# Constructing Bi-Order-Transformer-CRF With Neural Cosine Similarity Function for Power Metering Entity Recognition

**KAIHONG ZHENG**[1], **JINGFENG YANG**[2], **LUKUN ZENG**[1], **QIHANG GONG**[1], **SHENG LI**[1], AND **SHANGLI ZHOU**[1]

[1]Digital Grid Research Institute, China Southern Power Grid, Guangzhou 510663, China
[2]China Southern Power Grid Company Ltd., Guangzhou 510663, China

Corresponding author: Kaihong Zheng (zhengkh@csg.cn)

**ABSTRACT** In recent years, knowledge graphs are applied to provide knowledge support and data support for power grid monitoring and decision-making. To construct a power metering knowledge graph, the power metering entities should be effectively recognized and extracted. However, the existing machine learning models do not fully consider the situation that some power metering entities' names are partially overlapping and boundaries of some power metering entities are fuzzy. In this paper, we propose a Bi-order-Transformer-CRF to recognize power metering entities. Specifically, to alleviate the problem of fuzzy entity boundaries, we train our power metering word-vectors, and then we design Neural Cosine Similarity Function for distinguishing similar entities and Bi-order Feature Extracting Mechanism for recognizing overlapping entity names in the proposed Bi-order-Transformer-CRF. Moreover, we analyze the complexity of the proposed methods and verify that Bi-order-Transformer-CRF achieves better power metering entity recognition results compared with the commonly used machine learning methods in experiments.

**INDEX TERMS** Power metering, entity recognition, neural network, pattern recognition.

## I. INTRODUCTION

With the development of smart power grid, the requirements of analyzing and processing power big data are increasing. At present, lots of power metering information exists in discrete power subsystems, which makes it difficult to obtain effective knowledge for supporting decision-making [1]. It is challenging to integrate the discrete power metering data to build the power knowledge graph and provide effective guidance for supporting decision-making is an urgent problem.

Knowledge graphs describe the concepts, entities, and their relationships in a structured form [2]. The units in a knowledge graph are ''entity-relation-entity'' triples, and the relations of entities are organized in a networked knowledge structure. For the smart power grid, knowledge graphs can be used to provide knowledge support and data support for power grid operation monitoring and decision-making [3]. Before constructing a power metering knowledge

graph, the power metering entities should be recognized and extracted from the power metering texts. Power metering entity recognition aims to identify domain-specific entities and their categories, which plays an important role in metering content analysis and building power metering knowledge graphs. In recent years, lots of research has been done to develop effective methods for power metering entity recognition.

Power metering entity recognition is a kind of Named Entity Recognition (NER), and the early machine learning methods for NER can be classified as statistical machine learning models and methods based on dictionaries or rules. In recent years, deep learning models are widely used in power metering and applications [4], [5]. Some commonly used deep neural networks have been applied for power metering entity recognition, such as Long Short-Term Memory (LSTM) [4], Convolutional Neural Network [6] and Transformer [7]. However, these existing models do not fully consider the situation that some power metering entities' names are partially overlapping and some power entities'

The associate editor coordinating the review of this manuscript and approving it for publication was Meng Huang.

names are very similar, especially in Chinese corpus, the word segmentation process will affect Chinese entity recognition. In the field of electric power metering, an entity may be differently recorded, such as "electric current" and "electric current flow", moreover, the boundaries of some power metering entities are fuzzy and difficult to be defined for power metering entities. For example, "electric current loss" can be regarded as a power phenomenon entity, and "electric current" can be treated as a power object entity.

To alleviate the above questions, from the perspective of data expression and natural language processing, we introduce 4 kinds of power metering entity terms: power metering Index (I), power metering Object (O), power metering Phenomenon (P), and power metering Meter (M) as entity labels for metering reports and texts in both Chinese and English [3], and then we train our power metering word-vectors based on a constructed power dictionary to alleviate the problem of fuzzy entity boundary. From the perspective of building effective data processing models, we propose a Bi-order-Transformer-CRF to handle the problem of overlapping entity names and similar entity names based on the constructed power metering word-vectors. The Bi-order-Transformer-CRF is a specially designed Transformer, which designs Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism. The Neural Cosine Similarity Function is proposed to distinguish similar entities and the Bi-order Feature Extracting Mechanism is proposed to handle the problem of overlapping entity names and to alleviate the influence of word segmentation for Chinese entities. We analyze the complexity of the proposed Bi-order-Transformer-CRF, and in experiments, we verify that the proposed methods are effective for the power metering entity recognition task.

The main contributions of this manuscript can be summarized as follows:

1) To alleviate the problem of fuzzy entity boundaries, we train our power metering word-vectors based on a constructed power metering dictionary and the defined power metering entity terms by using Word2Vec model.

2) To handle the problem of overlapping entity names and alleviate the influence of word segmentation for Chinese entities, we propose a Bi-order-Transformer-CRF for power metering entity recognition task. The Bi-order-Transformer-CRF defines a Neural Cosine Similarity Function and a Bi-order Feature Extracting Mechanism to distinguish similar entities and learn the relations between adjacent entities to judge whether they constitute an entity. Moreover, we also analyze the algorithm complexity of the proposed method.

3) We combine both Chinese and English reports as our dataset and design Ablation Experiments to verify that the proposed Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism are both effective, and the Bi-order-Transformer-CRF achieves

better power metering entity recognition results compared with the commonly used machine learning methods.

This paper is organized as follows: the second part introduces the related work on the task of named entity recognition and power metering entity recognition. In the third part, a Bi-order-Transformer-CRF is proposed, and the effect of the Neural Cosine Similarity Function and a Bi-order Feature Extracting Mechanism is analyzed. The fourth part verifies the effectiveness of the algorithm on experiments. The last part is the conclusions.

## II. RELATED WORKS

Named Entity Recognition (NER) comes of Natural Language Processing task (NLP), and the early machine learning methods for NER can be classified as methods based on dictionaries or rules and statistical machine learning models. Machine learning models improve the NER results, and many early statistical machine learning methods are proposed. Most of early statistical machine learning models use linear statistical models to train the annotated corpus, such as Hidden Markov model (HMM) [8] and Conditional Random Field (CRF) [9], but these statistical models heavily depend on hand-designed features and task-specific training data. These task-specific statistical methods have high development cost and they are difficult to be transferred to new tasks or new fields.

With the development of Neural Networks and Deep learning theory, various kinds of deep learning methods are used to NER in recent years. The representative work mainly includes: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Bi-directional Long Short-Term Memory (Bi-LSTM), which are applied to extract effective features, and then they are combined with CRF to predict tags [10]. These methods extract features automatically and achieve better recognition results compared with the early statistical machine learning methods. Convolutional Neural Networks (CNNs) are widely used in image processing, meanwhile, they are also effective deep machine models for NLP. The CNN can be used in the text classification, and it can be combined with a CRF for the entity recognition task [11]. The CNN-CRF model can be trained end-to-end, which does not depend on resources, functional engineering or data preprocessing of specific tasks. Moreover, CNNs are also combined with some commonly used NER methods to improve their recognition results, such as RNN, LSTM (GRU), and Bi-LSTM [12]. In recent years, Attention Mechanism is proposed and widely used in NLP. Self-Attention can be treated as a kind of Attention Mechanism, which is effective to extract the correlation features of sequence data [13]. Based on Self-Attention Mechanism, Transformer and BERT are proposed, which greatly improve the results in the field of Natural Language Processing. For NER, Transformers are also used in some tasks and achieve satisfactory results [14]. In power metering entities recognition task, Transformers are introduced and combined with the existing effective methods,

such as Transformer-CRF model, Transformer-LSTM-CRF model, Transformer-Bi-LSTM-CRF, and so on [15]–[17].

However, the existing models do not fully consider the situation that names of some certain power metering entities are partially overlapping and parts of power entities have very similar names, especially in Chinese corpus. Moreover, the fuzzy power metering entity boundaries are usually difficult to be defined for power metering entities. For the combined English and Chinese power entity corpus, the Transformer based models and the LSTM based models are not specially designed to distinguish the similar power entities and recognize the entity boundaries according to the commonly used word-vectors.

## III. PROPOSED MODELS

In this paper, we train power metering word-vectors to distinguish entity boundaries and propose a Bi-order-Transformer-CRF to recognize overlapping and similar entity names based on the constructed power metering word-vectors. The Bi-order-Transformer-CRF contains Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism. The Neural Cosine Similarity Function is proposed to distinguish similar entities and the Bi-order Feature Extracting Mechanism is proposed to handle the problem of overlapping entity names and the influence of word segmentation for Chinese entities.

### A. CONSTRUCT POWER METERING WORD-VECTORS
Generally, the entity types of general domain contain people, places, organizations, etc., and their formats are relatively standardized. Although many named entity datasets of general domain are open on the Internet, for power metering field, no public dataset can be directly used for training machine learning models. At present, lots of power metering information exists in isolated power subsystems, which makes it difficult to obtain effective data support for decision-making from this discrete information. Moreover, power metering entity recognition is more complex than that in the general domain, as the fuzzy boundaries of power metering entities are difficult to be defined. For example, "electric current loss" can be regarded as the entity of power phenomenon, meanwhile "electric current" can be treated as a power object entity. "Electric quantity differential anomaly" can be regarded as electric phenomenon entity, meanwhile "electric quantity" can be treated as a power metering index entity and "differential anomaly" is an metering phenomenon entity.

To solve this problem, this paper constructs a textual corpus for power metering. The power metering corpus contains Chinese data and English data. Chinese corpus comes from https: //www.ceppedu.com/ and https://baike.baidu.com/. In addition, cooperative enterprises of China Southern Power Grid provide some corpus information such as business reports and statistical data of power metering. English corpus comes from Wikipedia. Moreover, some English power metering corpus information comes from https://www.se.com/ww/en/download/ and

https://energysaver.nsw.gov.au/. We clean the data, remove the irrelevant information, divide the sentences of the corpus with various punctuation marks, and then divide the corpus into entities.

To alleviate the problem of fuzzy entity boundaries, under the help of the power metering experts, we introduce 4 kinds of power metering entity terms: power metering Index (I), power metering Object (O), power metering Phenomenon (P), and power metering Meter (M) as power metering entities. The details are as follows: the statistical power data is marked as power metering Index entities, such as "power consumption" and "meter reading rate". Mark objects, personnel, regions, and institutions related to power metering as power Object entities, such as "Electric energy meter", "Guangzhou Power Supply Bureau", etc. The phenomenon generated by the specific subject in the process of power metering is labeled as power metering Phenomenon entities, such as "Electric energy meter stops", "Electric current imbalance", etc. A power metering operation that refers to a specific action is marked as a power metering Meter entity, such as "meter reading", "abnormal repair" and so on. In addition, most of the Indexes and Objects of power metering entities are nouns, most of the Phenomena of electric power metering are combinations of noun and verb, and most of metering prefers verbs. The power metering entities are trained as word-vectors by using the Word2Vec model.

### B. CONSTRUCT BI-ORDER-TRANSFORMER-CRF
The Bi-order-Transformer-CRF is specially designed to handle the problem of overlapping entity names and alleviate the influence of word segmentation for Chinese entities. The Bi-order-Transformer-CRF contains a Neural Cosine Similarity Function and a Bi-order Feature Extracting Mechanism. The Neural Cosine Similarity Function is proposed to measure the similarity between the two entities by adjusting the cosine similarity influence between them to distinguish similar entities, and the Bi-order Feature Extracting Mechanism is proposed to handle the problem of overlapping entity names and to alleviate the influence of word segmentation in Chinese texts.

#### 1) NEURAL COSINE SIMILARITY FUNCTION
Self-Attention Mechanism is the core algorithm of Transformer, and similarity functions are the important basis of Self-Attention Mechanism. Therefore, constructing suitable similarity functions for the specific tasks is the key point of building an effective Transformer based model. In the power metering entity recognition task, for similar entity names, the Neural Cosine Similarity Function is proposed to effectively reflect the difference between similar entities, and the constructed Self-Attention is called Neural Cosine Self-Attention layer in this paper.

A Neural Cosine Self-Attention layer contains 3 parts: Neural Cosine Similarity Function, Attention coefficient, and Attention feature. For Neural Cosine Similarity Function, first, we use cosine function to calculate the similarity

between two entities. Cosine similarity measures the similarity between the two input vectors by measuring the cosine value of the angle between them.

$$cosine\,(\theta) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{1}$$

where, subscript $i$ is the $i$-th component, $A$ and $B$ are two vectors. To model different influence caused by the similarity between different entities for specific tasks, this paper defines the Neural Cosine Similarity Function:

$$\begin{aligned} Similarity_{ij} &= Neural\,(Wf_i, Wf_j) \\ &= Neural\,(V \times cosine\,(Wf_i, Wf_j)) \end{aligned} \tag{2}$$

where, $W$ and $V$ are learnable weights, $f_i$ is the $i$-th word-vector, $f_j$ is the $j$-th word-vector, the cosine function is defined as Formula (1), and Neural () is a single layer Neural Network, which uses ReLU activation function to prevent the similarity gradient of Neural Cosine Similarity Function from disappearing.

Based on the defined Neural Cosine Similarity Function, we define Attention coefficient:

$$\alpha_{ij} = softmax\,(Similarity_{ij}) = \frac{\exp\,(Similarity_{ij})}{\sum_i \exp\,(Similarity_{ij})} \tag{3}$$

where, $\alpha_{ij}$ is the Attention coefficient between word-vector $i$ and $j$. According to the Attention coefficient, the Attention feature $f_i$ of the current node $i$ can be expressed as:

$$h_i = sigmoid\left(\sum_j \alpha_{ij} W f_j\right) \tag{4}$$

We aim to embed the Attention features defined by Neural Cosine Similarity Function to a Transformer. Based on Attention features, we introduce an average multi-head attention mechanism to build a Transformer block:

$$h_i = sigmoid\left(\frac{1}{K}\sum_{k=1}^{K}\sum_j \alpha_{ij}^k W^k f_j\right) \tag{5}$$

A Neural Cosine Self-Attention layer can be built following Formula (1)-(5), and several Neural Cosine Self-Attention layers are stacked to extract features of power metering entity word-vectors in this paper, the resulting Deep Neural Network is called Neural Cosine Transformer.

### 2) BI-ORDER FEATURE EXTRACTING MECHANISM

The Bi-order Feature Extracting Mechanism is proposed to handle the problem of overlapping entity names and to alleviate the influence of word segmentation in Chinese texts. In the training process of word-vectors, some adjacent words can be labeled as one power entity, while one of them can be treated as an isolated entity. In Chinese corpus, word segmentation methods may divide a power entity into 2 isolated entities. For example, a word segmentation method may divide "energy meter replacement" into "energy meter" and "replacement", where, "energy meter replacement" is treated as "M" while "energy meter" can be labeled as "O".

To learn the relation of adjacent entities and generate correct labels, we propose a Bi-order Feature Extracting Mechanism.

The Bi-order Feature Extracting Mechanism contains two parts, the first part traverses the trained word-vectors one by one to extract the word features of English power entity words and the segmented Chinese power entity words. The second part traverses the trained word-vectors in pairs to extract the combined features of adjacent word-vectors to alleviated the influence of word segmentation for Chinese entities. The two parts are embedded in two Neural Cosine Transformers respectively to extract the power entity word-vector features.

### 3) CONDITIONAL RANDOM FIELD

The entity recognition problem can be treated as a sequence labeling problem. Considering the correlation between adjacent labels and entities, jointly modeling the label sequence with CRF is an effective entity recognition method. Using BIO method to label the given the training data set, CRF model obtains conditional probability model by maximum likelihood estimation. Given the variable $x$, the CRF calculates the probability $P$ through the corresponding label $y$ based on potential functions:

$$\phi\,(x, y) = \exp\left(\sum_{i=1}^{n} \left(A_{y_{i-1}, y_i} + P_{i, y_i}\right)\right) \tag{6}$$

where, $A_{y_{i-1}, y_i}$ is the score of adjacent word-vectors and their labels, and $P_{i, y_i}$ is the score of $i$-th word-vector and its label. The objective of training the power entity recognition model proposed in this paper is continuously improving the conditional probability $P(y|x)$.

$$P\,(y|x) = \frac{\phi\,(x, y)}{\sum_y \phi\,(x, y)} \tag{7}$$

When the probability approaches 1, the prediction results of the current neural network are consistent with the correct markings, which means that the power entity recognition model has been effectively trained. The structure of proposed Bi-order-Transformer-CRF cane be expressed as Figure 1.

### C. COMPLEXITY ANALYSIS OF THE PROPOSED BI-ORDER-TRANSFORMER-CRF

The Complexity of the proposed Bi-order-Transformer-CRF can be treated as a combination of four parts. The first part is calculating word-vectors, which is related to the size of corpus, and the word-vectors are generated by the Word2Vec model based on genism package. Therefore, the complexity of word-vectors generation can be expressed as $O(K(N)))$, and $K(N)$ is the complexity of Word2Vec that has $N$ words in its corpus. The second part is the computation of the proposed Neural Cosine Self-Attention layer, which introduces an extra single-layer Neural Network for calculating the similarity of entities. Assuming that the complexity of the single-layer Neural Network is $O(P(Num))$, where $Num$ is the number of nodes and $P$ is the complexity of the Neural Network, then the complexity of the Neural Cosine Self-Attention layer $O(C)$ can be expressed as $O(C) = O(S) + O(P(Num))$, where $O(S)$
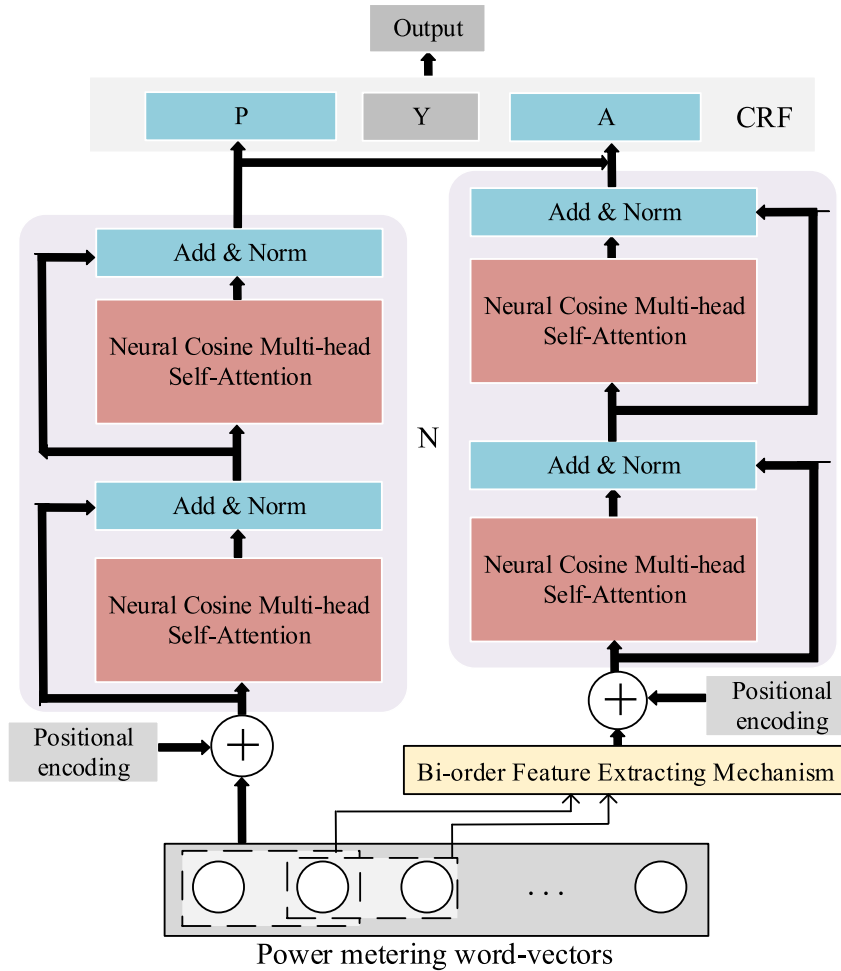
**FIGURE 1.** The structure of the proposed Bi-order-Transformer-CRF.

is the complexity of a Self-Attention layer. The third part is the complexity of the Bi-order Feature Extracting Mechanism, which can be realized by a Convolutional layer, and its complexity can be expressed as $O(Con)$. The fourth part is the complexity of a CRF, $O(CRF)$. The total complexity of the proposed Bi-order-Transformer-CRF can be expressed as follow:

$$O_{algorithm} = O(K(N)) + Layers \times O(C) \\ + (O(con) + Layers \times O(C)) + O(CRF) \quad (8)$$

The complexity of the proposed Bi-order-Transformer-CRF is a little higher than traditional Transformer-CRF model, but the increased complexity is constant. In experiments, we introduce Ablation Experiments to verify the proposed Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism are both effective, and the Bi-order-Transformer-CRF achieve comparable or even better power metering entity recognition results compared with the commonly used machine learning methods.

## IV. EXPERIMENTS

Experiments are divided into 3 parts, first, we construct power metering word-vectors, and then we show that the Bi-order-Transformer-CRF achieve better power metering entity recognition results compared with the commonly used machine learning methods mentioned in Introduction and Related works. Lastly, we discuss the different dimensional word-vectors, and ablation experiments are done to verify the proposed Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism are both effective.

## A. WORD-VECTORS CONSTRUCTION AND EVALUATION INDICATOR

The power metering corpus contains Chinese data and English data. Chinese corpus comes from https: //www. ceppedu.com/ and https://baike.baidu.com/. In addition, cooperative enterprises of China Southern Power Grid provide some corpus information such as business reports and statistical data of power metering. English corpus comes from Wikipedia. Moreover, some English power metering

corpus information comes from https://www.se.com/ww/en/download/ and https://energysaver.nsw.gov.au/. We clean the data, remove the irrelevant information, divide the sentence structure of the corpus with various punctuation marks, and then divide the corpus into entities. At the same time, because the batch segmentation is based on sentences, it is necessary to split the long sentences. Finally, 16454 sentences are constructed.

For the power metering entity data, under the guidance of experts and their entity classification mechanism in the field of power grid, this paper uses BIO method to mark the power metering entities as four categories I, O, P, and M. For Chinese entities, "YEDDA" is used for labeling, and based on the visual interface, by selecting the text of the entity part to be recognized and using quick annotation key, a large number of corpus can be annotated efficiently, and its entity recommendation function can significantly reduce the manual errors. The number of various entities is shown in Table 1.

**TABLE 1.** Statistics of corpus entity count.

| Entity type | Entity number |
|---|---|
| Index | 5251 |
| Object | 6039 |
| Phenomenon | 2934 |
| Meter | 4593 |

As table 1 shows, the dataset is unbalanced. For the power metering entities recognition task, we need to train our power metering word-vectors. For the English word vector, the "genism" is used to train two million entries in "Baidu baike" and 1000 metering terms in "An English-Chinese and Chinese-English Technical Dictionary of Electric Power Engineering". For Chinese word vector, the data includes the two million entries in "Baidu baike", "An English-Chinese and Chinese-English Technical Dictionary of Electric Power Engineering", and the metering terms confirmed by experts in 20 power metering reports of China Southern Power Grid. The mentioned Chinese data is used as the self-definition Dictionary of "jieba" word segmentation tool. These English data and Chinese data are trained as power metering word-vectors by using Word2Vec model with "gensim."

To evaluate the trained models, we introduce $F1$-score which is defined according to Precision and Recall. We use $F1$-score as the metric for evaluating the performance of the models for the following reason: the power metering dataset is skewed towards "Index" and "Object", and a classifier that correctly predicts the majority class can get a good accuracy, while F1-score balances the classes and considers Precision as well as Recall, which is more suitable for the power metering entity recognition task in this paper. In this paper, we also use micro-averaged indicators named $F1micro$ and macro-averaged $F1macro$ scores. The evaluating

indicators are defined as follows:

$$Precision_{micro} = \frac{\sum_{k=1}^{c} tp_k}{\sum_{k=1}^{c} tp_k + \sum_{k=1}^{c} fp_k} \quad (9)$$

$$Precision_{macro} = \frac{\sum_{k=1}^{c} Precision_{micro}}{c} \quad (10)$$

$$Recall_{micro} = \frac{\sum_{k=1}^{c} tp_k}{\sum_{k=1}^{c} tp_k + \sum_{k=1}^{c} fn_k} \quad (11)$$

$$Recall_{macro} = \frac{\sum_{k=1}^{c} Recall_{micro}}{c} \quad (12)$$

$$F1_{macro} = \frac{2 * Precison_{macro} * Recall_{macro}}{Precison_{macro} + Recall_{macro}} \quad (13)$$

where, $tp_k$ (true positives) refer to the number of entities which are correctly classified in class $k$, $fp_k$ is the number of entities which are incorrectly classified in class $k$, and $fn_k$ is the number of entities that belong to class $k$ but were not classified as such.

The main parameters of the proposed models are listed in Table 2.

## B. COMPARATIVE EXPERIMENTS AND ANALYSIS

In comparative experiments, we add some commonly used methods in the technical literature for comparison, and the results of the proposed model and the commonly used models are listed in Table 3.

The most commonly used NER models are LSTM based models and Transformer based models. As table 3 shows, the proposed Bi-order-Transformer-CRF achieves the best result compared with the Transformer based models and LSTM based models. Compared with CRF model, the $F1$ score of the Bi-order-Transformer-CRF is increased by 0.154, which shows that the Deep Neural Network has advantages in power metering entity recognition, and the proposed deep learning model performs better than this traditional machine learning method.

Compared with the LSTM and Bi-LSTM based models, the $F1$ score of the Bi-order-Transformer-CRF is increased by 0.063 and the Transformer-CRF achieves higher $F1$ score than the CNN-Bi-LSTM-CRF, which shows that the Transformer has advantages in power metering entity recognition compared with LSTM. Moreover, compared with the Transformer based models, the $F1$ score of the Bi-order-Transformer-CRF is increased by 0.005. To further analyze the performance of the proposed Bi-order-Transformer-CRF, we generate the results for four categories. The confusion matrix and the normalized confusion matrix are shown as Table 4 and Figure 2.

The confusion matrix for the power entity recognition task shows the error distribution of the four categories. As we can see from the Figure 2, although the proposed model correctly recognizes most entities, it is more likely to misclassify the Index entities and the Meter entities. Moreover, the number of Phenomenon entities is minimal, and the performance of the proposed Bi-order-Transformer-CRF on recognizing

**TABLE 2.** The main parameters in the proposed model.

| Learning rate | word-vectors dim | Optimizer | Batch size | Num of Blocks | Transformer embedding |
|---|---|---|---|---|---|
| 1e-5 | 100 | Adam | 32 | 6 | 500 |

**TABLE 3.** The test results of the proposed model and other commonly used models.

| No. | Models | P | R | Micro F1 |
|---|---|---|---|---|
| 1 | CRF [9] | 0.720 | 0.732 | 0.726 |
| 2 | HMM [8] | 0.709 | 0.711 | 0.710 |
| 3 | CNN [6] | 0.723 | 0.752 | 0.737 |
| 4 | CNN-CRF | 0.759 | 0.776 | 0.767 |
| 5 | LSTM-CRF | 0.791 | 0.799 | 0.795 |
| 6 | Bi-LSTM-CNN [6] | 0.801 | 0.802 | 0.801 |
| 7 | Bi-LSTM-CRF [4] | 0.808 | 0.803 | 0.805 |
| 8 | CNN-Bi-LSTM-CRF [3] | 0.816 | 0.819 | 0.817 |
| 9 | Transformer-CRF [15] | 0.868 | 0.869 | 0.869 |
| 10 | Transformer-LSTM-CRF [16] | 0.871 | 0.876 | 0.873 |
| 11 | PBERT-Bi-LSTM-CRF [14] | 0.874 | 0.878 | 0.876 |
| 12 | Transformer-Bi-LSTM-CRF [17] | 0.875 | 0.877 | 0.876 |
| 13 | **Bi-order-Transformer-CRF** | **0.880** | **0.880** | **0.880** |

**TABLE 4.** The confusion matrix of the Bi-order-Transformer-CRF.

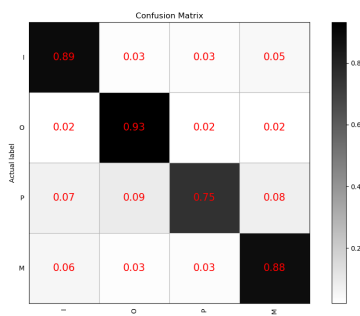| | Predict | | | |
|---|---|---|---|---|
| **Truth** | Index | Object | Phenomenon | Meter |
| Index | 4674 | 152 | 171 | 254 |
| Object | 128 | 5627 | 147 | 137 |
| Phenomenon | 218 | 272 | 2207 | 237 |
| Meter | 251 | 139 | 146 | 4057 |



**FIGURE 2.** The normalized confusion matrix of Table 4.

such entities is worse than others. To analyze this situation quantitatively, we calculate the $F1$ score on four categories, and the results are shown in Table 5.

As Table 5 shows, the $F1$ scores of Index, Object, and Meter are higher than that of Phenomenon, which shows that the imbalance of data sets affects the final recognition results. As mentioned above, the Bi-order-Transformer-CRF contains Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism. To verify that the proposed two methods in the Bi-order-Transformer-CRF improve the recognition results, we add ablation experiments.

### C. ABLATION EXPERIMENTS

In next experiment, we replace the Neural Cosine Similarity Function in the Bi-order-Transformer-CRF with other commonly used similarity functions and test the entity recognition performance, the $F1$ scores are shown in Table 6.

As we can see in Table 6, the Bi-order-Transformer-CRF with commonly used similarity functions performs better than the traditional Transformer-CRF, which verifies that the Bi-order Feature Extracting Mechanism is effective. Moreover, the Bi-order-Transformer-CRF with Neural Cosine Similarity Function achieves the best recognition results, which verify that Neural Cosine Similarity Function is effective.

For the Bi-order Feature Extracting Mechanism, we also test higher order Feature Extracting Mechanism and their combinations. The test $F1$ scores are shown in Table 7. In this experiment, all models use the proposed Neural Cosine Similarity Function, so the Transformer-CRF in this experiment is litter different from that in above experiments.

**TABLE 5.** The F1 scores on four categories of the Bi-order-Transformer-CRF.

| Entity type (Bi-order-Transformer-CRF) | P | R | F1 |
|---|---|---|---|
| Index | 0.887 | 0.890 | 0.888 |
| Object | 0.909 | 0.932 | 0.920 |
| Phenomenon | 0.826 | 0.752 | 0.786 |
| Meter | 0.866 | 0.883 | 0.875 |

**TABLE 6.** The *F*1 scores on 5 commonly used similarity functions.

| No. | Similarity function | P | R | Micro F1 |
|---|---|---|---|---|
| 1 | Cosine | 0.871 | 0.875 | 0.873 |
| 2 | Neural Networks | 0.876 | 0.878 | 0.877 |
| 3 | Dot Product | 0.868 | 0.872 | 0.870 |
| 4 | Concat function | 0.869 | 0.874 | 0.871 |
| 5 | Neural Conise | 0.880 | 0.880 | 0.880 |

**TABLE 7.** Test *F*1 scores of combinations on different order feature extracting mechanism.

| No. | Order combination | P | R | Micro F1 |
|---|---|---|---|---|
| 1 | Transformer-CRF | 0.875 | 0.876 | 0.875 |
| 2 | Bi-order-Transformer-CRF | 0.880 | 0.880 | 0.880 |
| 3 | 1+2+3-Transformer-CRF | 0.879 | 0.882 | 0.880 |
| 4 | 1+2+3+4-Transformer-CRF | 0.880 | 0.882 | 0.881 |
| 5 | 1+3-Transformer-CRF | 0.878 | 0.877 | 0.877 |

As Table 7 shows, the $1 + 2 + 3 + 4$-Transformer-CRF with the combination of Neural Cosine Transformer, Bi-order Transformer, 3-order Transformer and 4-order Transformer achieves the best test $F1$ score. However, the complexity of this model is higher and the improvement is very limited. Therefore, we use the Bi-order-Transformer-CRF as our final model.

The last experiment shows the test $F1$ on different dimensional word-vectors, and the results in Table 8 show the reason why we use 100-dim word-vectors.

**TABLE 8.** Test F1 scores on different dimensional word-vectors.

| No. | Word vectors | P | R | Micro F1 |
|---|---|---|---|---|
| 1 | Word2Vec-50 | 0.874 | 0.877 | 0.875 |
| 2 | Word2Vec-100 | 0.880 | 0.880 | 0.880 |
| 3 | Word2Vec-200 | 0.877 | 0.879 | 0.878 |
| 4 | Word2Vec-400 | 0.878 | 0.879 | 0.878 |

As above experiments show, the proposed Bi-order-Transformer-CRF improves the results of the power metering entities recognition task. We compare the proposed model with the commonly used machine learning models mentioned in the Introduction and Related works. Moreover, we verify that the designed Neural Cosine Similarity Function and the Feature Extracting Mechanism are both effective for the power metering entities recognition task.

## V. CONCLUSION

In this paper, to handle the problem of overlapping entity names and alleviate the influence of word segmentation, we propose a Bi-order-Transformer-CRF for power metering entity recognition task. The Bi-order-Transformer-CRF defines a Neural Cosine Similarity Function and a Bi-order Feature Extracting Mechanism to distinguish similar entities and learn the relation between adjacent entities to judge whether they constitute an entity. Moreover, we analyze the algorithm complexity of the proposed method. In experiments, we verify the proposed Neural Cosine Similarity Function and Bi-order Feature Extracting Mechanism are both effective, and the Bi-order-Transformer-CRF achieve comparable or better power metering entity recognition results compared with the commonly used machine learning methods.

In our research, we find that the word segmentation method affects the Chinese word-vectors construction, and thus we design a Bi-order Feature Extracting Mechanism to alleviate this problem, but it also brings higher computational complexity. If an accurate Chinese word segmentation method can effectively segment words according to the predesigned entity terms, the power entity recognition task may be handled more effectively, and it is our direction for future research.

## REFERENCES

[1] J. Li and K. Chen, "Building power knowledge map based on graph database," in *Proc. 2nd Smart Grid Conf.*, Beijing, China, 2018, pp. 77–81.

[2] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, 2017.

[3] Y. Xiao, "Chinese named entity recognition in electric power metering domain based on neural joint learning," *J. Zhejiang Univ. (Sci. Ed.)*, vol. 48, no. 3, pp. 321–330, 2021.

[4] Z. Ji, X. Wang, C. Cai, and H. Sun, "Power entity recognition based on bidirectional long short-term memory and conditional random fields," *Global Energy Interconnection*, vol. 3, no. 2, pp. 92–98, 2020.

[5] C. Nogueira dos Santos and V. Guimarães, "Boosting named entity recognition with neural character embeddings," 2015, *arXiv:1505.05008*. [Online]. Available: http://arxiv.org/abs/1505.05008

[6] J. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Trans. Assoc. Comput. Linguistics*, vol. 4, pp. 357–370, Apr. 2016.

[7] H. Yan, B. Deng, X. Li, and X. Qiu, "TENER: Adapting transformer encoder for named entity recognition," 2019, *arXiv:1911.04474*. [Online]. Available: http://arxiv.org/abs/1911.04474

[8] S. Morwal, "Named entity recognition using hidden Markov model (HMM)," *Int. J. Natural Lang. Comput.*, vol. 1, no. 4, pp. 15–23, Dec. 2012.

[9] A. Goyal, V. Gupta, and M. Kumar, "Recent named entity recognition and classification techniques: A systematic review," *Comput. Sci. Rev.*, vol. 29, pp. 21–43, Aug. 2018.

[10] X. Dong, S. Chowdhury, L. Qian, Y. Guan, J. Yang, and Q. Yu, "Transfer bi-directional LSTM RNN for named entity recognition in Chinese electronic medical records," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Oct. 2017, pp. 1–4.

[11] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Trans. Assoc. for Comput. Linguistics*, vol. 4, pp. 357–370, Dec. 2016.

[12] S. Thattinaphanich and S. Prom-on, "Thai named entity recognition using Bi-LSTM-CRF with word and character representation," in *Proc. 4th Int. Conf. Inf. Technol. (InCIT)*, Oct. 2019, pp. 149–154.

[13] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1–11.

[14] C. Jiang, Y. Wang, and J. Hu, "Power entity information recognition based on deep learning," *Power Syst. Technol.*, vol. 45, no. 6, pp. 2141–2149, Jun. 2021.

[15] H. Zhuang, F. Wang, S. Bo, and Y. Huang, "A BERT based Chinese named entity recognition method on ASEAN news," *in Proc. J. Phys., Conf.*, Apr. 2021, vol. 1848, no. 1, Art. no. 012101.

[16] T. P. Adewumi, F. Liwicki, and M. Liwicki, "Exploring Swedish & English fastText embeddings for NER with the transformer," 2020, *arXiv:2007.16007*. [Online]. Available: http://arxiv.org/abs/2007.16007

[17] J. Luoma and S. Pyysalo, "Exploring cross-sentence contexts for named entity recognition with BERT," 2020, *arXiv:2006.01563*. [Online]. Available: http://arxiv.org/abs/2006.01563

**JINGFENG YANG** received the master's degree from the South China University of Technology, China, in 2009. He is currently a Senior Engineer with China Southern Power Grid. His research interests include electrical energy measurement automation systems, electricity information acquisition systems, electric energy measurement, and big data analysis technology.

**LUKUN ZENG** was born in 1992. He received the B.S. degree in communication engineering and the Ph.D. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2013 and 2019, respectively. He is currently working with Digital Grid Research Institute, China Southern Power Grid. His research interests include electrical energy measurement automation systems and data mining technology.

**QIHANG GONG** received the master's degree from Zhejiang University, China, in 2020. He is currently an Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

**SHENG LI** received the master's degree from Hunan University, China, in 2019. He is currently an Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

**KAIHONG ZHENG** received the master's degree from Zhejiang University, China, in 2016. He is currently an Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

**SHANGLI ZHOU** received the bachelor's degree from Tsinghua University, China, in 1991. He is currently a Professor level Senior Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

• • •