Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **7**, Number **4**, November **2022**

Chinese Named Entity Recognition Method in Electricity Based on Combining Character Sequence and Word Sequence

Shuaisong Yang

School of Computer Science
Northeast Electric Power University
Changchun Road, Jilin City, Jilin Province, 132012, China
yangshuaisong@163.com

Yankun Gao

Beijing Institute of computer technology and Application
Yongding Road, Haidian District, Beijing City, 100039, China
jlugyk@163.com

Jingdong Wang*

School of Computer Science
Northeast Electric Power University
Changchun Road, Jilin City, Jilin Province, 132012, China
*Corresponding Author: wangjingdong@neepu.edu.cn

Fanqi Meng

School of Computer Science
Northeast Electric Power University
Changchun Road, Jilin City, Jilin Province, 132012, China
mfq81@163.com

Shuqiang Guo

School of Computer Science
Northeast Electric Power University
Changchun Road, Jilin City, Jilin Province, 132012, China
guoshuqiang@gmail.com

Lina Zhou

Beijing Institute of computer technology and Application
Yongding Road, Haidian District, Beijing City, 100039, China
1764413771@qq.com

*Corresponding author: Jingdong Wang

Received March 10, 2022, revised April 22, 2022, accepted June 30, 2022.

Abstract. *Chinese named entity recognition in the power field is critical in building a high-quality knowledge graph of power equipment fault. Since the existing entity recognition methods only focus on the features of character sequence, it isn't easy to achieve excellent recognition results in a professional and complex electrical knowledge corpus. This paper proposes a method for Chinese named entity recognition based on character sequence and word sequence to solve this problem. The innovation lies in co-encoding character sequences and word sequences to identify various types of entities using the improved Transformer structure. First, an electricity glossary is constructed using an N-gram based unsupervised method. Then, the electricity glossary is imported into the custom dictionary of the word segmentation tool to correct the word segmentation results, and the Word2Vec model is used to train the lexicon of electricity word embedding. Finally, a model based on the combination of character sequences and word sequences is used to identify Chinese named entities. In this paper, four experiments are carried out with the corpus of electrical equipment fault diagnosis as the research object. The experimental results show that compared with the BiLSTM-CRF model and the BERT-BiLSTM-CRF model, the F1 score is increased by 23.15% and 10.62%, respectively. At the same time, compared with the control experiment using only character sequence features, the F1 score is increased by 2.96%, and the precision is increased by 4.65%, which proves the effectiveness of the method proposed in this paper.*
Keywords: named entity recognition, Transformer, character sequence, word sequence, Word2Vec

---

1. **Introduction.** Chinese named entity recognition in electricity is research on the specific domain. Its purpose is to identify various entities contained in the electric power Chinese corpus [1]. Named entities refer to text fragments with a particular meaning or strong referentiality, usually including person, place, organization, dates and times, proper nouns, etc. [2]. In addition, one can recognize other categories of entities according to actual requirements [3-4]. Therefore, the concept of entity can be comprehensive. As long as a particular text fragment is required by the business, it can be considered an entity. In the Chinese named entity recognition in electricity research, this paper divides the entities into four categories: equipment, parts, faults, and operations according to actual business needs. Extracting the four types mentioned above of entities can lay the foundation for the subsequent research on the construction of the knowledge graph of power fault [1,5].

In the field of electric power, compared with a small amount of existing structured data, the widely existing unstructured data represented by text type contains massive amounts of information [6]. Among them, it includes the information describing the defects of equipment and parts in detail, the cause of the defect, and the processing operation of the defect [7]. Therefore, if the structured information can be extracted from the power fault text effectively, it can provide historical experience for the fault diagnosis and maintenance of equipment and provide a reference guide for the maintenance of similar equipment [8]. Moreover, Chinese named entity recognition is a research at the lowest level in Natural Language Processing. It is an essential prerequisite for high-level applications such as information extraction, information retrieval, knowledge graph, machine translation, and question answering systems in the electric power field [9-11].

The Chinese entity recognition task in the power field usually consists of two sub-tasks, one is to identify the entity's boundary, and the other is to classify the entity correctly [12]. The entity length span in the power corpus is large, and the power failure corpus is highly specialized and complicated to be understood by machines. Hence, named entity recognition in electricity is often complex. Through observation data and previous experiments, the named entity recognition in electricity has the following difficulties:

(1) It is difficult to identify the boundaries of entities correctly. In contrast to text classification [13], the objects of entity recognition are specific types of text fragments. It makes the goal of named entity recognition more refined and the task more complex. At the same time, due to the professionalism and complexity of the fault text and the particularity of the target in the power entity recognition task, the existing entity recognition methods applied in the open field are not strong enough to recognize professional power terminology.

(2) There is a phenomenon of polymorphism in electric power text. Entities describing the same equipment, parts, faults, and operations have many different expressions in the power text. It possibly affects the recognition performance.

Based on the above analysis, to better identify the entities in the Chinese electric power text, this paper proposes a Chinese named entity recognition method in electricity based on combining character sequence and word sequence. The innovation is to use the improved Transformer structure to encode the character sequence and the word sequence together and then combine the features of the word sequence to recognize automatically various types of entities in the power corpus. On this basis, combined with the BERT and CRF model, it effectively solves the problem that the traditional method only considers the character sequence and ignores the word sequence.

The main contributions of this paper are as follows:

(1) An unsupervised method based on the N-Gram model and statistical indicators is used to construct an electricity glossary from the electric power Chinese corpus. Solve the time-consuming and laborious problem of manually constructing the electricity glossary and provide the support of the professional dictionary package for correcting word segmentation errors.

(2) Use the Word2Vec model to train the lexicon of electricity word embedding to provide the word sequence feature source for the model. In addition, the electricity glossary was imported into the custom dictionary of the Chinese word segmentation tool (Jieba) during training to correct word segmentation results.

(3) The improved Transformer structure encodes the character sequence and the word sequence together to recognize various types of entities in the power corpus.

(4) In the task of Chinese named entity recognition in electricity, control experiments with mainstream models and only using character sequence features were carried out. The experimental results verified the effectiveness of the method in this paper.

The rest of this paper is as follows: Section 2 reviews related work; Section 3 describes the Chinese named entity recognition method in electricity based on combining character sequence and word sequence; Section 4 verifies the effectiveness of the method by experiments; Section 5 summarizes what has been done work and describe future research directions.

2. **Related Work.** For named entity recognition tasks, traditional methods usually employ dictionary matching or machine learning, such as Hidden Markov Model (HMM) [14], Conditional Random Field (CRF) [15], Support Vector Machine (SVM) [16]. These methods are limited to a two-step process of artificially modeling words and grammatical features followed by entity recognition. Manual feature modeling requires many experts to write templates and rules, which is time-consuming and labor-intensive with limited coverage.

To avoid the complicated process of manually extracting text features and automatically extracting text features, researchers have gradually adopted deep learning to solve the problem of NER. Bidirectional Long Short Term Memory (Bi-LSTM) solves the problem of gradient disappearance in Recurrent Neural Network (RNN) and extracts long sequence features by retaining important text information and forgetting unimportant information [17]. CRF can automatically learn constraints between labels during training to ensure the validity of the final predicted label sequence. Since the respective characteristics of Bi-LSTM and CRF are very suitable for sequence labeling tasks, the BiLSTM-CRF model has become a classic model for NER tasks [18]. Subsequently, many improved models based on BiLSTM-CRF appeared, which combined other models and methods based on BiLSTM-CRF to solve specific NER problems in various fields. For example, in network security, Qin et al. [19] proposed a named entity recognition method based on FT-CNN-BiLSTM-CRF. They first used feature templates to extract local contextual features and then automatically extracted character features and global text features through a neural network model. This method solved the problem of insufficient extracted features and difficulty in accurately identifying Chinese-English mixed entities. In biomedicine, Wei et al.[20] proposed an Attention-based BiLSTM-CRF model to improve the vector representation in BiLSTM through the attention mechanism. The model designs different Attention weight redistribution methods and fuses them, effectively preventing significant information loss during feature extraction.

Based on the Transformer model [21], Google proposed the BERT language model (Bidirectional Encoder Representation from Transformers) [22], which combines multiple layers of Transformers in series to generate word vectors based on contextual semantic information dynamically. The BERT model pre-trained with a large corpus can achieve a better initialization effect, so BERT has achieved excellent performance in various NLP tasks [23]. At present, the BERT-BiLSTM-CRF model has become a standard model for Chinese NER, and its effect is significantly improved compared to the BiLSTM-CRF model [24]. In biomedicine, Song et al. [25] proposed a drug name recognition method based on BERT-BiLSTM-CRF, which can effectively improve the evaluation index of local drug name recognition in Xinjiang. In the judicial field, Gu et al. [26] used the BERT pre-trained language model to generate word vectors according to the context of the words to enhance the semantic representation, realized the recognition of named entities in judicial documents, and laid the foundation for the realization of trial automation. In the military field, Lu et al. [27] constructed a BERT-BILSTM-CRF model to process named entity recognition tasks in army texts to conduct subsequent military intelligence analysis and combat informatization research.

In electric power, some research on Chinese NER has been carried out. For example, Xiao et al. [28] proposed a Chinese power metering named entity recognition technology based on joint learning. The technology combines the CNN-BLSTM-CRF model and the word segmentation model to share entity categories and confidence levels; simultaneously, the sequential computing order of the two models is changed to parallel computing, which reduces the accumulation of recognition errors. The experimental results show that the method's precision, recall, and F1 score are significantly improved without manually constructing features. Zheng et al. [29] proposed a new power Chinese NER model based on AttCNN-BiGRU-CRF, consisting of the following five layers. The Att indicates that the model is based on the attention mechanism. The BERT-based joint feature embedding layer combines character and word embedding to obtain more semantic information. The convolutional attention layer combines the local attention mechanism and CNN to capture local contextual relations. The BiGRU layer extracts high-level features of power metering text. And finally, the CRF layer obtains the output label sequence. Li et al. [30] adopted a remote supervision based manual annotation method to get a pseudo-annotated domain corpus. Then several popular methods are compared on the power dataset, and experiments show that the pre-trained neural network model and softmax classifier achieve better performance.

In the research of named entity recognition, the more mainstream methods are dictionary matching, machine learning, and deep learning. The method based on dictionary matching has high accuracy, but entity recognition is limited to the range contained in the dictionary, which has limited coverage and needs to build a high-quality dictionary in advance. Machine learning has specific generalization capabilities and is not limited to a fixed range of applications but requires manual modeling and relies on many experts to write templates and rules, which is time-consuming and labor-intensive. Among the methods based on deep learning, various optimization models based on the BiLSTM-CRF model, such as BERT-BiLSTM-CRF and other models, have achieved good results and have been widely used. However, most existing methods are limited to using only character sequence features and cannot effectively utilize word sequence features. Moreover, unlike English, Chinese does not have spaces as natural word segmentation boundaries, so correctly identifying entity boundaries is crucial for Chinese named entity recognition tasks. Especially in professional field texts, entities and sentence patterns are more complex, and it is challenging to identify entities correctly.

3. **Method.** The overall framework of the Chinese named entity recognition method in electricity based on combining character sequence and word sequence is shown in Figure 1, consisting of two main parts: building a lexicon of electricity word embedding and using character-word sequence features to identify entities. The first step of the method is to provide word sequence information to the model in the second step. And in the second step, the entity in the power corpus is recognized by combining character sequence and word sequence.

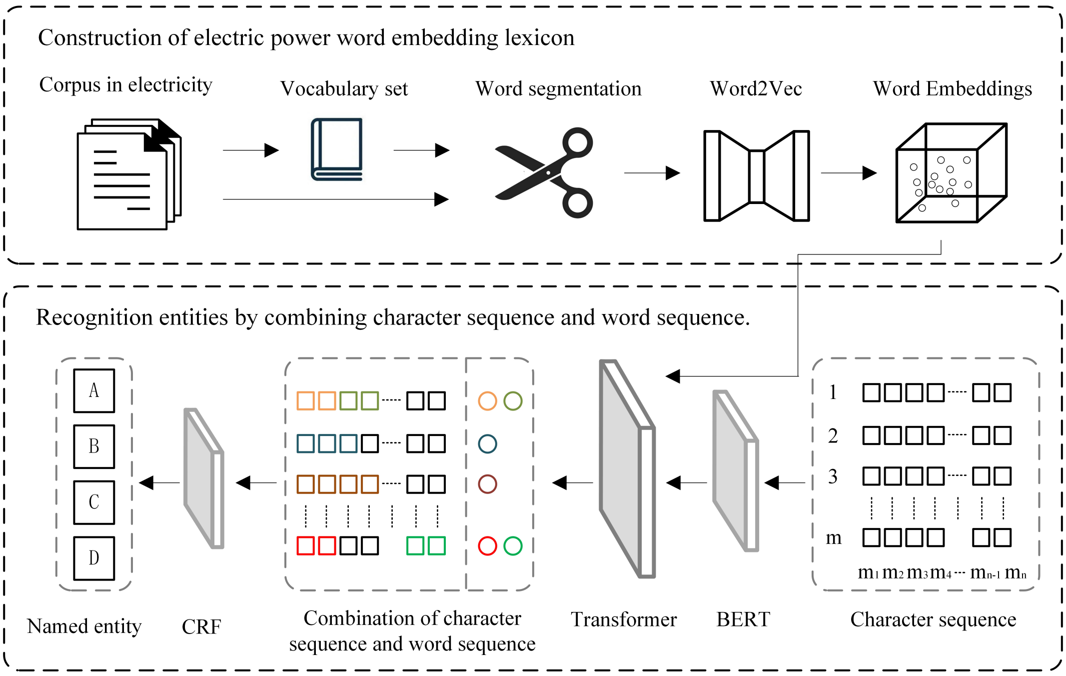

FIGURE 1. The overall framework of the Chinese named entity recognition method in electricity based on combining character sequence and word sequence

The first part uses an unsupervised method based on the N-Gram model and statistical indicators to construct an electricity glossary from the electric power Chinese corpus. Then, the electricity glossary is imported into Jieba's custom dictionary to optimize the word segmentation results on the power corpus. Finally, the Word2Vec model trains and generates a lexicon of electricity word embedding to realize the data basis for providing word sequence information for the model.

The second part identifies entities with the character and word sequences. We first annotate the power corpus using BIO format and feed it to BERT to provide character sequence features (the small squares in Figure 1 represent each character). Then, the improved Transformer model is used to match the character sequence to the words in the obtained lexicon of electricity word embedding, and the matched words are placed at the end of the sentence (the circles in Figure 1 represent the matched words). Adding a start position marker and an end position marker for all characters and words, the character sequence and word sequence are co-coded to realize the combination of word sequence features. The combined sequence is then passed into the CRF model, and the model is repeatedly revised according to the experimental results.

3.1. **Construction of lexicon of electricity word embedding.** Using the Word2Vec model to train and generate a lexicon of electricity word embedding to provide word sequence features to the entity recognition model. Word segmentation is a crucial step, and the accuracy of word segmentation affects whether the vocabulary of the trained pre-embedded lexicon and its mapping to the space are correct. In the professional field, there are many professional terms in the text that are not included in the built-in dictionary of the word segmentation tool, so it is easy to cause word segmentation errors. Although the commonly used Chinese word segmentation tools (such as Jieba) have a new word discovery function, it is still recommended to use a custom dictionary to ensure word segmentation accuracy. Therefore, it is necessary to use unsupervised methods to mine specialized vocabulary from power texts.

3.1.1. *Mining of electricity glossary.* The steps of the mining electricity glossary are shown in Figure 2:

(1) We use Jieba to perform fine-grained word segmentation on the power corpus and get the initial word segmentation results. Figure 2 uses the squares written with "A, B, C, D" to represent the initial word segmentation sequence.

(2) The N-Gram model is used to implement the second-order sliding window combination and the third-order sliding window combination for the adjacent words after word segmentation, and the combination result constitutes a candidate vocabulary set.

(3) The candidate words' statistical indicators such as word frequency, information entropy, and mutual information are calculated.

According to the candidate words' score, a threshold is set to filter the candidate words, and the screening results are formed into the electricity glossary.

Professional vocabulary mining mainly uses three indicators: word frequency, information entropy, and mutual information [31]. The word frequency refers to the number of occurrences of a candidate word. The candidate word may be a professional vocabulary when the word frequency value is enormous. If we do not set the word frequency filter, the calculation time will be significantly increased. Meanwhile, the proportion of professional words in candidate words that only appear once is not high. Therefore, this paper first screened out candidate words with a word frequency greater than or equal to 2. In addition, we use information entropy to describe the uncertainty of adjacent words before and after a candidate word—the greater the uncertainty, the greater the value of the information entropy. At the same time, the more likely that the candidate's word is a professional vocabulary. The formula of information entropy is shown in Equation (1) and (2).

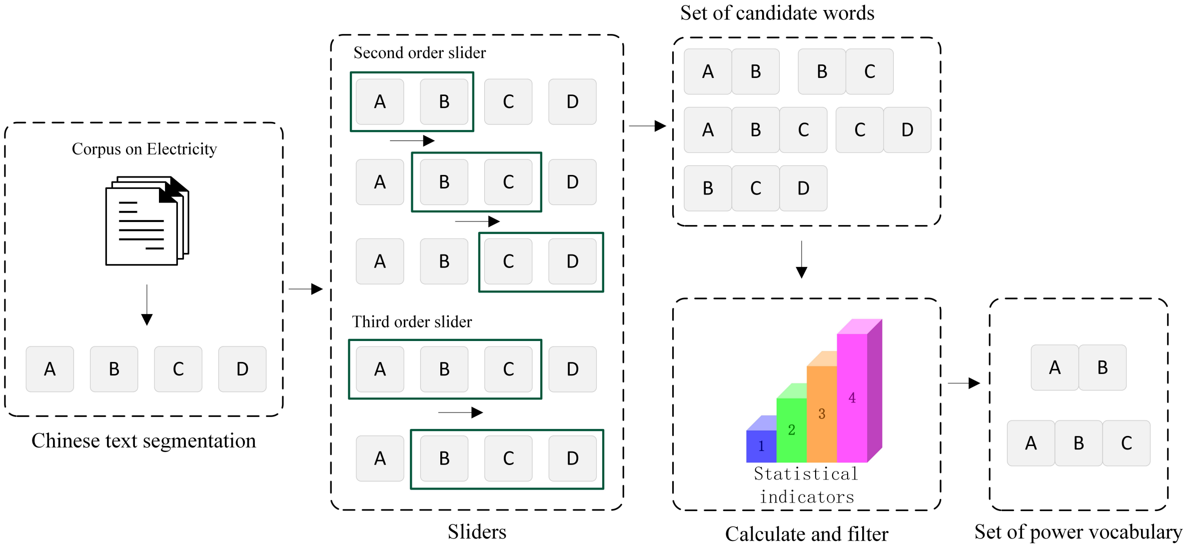

FIGURE 2. The steps of unsupervised mining of electricity glossary

$$H_L(w) = \sum_{n=1}^{N} -P(w_n)\log_2 P(w_n) \tag{1}$$

$$H_R(w) = \sum_{m=1}^{M} -P(w_m)\log_2 P(w_m) \tag{2}$$

Where: w is the candidate word, N is the total number of adjacent words on the left side, and M is the total number of adjacent words on the right side. $P(w_n)$ is the probability of the adjacent word on the left side, $P(w_m)$ is the probability of adjacent words on the right side. $H_L(w)$ is the information entropy value on the left side of the candidate word, and $H_R(w)$ is the information entropy value on the right side of the candidate word.

The greater the information entropy of a candidate word, the more likely it is to be a specialized word. Information entropy is an external indicator of candidate words. It is not enough to only consider the left and right information entropy, and we also need to consider the degree of cohesion within the candidate words. Mutual information can describe the degree of solidification. The greater the degree of solidification, the more likely the word is a professional vocabulary. The mutual information formula of the second-order slider combination is shown in Equation (3).

$$I(x;y) = \log_2 \frac{p(x,y)}{p(x)\,p(y)} \tag{3}$$

Where $p(x)$ and $p(y)$ are the probabilities of x and y appearing in the document, respectively, and $p(x,y)$ is the probability that the candidate word for the combination of x and y appears in the document. If the value of mutual information is significant, the candidate word composed of x and y is more likely to be a professional vocabulary. Conversely, it means a higher probability that x and y cannot be combined. Similar to the mutual information of the second-order slider combination, the mutual information formula of the third-order slider combination is given as Equation (4).

$$I(x;y;z) = \min\left\{\log_2 \frac{p(x,y,z)}{p(x,y)\,p(z)}, \log_2 \frac{p(x,y,z)}{p(x)\,p(y,z)}\right\} \tag{4}$$

Where $p(x, y, z)$ is the probability that the candidate word of the combination of x, y, z appears in the document; $p(x, y)$ is the probability that the candidate word of the combination of x and y appears in the document; $p(y, z)$ is the probability that the candidate word for the combination of y and z appears in the document; $p(x)$ and $p(z)$ are the probability that the word x and z appear in the document, respectively.

The final score formula of candidate words is shown in Equation (5):

$$score = I + \min\{H_L(w), H_R(w)\} \tag{5}$$

After calculating the scores of all candidate words, sort them according to the scores from large to small. Set the number to n, and construct the top n words in the score as the electricity glossary.

3.1.2. *Training lexicon of electricity word embedding.* Words in the text are in symbolic form, while mathematical models only accept numerical input. Therefore, converting the word into a numerical form of a particular dimension is necessary. The transformed word vector needs to spatially represent features such as the original word's meaning and part of speech. The above process is called word embedding, and the Word2Vec used in this paper is one of the current commonly used word embedding models. The Word2vec model mainly consists of the Skip-gram model and the CBOW model [32]. If a word is used as input to predict the context around it, the model is called a Skip-gram model. And if the context of a word is used as input to predict the word itself, it is a CBOW model. The model structure of Word2vec is shown in Figure 3.

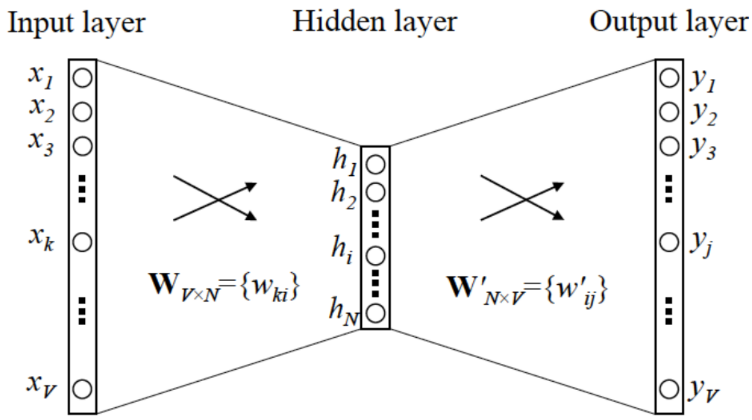

Figure 3. Structure of the Word2vec model

Mathematical models only accept numeric input, and Word2vec is no exception. Therefore, Word2Vec first maps the result of text segmentation into an N-dimensional vector through one-hot encoding as its original input. In Figure 3, $x_1, x_2, x_3, \ldots, x_k, \ldots, x_V$ are the input of V words in the form of one-hot encoding, $y_1, y_2, y_3, \ldots, y_k, \ldots, y_V$ is the probability of outputting on V words. The Skip-gram model and the CBOW model are used for multiple rounds of training so that the input and output of the model are as close as possible to the actual situation. When we complete the model training, we finally get the weight of the neural network, which is used as the vector corresponding to the word. The dimension of the word vector (consistent with the number of hidden layer nodes) is generally much smaller than the size of the total number of words, so Word2vec is essentially a dimensionality reduction operation. It reduces the dimensionality of words from one-hot encoded representations to Word2vec representations.

First of all, this paper imports the constructed electricity glossary into the Jieba custom dictionary to help correct the word segmentation results of the electric power corpus. Then the meaningless characters such as punctuation marks in the word segmentation results are cleaned. At the same time, we filter out words with more minor word frequencies, and high-frequency words are retained. Because words with lower word frequency may be useless words, misclassified words, and the linguistic meaning they reflect may be inaccurate; the higher the word frequency, the more context information is captured, which is generally credible. Finally, the Word2Vec model trains the lexicon of electricity word embedding.

3.2. **The Model for Named Entity Recognition.** The model's overall structure is shown in Figure 4, which is divided into three layers. We use the BERT layer pre-trained with a large amount of data to correct the words and the positions of the words mapped to the vector space to a certain extent. The role of the Transformer layer is to match the input character sequence with the trained lexicon and generate a combination structure of character sequence and word sequence. The role of the CRF layer is to constrain the labels of the output sequence by learning the sequence rules of the labels.

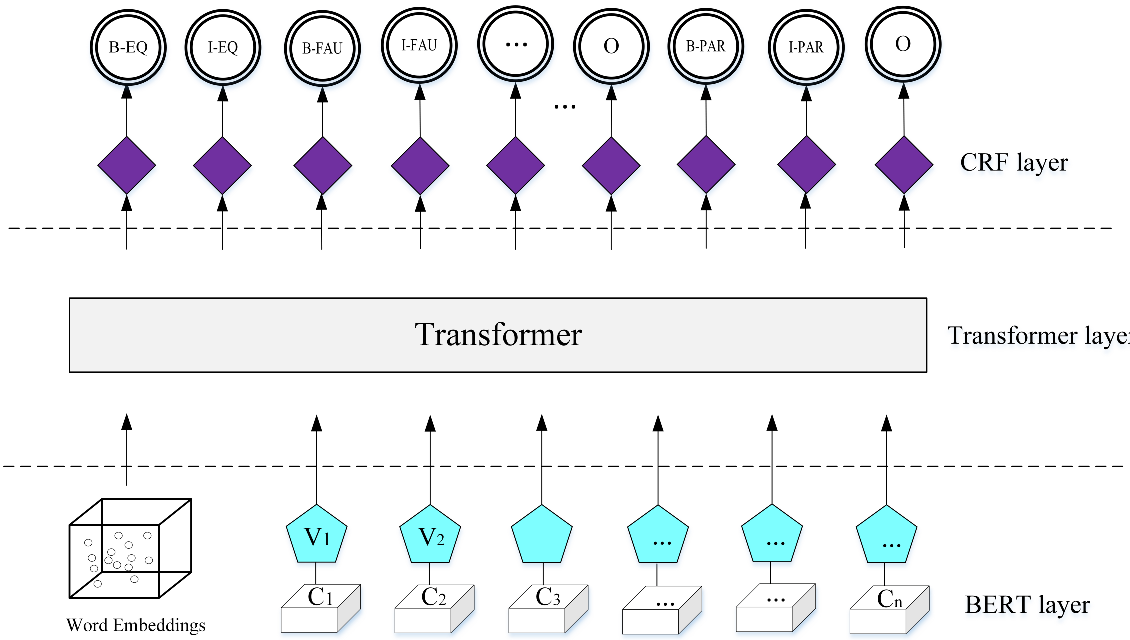

FIGURE 4. Overall structure of the model

3.2.1. *BERT.* BERT performs feature extraction and training through a multi-layer neural network and converts the input text into word vectors, enabling the BiLSTM layer to learn contextual features [33]. The BERT model converts the input sequence into a composite embedding of word, sentence, and position, then passed to the next-layer model. The input structure of BERT is shown in Figure 5.

The most important module of the BERT model is the bidirectional Transformer encoding structure for feature extraction, which uses a self-Attention mechanism and fully connected layers to model the input text. The dynamic word vectors trained by the BERT model can express different semantics in different contexts. Compared with the

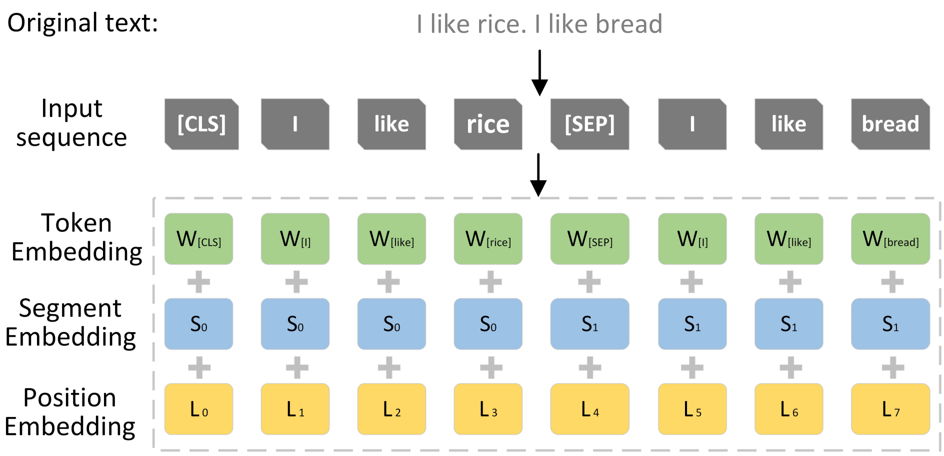

FIGURE 5. Input structure of the BERT model

static word vector obtained by the traditional model, the word vector contains the meaning of the word itself and its context. Therefore, it can also capture latent features at the sentence level.

3.2.2. *Transformer.* The encoder structure of the Transformer model consists of 6 identical base layers stacked, and each base layer consists of two sub-layers. The first is a Multi-Head Attention layer, and the second is a fully connected feed-forward neural network layer. A residual connection is then used in both sub-layers, followed by a layer normalization operation. The decoder structure is similar to the encoder structure, and it is also composed of 6 identical primary layers stacked. In addition to the multi-head attention and feed-forward neural network layers, each layer also has a hidden multi-head attention layer. This layer is Used to perform multi-head attention operations on the output of the encoder layer. Each sub-layer of the decoder also takes residual connections, followed by normalization operations. In summary, the core of the Transformer model architecture is the attention mechanism.

The attention mechanism is essentially a resource allocation model that focuses on the critical points of things at a specific moment. Self-attention is the correlation calculation for different positions in the sequence. Precisely, the input information is linearly mapped to three separate spaces, and a query and scoring mechanism is established to calculate the degree of correlation between words in a sentence. By assigning a higher weight to essential words, the model pays more attention to words that carry important information. Suppose the input is a matrix $A \in R^{nd}$, n is the sequence length, and d is the input dimension. A is mapped to different spaces Q, K, V through three different weight matrices W_Q, W_K, W_V, and the dimensions of the weight matrix are all $R^{d\times d}$. The formulas for calculating the attention mechanism using the scaled dot product are Equation (6) and Equation (7).

$$Q, K, V = AW_Q, AW_K, AW_V \tag{6}$$

$$Attention\left(Q, K, V\right) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

Where d_k is the dimension of the self-attention layer, $\sqrt{d_k}$ can prevent the inner product of QK^T from being too large. After self-attention processing, the vector at a specific position contains the information of the word itself and the correlation information with other words. Therefore, it is richer in feature expression.

This paper uses the improved Transforme model to encode the input sequence, and its structure is shown in Figure 6. Character sequences are provided by BERT, and word sequences are provided by word-by-word matching lexicon of electricity word embedding. If the match is successful, place all the matched words at the end of the sentence in turn, and add a start position marker and an end position marker to the position of each character and word. The corpora are encoded into the structure in the Transformer model to realize the combination of character sequences and word sequences to recognize electric Chinese named entities.

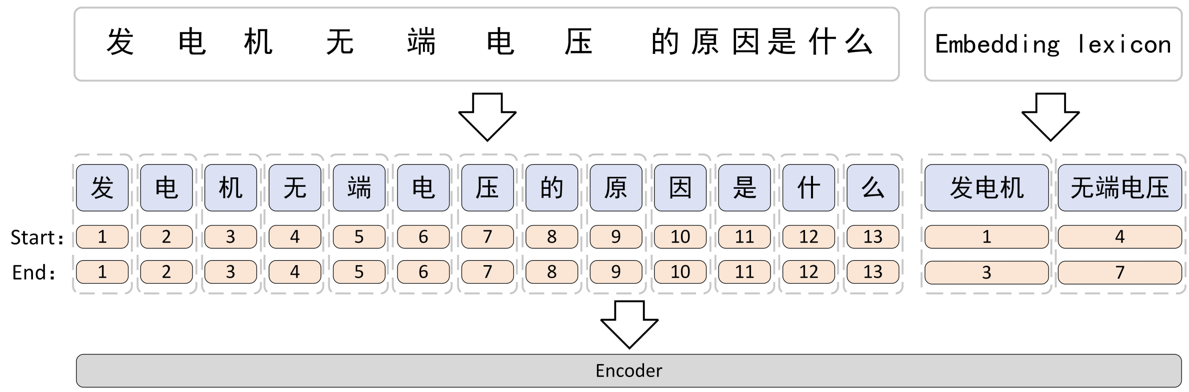

FIGURE 6. The structure of the combination of character sequences and word sequences.

3.2.3. *CRF.* The CRF model is a graphical model for calculating the joint probability distribution. By combining the correlations of the contextual sequence labels, we can calculate the probability distribution of the overall sequence. It can normalize local features to global features, obtain the optimal global solution, and outputs the final label sequence. When the CRF layer performs label training, it can receive the hidden constraint rules of the label. According to the tasks and research data in this paper, the labeling rules can be obtained as follows:

(1) The entity's label should start with B, not I. For example: if the entity MN appears "$M \rightarrow$ I-EQ, $N \rightarrow$ I-EQ", it violates the constraint rule because it should be "$M \rightarrow$ B-EQ, $N \rightarrow$ I-EQ";

(2) The labels of all characters of an entity must belong to the same class. For example, "B-FAU I-FAU" is a legal sequence, but "B-FAU I-EQ" is an illegal tag sequence;

(3) The label "B" and "O" must be used as the beginning of the sequence labeling.

Using the CRF model to identify the label sequence is selected according to the context and implicit label rules instead of choosing the sequence with the highest scoring function.

4. Experiments and analysis.

4.1. **Experimental Data.** We conduct four experiments to verify the feasibility of identifying various types of entities in text from the power equipment fault corpus and analyze the method's accuracy proposed in this paper. The data used in this experiment is the power failure QA corpus, and the corpus size is 335k. There are four types of entities: equipment, parts, faults, and operations. After labeling the corpus, we divided it into the training set and test set in a ratio of 8:2. The distribution of various types of entities in the dataset is shown in Table 1.

This paper uses the BIO format to label the dataset, the labeling tool is YEDDA, and the schematic diagram of the labeling is shown in Figure 7. Each character corresponds

Table 1. Distribution of dataset entity types

Label	Train	Test
Equipment	582	105
Part	1465	592
Fault	1071	330
Operate	796	319

to a label, and all labels can be divided into entity and non-entity. All non-entity labels are "O", and Entity tags are divided into location and type. In the location annotation, "B" means that the character is the entity's beginning, and "I" means that the character is at the non-beginning position of the entity. Moreover, in the type annotation, we use "EQ" to represent the entity of the equipment type; "PAR" to describe the entity of the part type; "FAU" to describe the entity of the fault type; "OPE" to mean the entity of the operation type.

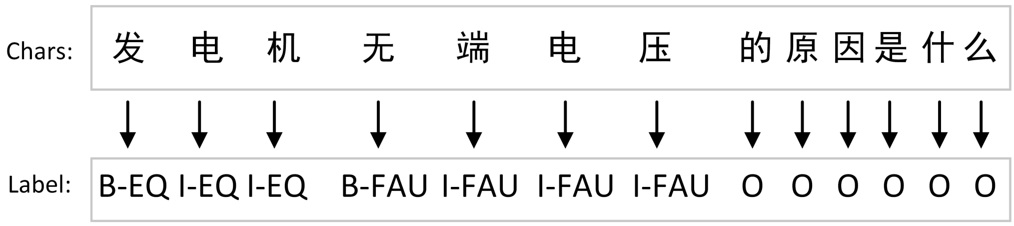

Figure 7. Example of BIO format annotation

4.2. **Evaluation Index.** We use the confusion matrix composed of the predicted results and the actual results to explain the indicators used in the evaluation experiments. Table 2 is the confusion matrix.

Table 2. Confusion matrix composed of predicted results and actual results

	The predicted result is 1	The predicted result is 0
The actual result is 1	TP	FN
The actual result is 0	FP	TN

When the predicted result is 1 or 0, the confusion matrix takes P or N, respectively. P represents the prediction result as positive, and N represents the prediction result as negative. If the actual result is the same as the predicted result, the confusion matrix gets T, which means the prediction is correct. Conversely, if the actual result is different from the predicted result, the confusion matrix gets F, and F represents the prediction error. Therefore, the confusion matrix is obtained. In the experiments of this paper, TP represents the correctly recognized entity; FN represents the incorrectly recognized entity; FP represents the non-entity recognized as an entity; TN represents the correctly recognized part of the non-entity.

The precision formula is shown in Equation (8), which represents the proportion of the predicted positive values that are positive. The precision reflects the degree of precision in the entities identified by the model. The higher precision, the more likely the entities identified by the model are correct.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The formula of recall rate is shown in Equation (9), which represents the proportion of correctly recognized entities to all entities. The recall rate reflects the recall ability of the model for entities. The higher the recall rate, the more comprehensive the entities identified by the model, and the more likely it is to identify as many entities as possible.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The higher the precision and recall of the model, the better the recognition effect. However, there is often a specific restrictive relationship between precision and recall in practice. One of the indicators may be too large to affect the other, making it difficult for both indicators to obtain high values simultaneously. Therefore, we introduce the $F1 - score$ to integrate the precision and recall rates to evaluate the experiment results comprehensively. Equation (10) is the formula of the $F1 - score$.

$$F1 - score = \frac{2Precision * Recall}{Precision + Recall} \tag{10}$$

The precision rate can reflect the prediction precision of the model's positive sample results, and the recall rate can reflect the model's degree of recall and recognition of positive samples. However, only using these two indicators to evaluate model performance is one-sided. Therefore, to comprehensively evaluate the effect of the model, the $F1-score$ is mainly used in our paper to evaluate the model's performance.

4.3. **Experimental Parameters.** To make our model perform better in the task of electric Chinese entity recognition, we need to determine the optimal parameters of the model. Therefore, we conducted comparative experiments with multiple sets of different batch-size and learning-rate. When LR is $6e^{-3}, 6e^{-4}, 6e^{-5}$, respectively, we carry out experiments under different batch-sizes, and the results are shown in Figure 8.

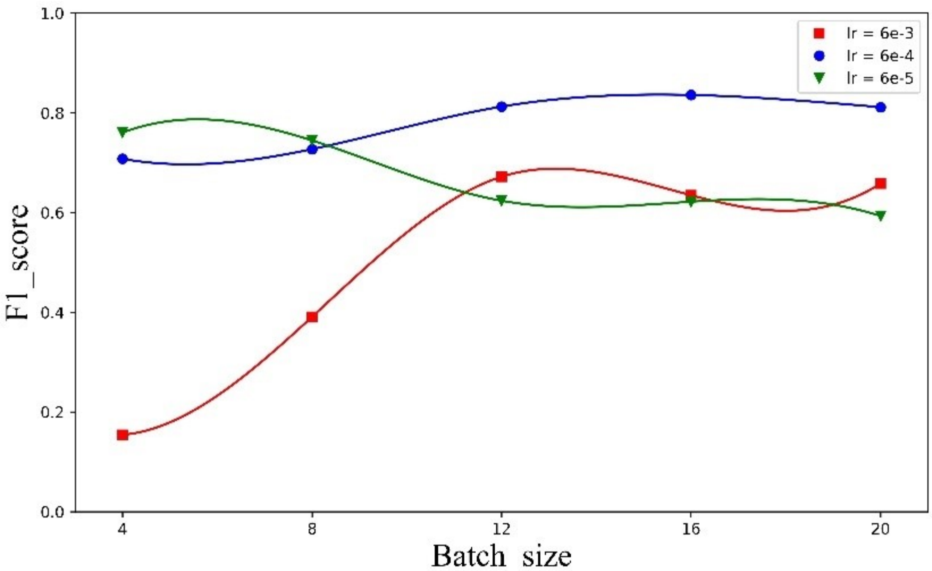

FIGURE 8. Experimental results under different learning-rate and batch-size

When the learning-rate is $6e^{-3}$, the effect curve increases with the batch-size within a specific range and then maintains a high position. When the learning-rate is $6e^{-4}$, the

Table 3. Main parameters of the model

parameters	value
batch-size	16
learning-rate	6e-4
epoch	120
layer	1
head	8
head dim	20
warmup	0.1
dropout	0.5

model works well when the batch-size is small, and it will increase to a certain extent as the batch-size increases and then maintain a high position. When the learning-rate value is $6e^{-5}$, the effect of the model is better when the batch-size is small, but it decreases as the batch-size increases and then maintains a lower position. So we can conclude that the values of batch-size and learning-rate have a strong correlation. When learning-rate takes a larger value, batch-size should also take a larger value to obtain better experimental results. Conversely, when the value of learning-rate is small, batch-size also needs to be set to a small value to get better results. Therefore, we must choose matching batch-size and learning-rate to optimize the experimental results. According to the results of the comparative experiment, we finally select learning-rate as $6e^{-4}$ and batch-size as 16. Besides learning-rate and batch-size, we also determined the remaining parameters through other comparative experiments. The final parameters of the model are shown in Table 3.

4.4. **Experimental Results.** We conducted four experiments to verify the effectiveness of our proposed method in the task of electricity Chinese named entity recognition by comparing the effects of each model. Experiment 1 adopted the BiLSTM-CRF model. Experiment 2 adopts the BERT-BiLSTM-CRF model. Experiment 3 uses the BERT-Transformer-CRF model but does not contain word sequence information. Experiment 4 uses the BERT-Transformer-CRF model, incorporating both word sequences and word sequences. In constructing the lexicon of electricity word embedding, we identified 13536 specialized vocabularies, and the constructed glossary contains 2465 vocabulary. Each vocabulary can be mapped into a 50-dimensional vector space. We use the lexicon to provide word sequence information in Experiment 4. The indicators for evaluating the experimental results are precision, recall and F1 score. The results of the four experiments are shown in Figure 9.

Firstly, comparing the precision, recall, and F1 score of Experiment 1 and Experiment 2. We can find that the BERT-BiLSTM-CRF model is improved by 11.61%, 13.45%, and 12.53%, respectively, compared with the BiLSTM-CRF model. The BiLSTM-CRF model is a classic model in entity recognition tasks. Adding the BERT model to it can significantly improve the practical effect. The reason is that BERT is a model obtained through large-scale data pre-training, which can learn word vectors with context and the order of sentences. Then, comparing Experiment 3 and Experiment 2, we can see that the precision, recall and F1 value are improved by 8.74%, 6.6% and 7.66%, respectively. Because the Transformer is based on the self-attention mechanism, it can mine long-distance dependencies in the text. Furthermore, we improve the absolute position encoding of the Transformer in the form of co-encoding head and tail positions, making it more suitable for named entity recognition tasks. Finally, compare Experiment 4 and Experiment 3.

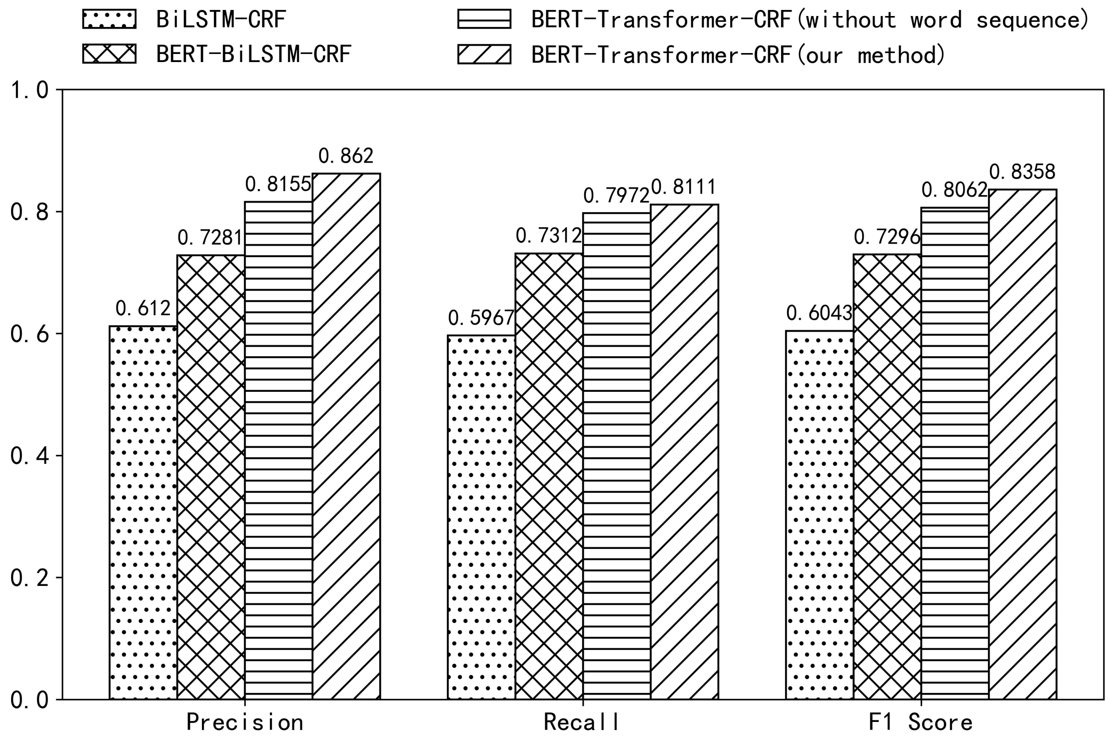

FIGURE 9. Results of four entity recognition experiments

When we added word sequence information to the BERT-Transformer-CRF model, the precision, recall and F1 values were improved by 4.65%, 1.39% and 2.96%, respectively. It shows that the BERT-Transformer-CRF model combined with word sequence can significantly improve the precision compared with only using character sequence and can substantially increase the effect of identifying entities.

By comparing the effects of the four experiments, our method is significantly better than the others, especially in terms of precision.

5. **Conclusion and future work.** To improve the effect of recognizing multiple types of entities from the Power Chinese corpus, we propose a Chinese named entity recognition method in electricity based on combining character sequence and word sequence. The novelty of this method is to improve the original absolute position encoding of the Transformer into the structure of co-encoding head position and tail position. Using a word-by-word matching lexicon can realize the combination of character sequence and word sequence so that the model can fully use the feature information in the text. Through experiments, it is found that this method is significantly better than other methods in the power Chinese named entity recognition task, especially the improvement of precision is more pronounced. It is shown that there are fewer errors in the entities identified from the electric power Chinese corpus based on this method, and the reliability and practicality of automatic identification of electric power entities are improved.

Identifying power entities accurately and comprehensively is the first step in building a high-quality knowledge graph in the power fault field. It provides nodes for building a knowledge graph of power faults. Subsequent researchers can carry out research related to relationship extraction by automatically extracting relationships between entities to provide edges connecting knowledge graph nodes. By integrating the knowledge of the power industry through the constructed power knowledge graph, operators can quickly respond to power equipment faults, discover faults and their causes in time, and ultimately improve the accuracy and efficiency of fault diagnosis. In addition, we can also integrate various sources and types of data in the power knowledge graph, such as equipment real-time monitoring information, historical fault handling information, etc. In this way, operators can diagnose equipment defeats and evaluate equipment status more comprehensively and accurately, which is also the direction of follow-up research.

Acknowledgment. This work is partially supported by the Natural Science Foundation of Jilin Province,China "Research on knowledge graph construction technology for fault diagnosis of electrical primary equipment"(NO. 20220101129JC). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] K. Zheng, L. Sun, X. Wang, S. Zhou, H. Li, S. Li, L. Zeng, Q. Gong, Named entity recognition in electric power metering domain based on attention mechanism, *IEEE Access*, vol. 9, pp. 152564-152573, 2021.

[2] M. Won, P. Murrieta-Flores, B. Martins, Ensemble named entity recognition (ner): evaluating ner Tools in the identification of Place names in historical corpora, *Frontiers in Digital Humanities*, 2018, https://doi.org/10.3389/fdigh.2018.00002

[3] A. Goyal, V. Gupta, M. Kumar, Recent named entity recognition and classification techniques: a systematic review, *Computer Science Review*, vol. 29, pp. 21-43, 2018.

[4] W. Yoon, C.H. So, J. Lee, J. Kang, Collabonet: collaboration of deep neural networks for biomedical named entity recognition, *BMC bioinformatics*, vol. 20, 249, 2019.

[5] J. Wang, X. Wang, C. Ma, L. Kou, A survey on the development status and application prospects of knowledge graph in smart grids, *IET Generation, Transmission & Distribution*, vol. 15, pp. 383-407, 2020.

[6] Z. Qu, X. Fan, H. Yu, N. Qu, Smart grid text knowledge acquisition model based on ontology, *Journal of Northeast Electric Power University*, vol. 34, no. 5, pp. 60-68, 2014.

[7] C. Silva, M. Saraee, Understanding causes of low voltage (LV) faults in electricity distribution network using association rule mining and text clustering, *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pp. 1-6, 2019

[8] C. Liu, R. Song, Y. Hou, Y. zhang, Design and implementation of high voltage transmission equipment auxiliary management system, *Journal of Northeast Electric Power University*, vol. 39, no. 4, pp. 86-90, 2019.

[9] H. Wu, L. Zhou, Y. Xin, Comparative study about personalized recommendation of home and abroad based on knowledge map, *Journal of Northeast Electric Power University*, vol. 32, no. 5, pp. 124-128, 2012.

[10] A. Kumar, D.P. Singh, S. Subham, Electricity bill management system, *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 6, pp. 6062–6067, 2021.

[11] D. Diefenbach, V. Lopez, K. Singh, P. Maret, Core techniques of question answering systems over knowledge bases: a survey, *Knowledge and Information systems*, vol. 55, no. 6, pp. 529-569, 2018.

[12] J. Li, A. Sun, J. Han C. Li, A survey on deep learning for named entity recognition, *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50-70, 2022.

[13] Y. Han, M. Li, X. Guo, The text classification attribute reduction algorithm based on the rough set theory, *Journal of Northeast Electric Power University*, vol. 36, no. 5, pp. 92-96, 2016.

[14] X. Yan, X. Xiong, X. Cheng, Y. Huang, H. Zhu, F. Hu, HMM-BiMM: Hidden markov model-based word segmentation via improved Bi-directional maximal matching algorithm, *Computers & Electrical Engineering*, vol. 94, 107354, 2021.

[15] J. Kim, Y. Ko, J. Seo, A bootstrapping approach with CRF and deep learning models for improving the biomedical named entity recognition in multi-domains, *IEEE Access*, vol. 7, pp. 70308-70318, 2019.

[16] M. Talha, S. Boulaknadel, D. Aboutajdine, Performance evaluation of svm-based amazighe named entity recognition, *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*, pp. 232–241, 2018.

[17] G. Xu, Y. Meng, X. Qiu, Z. Yu, X. Wu, Sentiment analysis of comment texts based on BiLSTM, *IEEE Access*, vol. 7, pp. 51522-51532, 2019.
[18] Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, *arXiv*, 2015, https://doi.org/10.48550/arXiv.1508.01991
[19] Y. Qin, G. Shen, W. Zhao, Y. Chen, M. Yu, X. Jin, A network security entity recognition method based on feature template and CNN-BiLSTM-CRF, *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 6, pp. 872-884, 2019.
[20] H. Wei, M. Gao, A. Zhou, F. Chen, W. Qu, C. Wang, M. Lu, Named entity recognition from biomedical texts using a fusion attention-based BiLSTM-CRF, *IEEE Access*, vol. 7, pp. 73627-73636, 2019.
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems*, pp. 5998-6008, 2017.
[22] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv*, 2018, https://doi.org/10.48550/arXiv.1810.04805
[23] A. Rogers, O. Kovaleva, A. Rumshisky, A primer in bertology: What we know about how bert works, *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 842-866, 2020.
[24] L. Xu, S. Li, Y. Wang, L. Xu, Named entity recognition of BERT-BiLSTM-CRF combined with self-attention, *International Conference on Web Information Systems and Applications*, pp. 556-564, 2021.
[25] Y. Song, S. Tian, L. Yu, A method for identifying local drug names in Xinjiang based on BERT-BiLSTM-CRF, *Automatic Control and Computer Sciences*, vol. 54, no. 3, pp. 179-190, 2020.
[26] L. Gu, W. Zhang, Y. Wang, B. Li, S. Mao, Named entity recognition in judicial field based on BERT-BiLSTM-CRF model, *2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI)*, pp. 170-174, 2020.
[27] Y. Lu, R. Yang, X. Jiang, C. Yin, X. Song, A military named entity recognition method based on pre-training language model and BiLSTM-CRF, *Journal of Physics: Conference Series*, 2020, https://doi.org/10.1088/1742-6596/1693/1/012161
[28] Y. Xiao, K. Zheng, X. Wang, B. Qian, L. Sun, Chinese named entity recognition in electric power metering domain based on neural joint learning, *Journal of ZheJiang University (Science Edition)*, vol. 48, no. 3, pp. 321-330, 2021.
[29] K. Zheng, L. Sun, X. Wang, S. Zhou, H. Li, S. Li, L. Zeng, Q. Gong, Named entity recognition in electric power metering domain based on attention mechanism, *IEEE Access*, vol. 9, pp. 152564-152573, 2021.
[30] R. Li, G. Chen, T. Zhu, Z. Liu, Y. Xiong, Recognizing named entity in electric power documents, *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pp. 652-660, 2020.
[31] H.Y. Zhu, X. Yin, S. Zhang, Z. Wei, M.Y. Hsieh, A discovery method for new words from mobile product comments, *Computer Systems Science and Engineering*, vol. 35, no. 6, pp. 399-410, 2020.
[32] K.W. Church, Word2Vec, *Natural Language Engineering*, vol. 23, no. 1, pp. 155-162, 2017.
[33] R. Cai, B. Qin, Y. Chen, L. Zhang, R. Yang, S. Chen, W. Wang, Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM, *IEEE Access*, vol. 8, pp. 171408-171415, 2020.