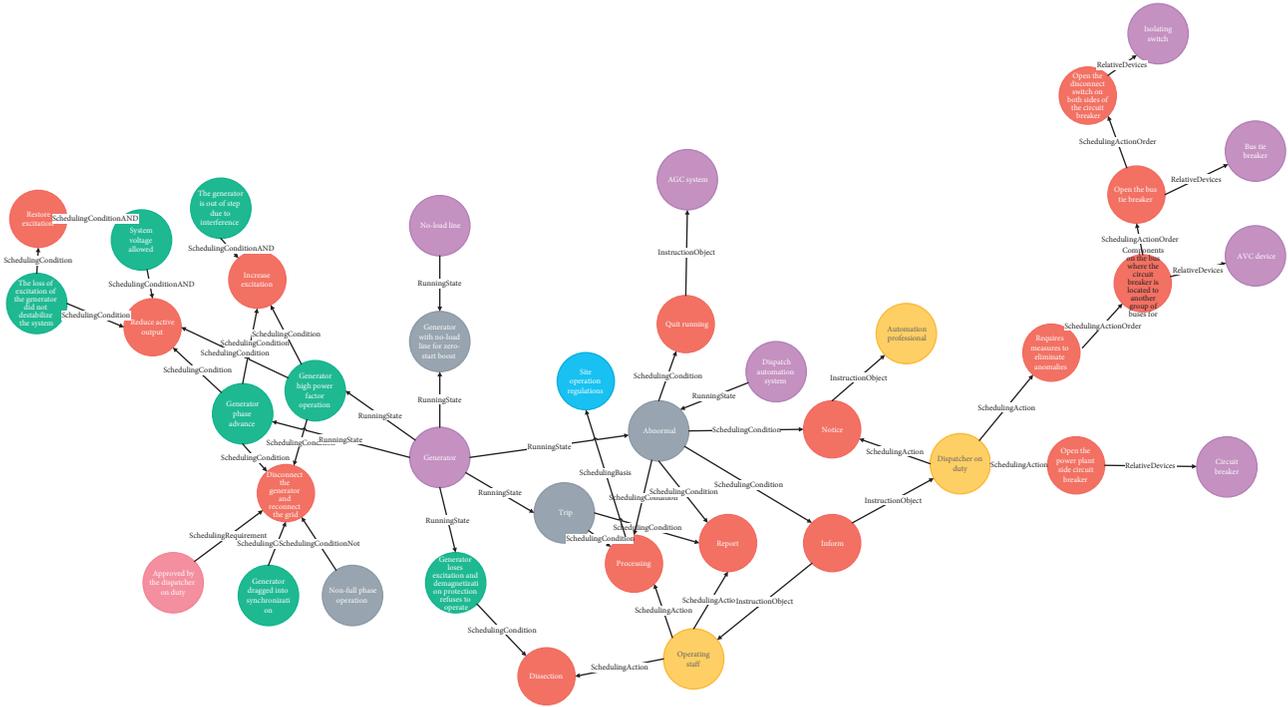


FIGURE 5: Partial knowledge graph of power dispatching behavior.

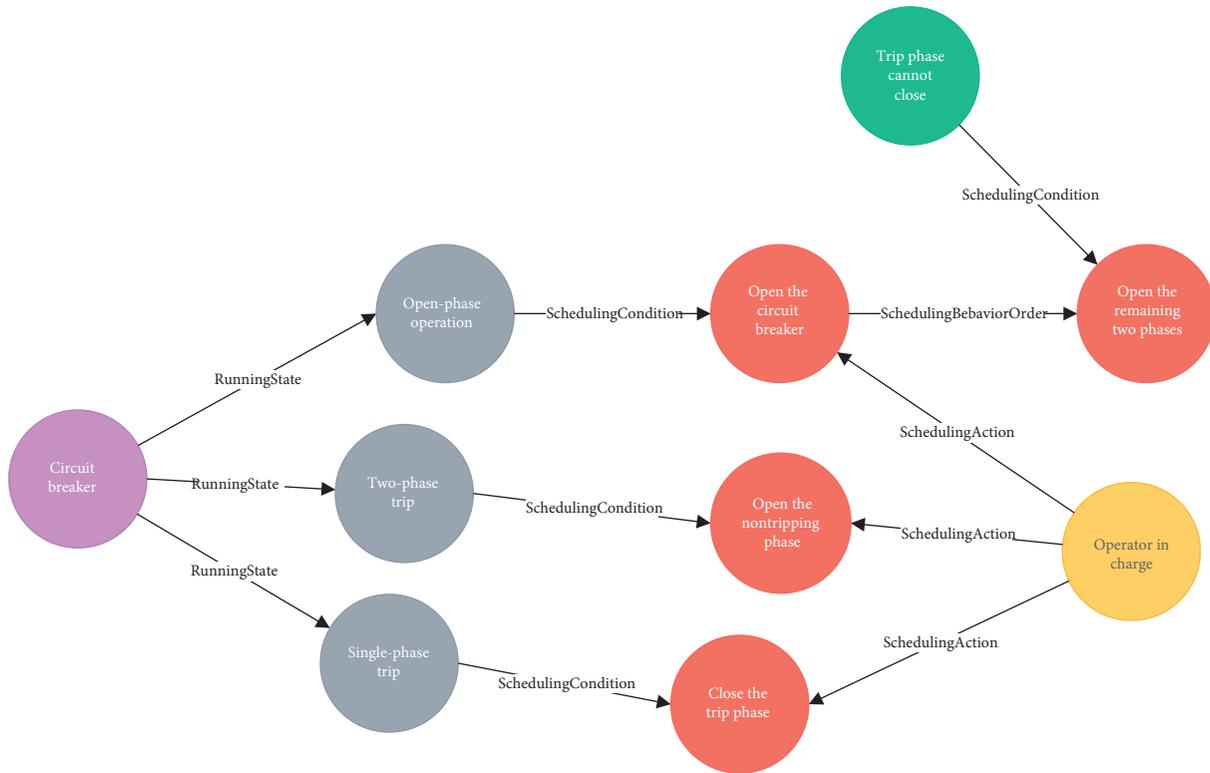


FIGURE 6: Power dispatching behavior graph under the circuit breaker non-full phase operation scenario.

paper can extract the entity vocabulary of power dispatching behavior from the power dispatching sentence and classify the entities.

4.2.2. Construction of Knowledge Graph of Power Dispatching Behavior. According to the relationship extraction method mentioned above, we extract the entity-relationship of the

power dispatching behavior based on the power dispatching texts and entity recognition results and form triples with the entity pairs. The graph database Neo4j is used to store the data and construct a knowledge graph structure. The result of the knowledge graph construction of power dispatching behavior is shown in Figure 5.

The nodes of different colors represent entities of different entity categories. Entities are connected by directed edges that represent relationships between entities to form the graph structure of the knowledge graph. The knowledge graph can store knowledge information such as knowledge entities and relationships. It is easy to see that, compared with other forms of databases such as original text and tables,

---

```
match p = () - [* {reSta: "non - full phase operation occurs during circuit breaker operation"}] - ()return p. (6)
```

---

According to the query statement to get the power dispatching behavior knowledge in this scenario, the query result is shown in Figure 6. It can be seen from a simple retrieval example that the power dispatch behavior knowledge graph constructed in this paper has both semantic information and relationship information, which can retrieve richer information and return intuitive visualization results. In addition to the example retrieval method, the knowledge graph query method is very flexible and can be queried based on entity node attributes, relationship attributes, path depth, etc., to obtain richer knowledge information. In the face of complex power dispatching business, the knowledge graph constructed in this paper will provide knowledge about the dispatching behavior of related businesses and effectively help dispatchers to conduct power dispatching.

## 5. Conclusion

This paper explores the construction method of knowledge graph based on power dispatching behavior. In order to obtain the annotated corpus, a phrase extraction algorithm based on mutual information and left and right entropy is used in this paper to annotate the corpus, by which the corpus is constructed semiautomatically. Based on the bidirectional long and short time memory network and conditional random field model, the entity is trained and identified. The relations of entities are extracted according to the text of power dispatching behavior, to store the data and construct the knowledge graph of power dispatching behavior.

According to the constructed knowledge graph, we can search more efficiently the knowledge related to the power dispatching behavior, provide the underlying knowledge model for the dispatching automation system, and further improve the intelligence of the power dispatching. There are also some problems and threats in this paper. The data set we used is small, and the diversity of knowledge content requires more knowledge data support. In

knowledge graphs link discrete data, and knowledge representation and knowledge storage are more intuitive and efficient, without the need for intermediate data conversion and processing.

This paper adds a “scheduling scenario” attribute to the relationship of the knowledge graph, to facilitate querying the possible scheduling behavior in a certain scheduling scenario in the knowledge graph. Taking the scheduling scenario of the “non-full phase operation occurs during circuit breaker operation” as an example, we executed the Cypher query language of the neo4j database to conduct the query. The specific query statement is as follows:

addition, due to the lack of updated data, we cannot study the update process of the knowledge graph, and the relationship extraction method in this article depends on text mode and rules. In the future, we will conduct further research and improvement on the existing problems, continue to explore a more efficient and automated relationship extraction model, and study a more effective construction method of knowledge graph based on power dispatching.

## Data Availability

The data set contains some books of Grid Dispatching Regulations published by STATE GRID Corporation of China and its subsidiaries, such as “Dispatching Regulation of Hunan Power Grid” for Hunan province of China.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

For this paper, Shixiong Fan conceived and designed the research study; Shixiong Fan, Zhifang Liao, Xingwei Liu, and Ying Chen collected data; Shixiong Fan, Xingwei Liu, Zhifang Liao, Ying Chen, and Yiqi Zhao designed the methodology and experiment; Shixiong Fan, Xingwei Liu, Ying Chen, Yiqi Zhao, and Huimin Luo completed the experiment; Shixiong Fan and Haiwei Fan conducted application deployment; Ying Chen, Yiqi Zhao, and Huimin Luo wrote and modified the initial paper; Zhifang Liao, Ying Chen, and Huimin Luo revised the paper. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This work was supported in part by the Basic Prospective Project of SGCC (no. 5442DZ180017) and in part by the

Science and Technology Research Foundation of SGCC (5442DZ180024-1).

## References

- [1] Y. Bi, L. Jiang, X. Wang, and L. Cui, "Design and investigation on service-oriented architecture-based smart grid dispatching and control System," *Automation of Electric Power Systems*, vol. 39, no. 2, pp. 92–99, 2015.
- [2] X. Li, J. Xu, Z. Guo, W. Ning, and Z. Wang, "Construction and application of knowledge graph of power dispatch automation system," *China Electric Power*, vol. 52, no. 2, pp. 70–77, 2019.
- [3] T. Steiner, "Adding realtime coverage to the google knowledge graph," *Proceedings of the 11th International Semantic Web Conference (ISWC 2012)*, Boston, MA, USA, September 2012.
- [4] Z. Liao, Z. Zeng, Y. Zhang, and X. Fan, "A data-driven game theoretic strategy for developers in software crowdsourcing: a case study," *Applied Sciences*, vol. 9, no. 4, p. 721, 2019.
- [5] Z. Liao, Z. Wu, Y. Li, Y. Zhang, X. Fan, and J. Wu, "Core-reviewer recommendation based on Pull Request topic model and collaborator social network," *Soft Computing*, vol. 24, no. 8, pp. 5683–5693, 2020.
- [6] Z. Liao, B. Zhao, S. Liu et al., "A prediction model of the project life-span in open source software ecosystem," *Mobile Networks and Applications*, vol. 24, no. 4, pp. 1382–1391, 2019.
- [7] Z. Liao, L. Deng, X. Fan et al., "Empirical research on the evaluation model and method of sustainability of the open source ecosystem," *Symmetry*, vol. 10, no. 12, p. 747, 2018.
- [8] N. Wang, "Company name identification in Chinese financial domain," *Journal of Chinese Information Processing*, vol. 16, no. 2, pp. 1–6, 2002.
- [9] X. Luo, Y. Li, W. Wang, X. Ban, J.-H. Wang, and W. Zhao, "A robust multilayer extreme learning machine using kernel risk-sensitive loss criterion," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 1, pp. 197–216, Jan. 2020.
- [10] G. Lample, "Neural architectures for named entity recognition," 2016, <https://arxiv.org/abs/1603.01360>.
- [11] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357–370, 2016.
- [12] X. Han, Y. Zhang, W. Zhang, and T. Huang, "An attention-based model using character composition of entities in Chinese relation extraction," *Information*, vol. 11, no. 2, p. 79, 2020.
- [13] J. Leng and P. Jiang, "A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm," *Knowledge-Based Systems*, vol. 100, pp. 188–199, 2016.
- [14] Y. Tang, "Enhancement of power equipment management using knowledge graph," in *Proceedings of the 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, IEEE, Chengdu, China, May 2019.
- [15] Y. Yang, "Multi-source heterogeneous information fusion of power assets based on knowledge graph," *Proceedings of the 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, IEEE, Zhengzhou, China, October 2019.
- [16] H.-F. Wang, "An error recognition method for power equipment defect records based on knowledge graph technology," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 11, pp. 1564–1577, 2019.
- [17] X. Luo, J. Sun, L. Wang et al., "Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4963–4971, 2018.
- [18] M. Chen, Y. Li, X. Luo, W. Wang, L. Wang, and W. Zhao, "A novel human activity recognition scheme for smart health using multilayer extreme learning machine," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1410–1418, 2019.