# Fine-Grained Drug Interaction Extraction Based on Entity Pair Calibration and Pre-Training Model for Chinese Drug Instructions

Xiaoliang Zhang, The First Affiliated Hospital of Nanjing Medical University, China

Feng Gao, Wuhan University of Science and Technology, China

https://orcid.org/0000-0002-2396-1360

Lunsheng Zhou, Wuhan University of Science and Technology, China

Shenqi Jing, Nanjing Medical University, China

Zhongmin Wang, Nanjing Medical University, China

Yongqing Wang, Nanjing Medical University, China

Shumei Miao, Nanjing Medical University, China

Xin Zhang, Nanjing Medical University, China

Jianjun Guo, Nanjing Medical University, China

Tao Shan, Nanjing Medical University, China

Yun Liu, Nanjing Medical University, China*

https://orcid.org/0000-0002-2396-1360

## ABSTRACT

Existing pharmaceutical information extraction research often focus on standalone entity or relationship identification tasks over drug instructions. There is a lack of a holistic solution for drug knowledge extraction. Moreover, current methods perform poorly in extracting fine-grained interaction relations from drug instructions. To solve these problems, this paper proposes an information extraction framework for drug instructions. The framework proposes deep learning models with fine-tuned pre-training models for entity recognition and relation extraction. In addition, it incorporates an novel entity pair calibration process to promote the performance for fine-grained relation extraction. The framework experiments on more than 60k Chinese drug description sentences from 4000 drug instructions. Empirical results show that the framework can successfully identify drug related entities (F1 ³ 0.95) and their relations (F1 ³ 0.83) from the realistic dataset, and the entity pair calibration plays an important role (~5% F1 score improvement) in extracting fine-grained relations.

## KEYWORDS

Chinese Drug Instruction, Entity Pair Calibration, Fine-Grained, Information Extraction, Pre-Training Model

*Corresponding Author

## INTRODUCTION

Drug instructions contain a wealth of drug knowledge, which can be extracted for decision support in clinical diagnosis, prescriptions, and healthcare management. However, mining drug instruction knowledge is a challenging task. Textual descriptions in drug instructions usually take the form of long, complicated sentences, which are difficult to be processed by man or machine. Existing information extraction methods have been used for drug and disease entity recognition (Lin, & Xie, 2020; Sun et al., 2021; Zhu et al., 2021), drug-disease relationship extraction (Bose, et al., 2021; Fatehifar, & Karshenas, 2021; Mingliang, Jijun, & Fei, 2021), etc. Recently, pre-training models have become prominent in Natural Language Processing (NLP) tasks due to its generalization capability (Vaswani et al., 2017). However, many existing methods still consider the knowledge extraction task as a pipeline process that consists of Named Entity Recognition (NER) and Relation Extraction (RE) tasks. These pipeline methods are prone to error propagation since not all entities generated from NER are valid for RE (Pawar, Palshikar, & Bhattacharyya, 2017). In drug instructions, it is very common to have multiple entities and relations in a single sentence, without proper entity pair checking, the performance of fine-grained relation extraction will be significantly affected by the invalid entity pairs (Gao, Zhou, & Gu, 2021). Siriwon Taewijit, & Thanaruk Theeramunkong (2021) proposed an approach to learn meaningful patterns by hyperbolic embedding and then extract adverse drug reactions from electronic medical records. As a result, manual effort is required to prepare entity pairs. In other cases, joint methods are proposed to extract entities and relations in a single process (Wei et al., 2020), however, these methods struggle with many-to-many relations that involve overlapping entities (Tiktinsky et al., 2022).

To overcome the limitations in the current approaches, this paper proposes a novel pipeline drug information extraction framework using an entity pair calibration module based on pre-training models. It starts by automatically gathering the textual descriptions from public drug instruction data sources. Then, it uses deep learning models to identify four categories of entities: drugs, diseases, body parts, and symptoms. After the entity name recognition, it gathers the sentences with multiple relevant entities and applies the entity pair calibration (EPC) process. The goal of EPC is to distinguish between the main entity (the primary drug which the instruction is intended for, e.g., the first drug mention in Table 1) and the secondary entities (the other entities mentioned in the sentence of the instruction, potentially associated with the primary drug entity, e.g., the second, third and fourth drug mentions in Table 1). This step can reduce the noise for further drug relation extraction and facilitate accurate extraction of fine-grained relations. Finally, it extracts fine-grained relations from those entity pairs.

The main contribution of this paper is two-fold:

1. A pre-training-based drug information extraction framework with a novel entity pair calibration process is proposed to extract drug-related entities and fine-grained relations from drug instructions.
2. To evaluate, an empirical study is carried out over a realistic dataset that contains over 60,000 sentences from 4,000 drug instructions.

The remainder of the paper is organized as follows. First, the related works in the field of drug information extraction is discussed. Second, the overall framework proposed in this paper is presented, including functionalities of each module and the design of the deep learning models used in the

Table 1. Sample drug instruction text with multiple entities and relations

| |
|---|
| Compound **paracetamol** should not be used together with **chloramphenicol**, **barbiturates** (such as **phenobarbital**), etc. |

modules. Third, the details of entity pair calibration process are elaborated on. The empirical study and result analysis is then presented, and then the conclusion and a discussion of future directions.

## RELATED WORKS

### Pipeline Methods

Named Entity Recognition (NER) is a well-established topic in NLP. Fang et al. (2002) proposed a method of named entity recognition based on rules, but this method relies heavily on predefined rules and is prone to error causing incomplete/wrong rule sets. Neural network models, such as BiLSTM (Miwa, & Bansal, 2016) have proven efficient in NER and RE tasks. Zhou et al.(2016) proposed Attention-Based Bidirectional Long Short-Term Memory Networks(AttBLSTM) to capture the most important semantic information in a sentence. Gong Lejun (2020) used a drug entity extraction method based on GRU and CNN for detecting adverse drug reactions (ADRs) and achieved 75% overall accuracy in the DDIE 2013 corpus (Herrero-Zazo et al., 2013). Beliga, Meštrović, and Martinčić-Ipšić (2016) proposed a novel Selectivity-Based Keyword Extraction (SBKE) method, which extracted keywords from the source text represented as a network. Wang Yongchao (2021) proposed an entity extraction method based on a pointer network, but the pointer-based entity annotation method does not provide for nested names.

Deep learning methods are also widely used in relation extraction Zeng et al. (2014) introduced convolutional neural networks to the relationship extraction task. However, CNNs cannot capture long-term semantic information in sentences. In Shen et al. (2015), an unsupervised hybrid framework called REACTOR was proposed to identify various types of relations among entities appearing automatically in enterprise data. Li and Ke (2018) found that recurrent neural networks (RNN) are good at capturing the contextual information of sentences and can reflect the higher order relationships of sentences, but RNNs cannot efficiently capture features across large time spans. Ningthoujam et al. (2019), demonstrates that Long and Short Term Memory networks (LSTM) have a better extraction performance on relation extraction. In Jagan, Parthasarathi, and Geetha (2019), a bootstrapping approach was proposed to extract generic relations that exist between different components of a sentence. Zhang, Du, Jia, and Li (2021) used a combination of BiLSTM and CNN for medical entity relationship extraction. Srinivasan and L D, (2019) proposed a twofold convolutional neural network approach with a new activation function, which generalizes faster with higher accuracy. Lv et al. (2020) proposed a distant supervised relation extraction model (DiSAN-2CNN), in which a multi-dimension self-attention mechanism is utilized to encode the features of the words together with the entire sentence. Siriwon Taewijit, & Thanaruk Theeramunkong (2021) found that DL can be very helpful for diagnosing neurological disorders by its symptoms, because DL can be used to identify patterns for each disorder for identification.

### Joint Methods

To avoid error propagation in pipeline methods, Bekoulis et al. (2018) transformed entity relationship extraction into a multi-headed selection problem. In order to solve the problem of overlapping relations, Li et al. (2019) proposed to treat the task of joint entity-relationship extraction as a multi-round question-and-answer type problem, but the model is more suitable for machine reading comprehension. Wei et al. (2020) proposes to model relationships as functions that map head entities to tail entities, rather than treating them as labels on pairs of entities.

Most joint entity-relationship extractions are still supervised learning approaches that require a large amount of high-quality training set (Lv et al., 2020). Recent research tried to deal with the problem of having a relatively small training sample size. Wang & Hebert (2016) demonstrated that using model regression networks have good generalization capabilities from small sample sizes, but it still requires prior category learning. Bao et al. (2019) have successfully classified English sentences

using the global distributional features of words over a small sample size, however this approach is language-specific and is difficult to transfer between languages. The MTB model proposed by Google (Soares et al., 2019) uses a metric learning approach and achieves good results on top of the small sample dataset.

## Pre-Training Models

Recently, attention mechanisms have proven efficient in NLP tasks. Vaswani et al., (2017) first proposed an attention-based transformer model, which can be decoupled from the traditional neural network structures such as CNN and RNN. This enables the use of transformers as a generic module across different NLP tasks. The bidirectional encoder representation from transformers (BERT) (Devlin et al., 2018) is developed based on the transformers, which pre-trains sentences by mask language model (MLM) and next sentence prediction (NSP). In order to extract to the fullest extent, the lexical, syntactic and semantic information from training corpora, Sun et al. (2020) propose a continual pre-training framework named ERNIE 2.0 which builds and learns incrementally pre-training tasks through constant multi-task learning.

Several efforts have been made to improve BERT. R-BERT (Wu & He, 2019) injects entity type information into the pre-trained model and can obtain high quality relation extraction results, proving the importance of entity information enhancement in relation extraction. Roberta (Y. Liu et al., 2019) proposes to use a dynamic mask instead of a static mask and can further improve the performance of the model.

The original BERT model is designed for English corpus. For Chinese corpus, the word-based "Mask" will lose some semantics. The BERT(wwm) (Cui et al., 2021) model uses a whole word mask and is more effective for Chinese text. To identify the whole word, a word segmentation tool is used to segment the text, and then the whole word segment is marked and masked, with the help of several other optimizations. The experimental results on the public data set CMRC 2018 showed that BERT(wwm) outperformed the original BERT. In this paper, the authors use BERT(wwm) as a basic model for the drug information extraction framework.

## Summary of Current Approaches

A comparison of current research strands in text-based knowledge extraction, (including NER and RE), are shown in Table 2. Rule-based models can achieve good accuracy, but the recall is typically limited due to the requirements of manual rule definitions. Pipeline methods can achieve relatively good results in NER and coarse-grained RE, but suffer from error propagation and hence cannot accurately extract fine-grained relations. Joint entity and relation extraction can avoid error propagation, but requires a large amount of training sets, and it struggles in extracting multiple relations with overlapping entities. Pre-training models can be used to enhance both pipeline and joint models, and can be used in multiple NLP tasks. In the authors' approach, they proposed a pre-training-based pipeline model with a novel entity pair calibration process, which eliminates errors caused by incorrect entity pairs.

## DRUG INFORMATION EXTRACTION FRAMEWORK

The drug information extraction framework proposed in this paper uses pre-training models as a key method to label sentences as well as entities and relations in sentences. In the following, the authors first present the definition of the problem. Then, they discuss the architecture of the framework and introduce the functions of its modules. Finally, they elaborate on the design choices of the pre-training-based neural models used in the framework.

**Table 2. Summary of current approaches**

| Category | Sample Approaches | Strength | Weakness |
|---|---|---|---|
| Rule-based | (Fang, & Sheng, 2002) | Good accuracy, easy implementation. | Poor recall due to manual rule definition requirement |
| Pipeline | (Zhang et al., 2021) | Good performance in NER and coarse-grained RE. | Poor performance in fine-grained RE due to noisy entity pairs, need to train models separately. |
| Joint Models | (Wei et al., 2020), (Soares et al., 2019), (Zheng et al., 2021) | Good performance in NER and RE, models trained at once | Not suitable for overlapping/n-ary relations, needs large training sets. |
| Pre-Training Models | R-BERT (Wu & He, 2019), Roberta (Liu et al., 2019), BERT-CLS (Devlin et al., 2018) | Can be used to enhance pipeline and joint models, can be used in different tasks | Language specific, needs large corpus, needs fine-tuning. |
| Pipeline with EPC | Our Approach | Uses pre-training over pipeline methods with a novel entity pair calibration, reduces noise in RE | Overhead for entity pair calibration training, beneficial only in multi-entity and relation extraction. |

## Problem Definition

Given the sentence $S = \langle t_1, t_2, \dots t_n \rangle$ as a sequence of tokens, a knowledge extraction framework F:S → T creates triples from the sentence, where T is a set of triples. $\forall t \in T, t = \langle s, p, o \rangle$ where $s, \circ \in E$, E is the set of tokens $\{t_i, \dots t_j\} \subseteq S$ referring to entities in the sentence, $p \in R$ refers to the relations among entities. A named entity cognition task $T_{ner} : S \rightarrow E$ labels tokens from the sentence with entity types, a relation extraction task $T_{re} : E \times E \rightarrow T$ labels entity pairs with relation types, hence creating the set of triples. Therefore, resulting in the pipeline extraction framework $F_P : T_{ner} \circ T_{re}$

An observation that was made from drug instructions is that a single sentence can have multiple entities and relations, as shown in Table 1. However, not all of the entities have relations. Hence, the cardinality of the input of $T_{re}$ (i.e., $|E \times E|$) is significantly higher that its output (i.e., $|T|$). This introduces a lot of noise in relation extraction and could lead to an imbalance in the samples and lower the extraction precision.

Another observation that was made is that some drug relations always appear over two different entity roles, e.g., drug-drug interactions always apply between the primary drug and the secondary drugs in the sentence. Sometimes the relation only appears in the same entity role, e.g., there is a hypernym relation between the third and fourth drug mentions in Table 1, and the hypernym relation mostly appears in same entity roles. A natural idea is to use this information and add an entity pair calibration task $T_{epc} : E \rightarrow (E_p | E_s)$, where $E_p$ and $E_s$ refer to primary and secondary entities, respectively, and the relation extraction task is modified into learning specific relations according to the entity roles received. This way, the EPC enabled pipeline extraction framework $F_{p\_epc} : T_{ner} \circ T_{epc} \circ T_{re}$ will receive much less noise during relation extraction, and its overall information extraction performance can be enhanced.

## Architectural Design

The proposed framework has four main modules: data collection and cleaning, named entity recognition, entity pair calibration, and entity relation extraction, as depicted in Figure 1. The

**Figure 1. Overall framework of information extraction for drug instructions**
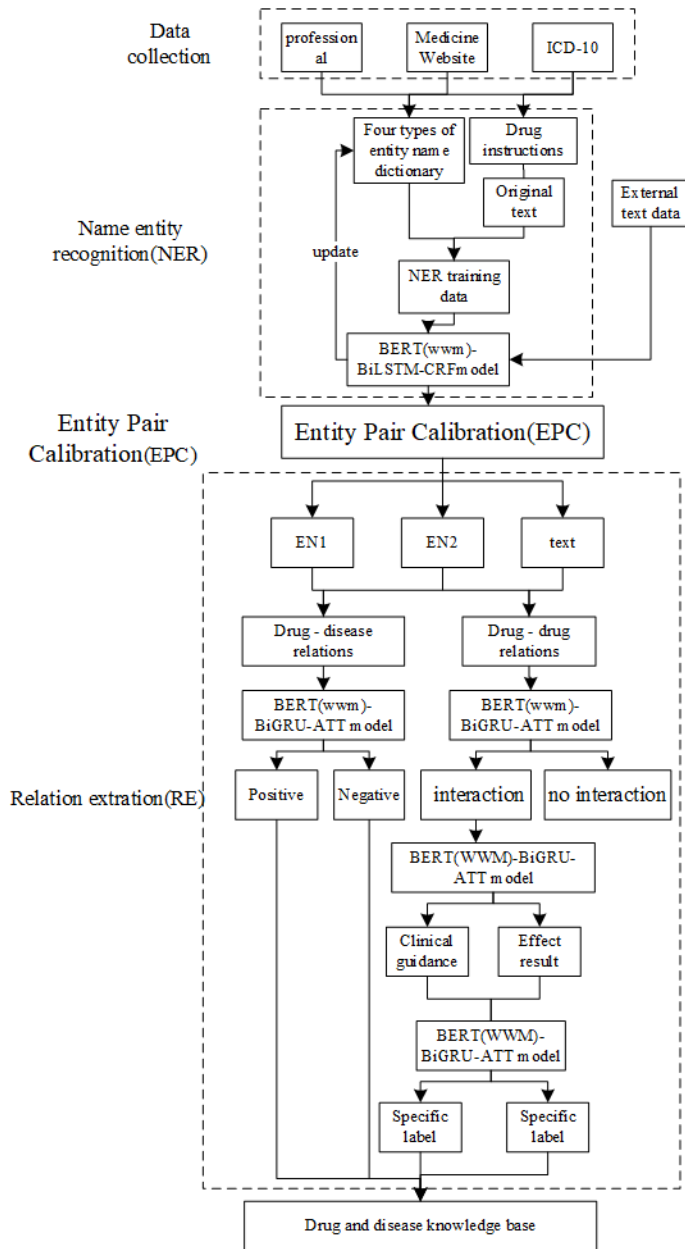


data collection is an automatic process that gathers drug instructions and organizes them into semi-structured text inputs. The entity recognition and entity pair calibration are semi-supervised learning modules that identifies valid entity pairs for the relation extraction module, which is a supervised learning module. The NER and RE modules are similar to existing pipeline methods (e.g., Zhang et al., 2021), with the improvement of using the BERT-wwm (Cui et al., 2021) as the pre-training model, whereas the EPC module is a novel module proposed in this paper that intermediates NER and RE.

### Data Collection and Cleaning

The data collection module automatically collects drug instructions, disease diagnosis, treatment guidelines, and other medical text data from public sources on the Internet. The authors carry out this operation regularly with a careful selection of authoritative data sources in order to provide rich and reliable data sources for drug information extraction. The structure of the collected and cleaned data is stored as spreadsheets, all textual descriptions of a drug instruction is retained and categorized into columns using the section titles that appear in the instructions, e.g., product name, primary chemical component, producer, drug reaction, drug interaction, etc.

### Named Entity Recognition

The Named Entity Recognition (NER) module is mainly responsible for identifying four kinds of named entities from the text, which are "drugs," "diseases," "symptoms" and "body parts," thus preparing datasets for further relationship extraction. This module uses an entity name dictionary as a starting point and applies a BERT(wwm)-BiLSTM-CRF deep learning model to learn new names. The entire procedure consists of three steps: First, the original sentences in the instruction are converted to the format required by the training dataset for the named entity recognition algorithm model. Second, the sentences are processed with the longest string matching method with the limited dictionary names collected manually in the early stage. The longest string matching method can effectively avoid the wrong annotation of nested entity names in the dictionary. Third, the deep learning model is trained by the dataset to generalize entity names; generalized entity names are added to the dictionary, so that the training data set of the model can be updated regularly, thus achieving a self-improvement solution.

### Entity Pair Calibration

When drug and disease entities have been extracted from sentences, the authors can select sentences with n (n> = 2) entities to further investigate entity relations. Since the authors are trying to learn binary relations, those n entities may produce $C\left(n,2\right)$ entity pairs. However, not all of those entity pairs are legitimate for relation extractions, e.g., the same drug/disease may have different names, occurring at different locations in the same sentence, or there may be a type-of/sub-class-of relation between them. In these cases, the entities pairs cannot enter the drug/disease interaction extraction module without causing noise. To reduce noisy data caused by wrong entity pairs, the Entity Pair Calibration (EPC) module is used to verify the primary entity (the main drug described by the drug instruction, i.e., the intended object for the instruction) and the secondary entity (the other drugs/diseases described in the sentence). In this paper, the authors use a modified word embedding method to splice the entity names with the original sentences and "Mask" entity names in the sentences. Finally, the masked text is fed into a BERT(wwm)-BiGRU-ATT model to train the entity pair calibration module so that it can identify the categories of entities in the text and remove entity pairs with unintended type-of/sub-class-of relations.

### Relationship Extraction

The Relationship Extraction (RE) module extracts the relationship between the entity pair and its text information from the entity pair calibration module, which is mainly realized by the BERT(wwm)-BiGRU-ATT model. The authors distinguish between two main relations: drug-disease or drug-drug relation, based on the type of the secondary entity in the entity pairs.

The relationship between "drugs" and "diseases" can be further divided into positive and negative correlations; "Positive correlation" means that the drug can treat or alleviate the disease; "negative correlation" means that the drug can cause or aggravate the disease. The relationships between drugs are firstly divided into exist-interaction and non-interaction (Sometimes a sentence states that there is no known interaction between two drugs.); then for the sentences describing drug pairs that do

have interactions, the authors further extract three types of relationships, including the interaction mechanism, clinical guidance, and interaction results. Each drug-drug interaction relationship type can be categorized into more fine-grained labels. For example, "clinical guidance" can be further divided into use with caution (UC), prohibition of joint use (PU), joint use as per prescription (PP), joint use over different time spans (TS), different bottle injections (BI), etc. The "interaction results" are further divided into decreased efficacy (ED), improved efficacy (EI), increased adverse reactions (RI), improved efficacy and increased adverse reactions (EIRI), decreased efficacy and increased adverse reactions (EDRI), etc. Ultimately, fine-grained knowledge of drug interaction and drug usage can be established through automatic mining of text data in drug instructions. The extraction process is exemplified in Figure 2.

## Pre-Training Based Neural Models

The NER, EPC, and RE modules are equipped with different deep learning models. All modules employ a variant of the BERT model as the pre-training model. The NER module combines the pre-training model with BiLSTM and CRF. The EPC and RE modules combine the pre-training model with a BiGRU and an attention mechanism. In this section, the specific configurations of deep learning models used in relevant modules are introduced.

**Figure 2. Examples of information extraction from drug instructions (translated from Chinese drug instruction)**
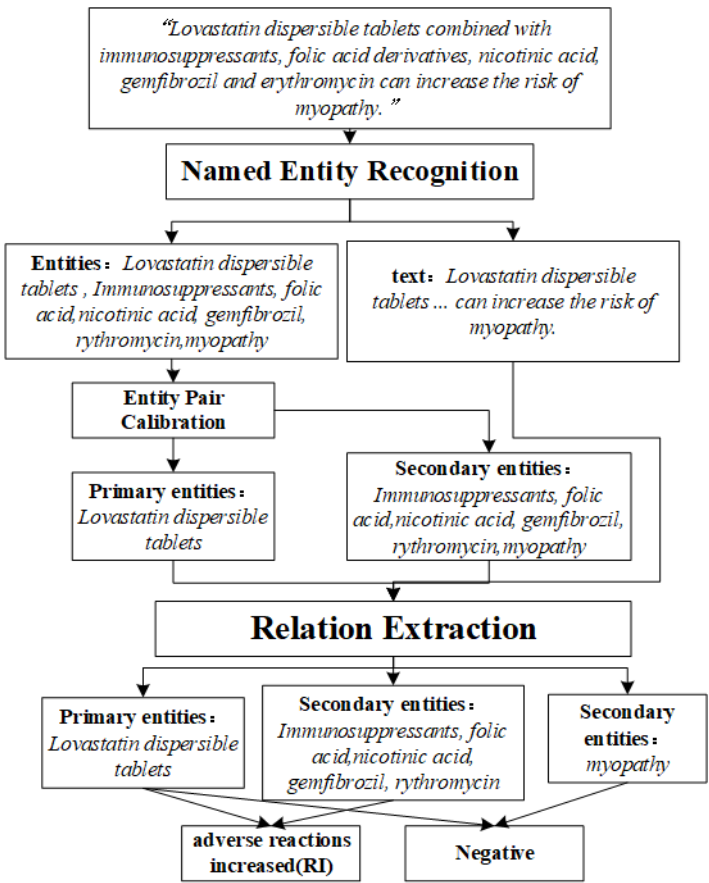
## *BERT(wwm) for Pre-Training*

Current study suggests that pre-training models can significantly improve the downstream tasks (Devlin et al., 2018). Also, additional information can be supplied to pre-training models for better semantic representation (Wu & He, 2019). In this paper, pre-training models are used to capture the semantics of entity roles. The BERT(wwm) pre-training model was selected as the first layer encoder. The BERT(wwm) is a variant of BERT, with improvements on the masking method for Chinese text, that is, masking the entire word in Chinese instead of a single character, so as to keep the word and phrase level semantics in Chinese, as shown in Table 3.

From the text semantics after "Mask," we can see that the text semantics after "WWM" is more complete, and the experimental effect of the final BERT(wwm) on the open dataset CMRC 2018 is also better than that of BERT, so BERT(wwm) was chosen as the pre-training model.

The BERT(wwm) model is built based on BERT transformers, which is a two-way language model of transformers. A transformer is a method different from the traditional neural network, which completely relies on a self-attention mechanism to calculate the vector representation of input and output. First, three vectors are created for each word embedding vector of the encoder, which are query vector, key vector, and value vector. These vectors are the weight matrix w obtained by combining the word vector with the three training vectors $W_Q$, $W_K$, $W_V$. The matrix is the parameter to be learned, and then the output is obtained by calculating the formula:

$$Z = Attention\left(Q, K, V\right) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

In formula (1): each line of Q, K, and V represent the query vector, key vector, and value vector; D is the length of the key vector, divided by dk1/2 to get a more stable gradient in the process of back propagation. The structure of the BERT(wwm) model is shown in Figure 3.

## *BERT(wwm)-BiLSTM-CRF for NER*

The BERT(wwm)-BiLSTM-CRF model is the deep learning model of the NER module. The purpose of this model is to recognize the names of drugs, diseases, symptoms, and body parts from the medical text, so as to provide data for the subsequent relationship extraction task. The BERT(wwm) pre-training model trains an intermediate word vector for the encoder, which better integrates the syntax and semantics of the context of the word. Then the intermediate word vector is used as the input of the downstream BiLSTM model. BiLSTM is a bidirectional recurrent neural network, which can better capture the dependency of context semantics in text. The prediction score of each tag is labeled by the training output sequence of the BiLSTM network, which will be used as the input of the CRF

**Table 3. The whole word mask example (translated from Chinese)**

| Categories | Examples |
|---|---|
| text | When sirolimus and voriconazole are used together, the blood drug level increases drastically, thus these should not be used jointly. |
| MASK | When sirolimus and voriconazole are used together, the blood drug level [MASK] increases drastically, thus [MASK] should be used jointly. |
| WWM | When sirolimus and voriconazole are used together, the blood drug level [MASK] drastically, thus [MASK] should be used jointly. |

**Figure 3. BERT(wwm) model architecture**
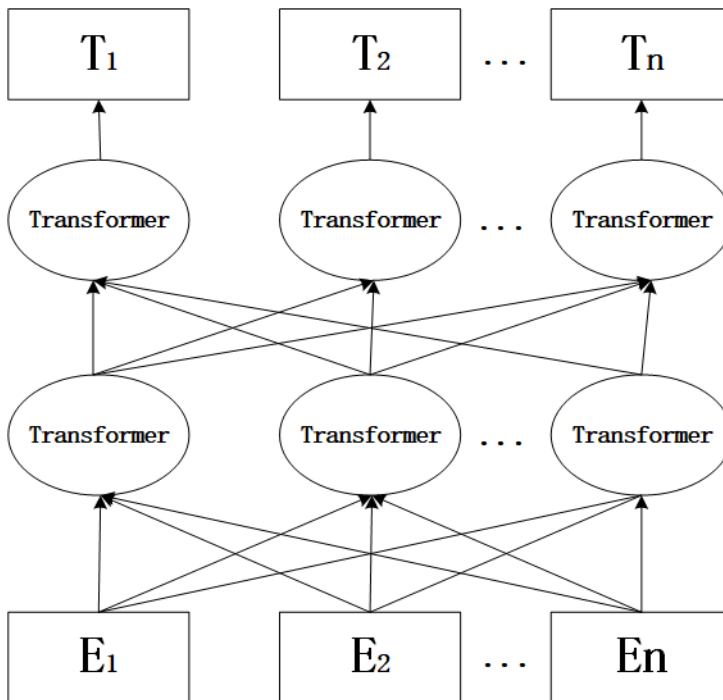


layer. The CRF layer can obtain the constraint rules of sequence labeling data from the training data, thus greatly reducing the probability of illegal sequence labels being predicted.

The overall model architecture is shown in Figure 4. In this figure, the input text is first transformed into a dynamic semantic vector by BERT(wwm). The vector is trained by the BiLSTM network and the constraint rules of the CRF layer, and the label corresponding to each word is finally output. "B-DRUG" represents the beginning of an entity, "I-DRUG" represents the middle and end of an entity, and "O" represents the part that is not an entity.

## BERT(wwm)-BiGRU-ATT for EPC and RE

The BERT(wwm)-BiGRU-ATT model is used in the EPC and RE modules. Different word embedding can be used for different tasks. In this model, BERT(wwm) is still used as the upstream encoder to encode the text as the intermediate word vector, and the downstream uses a bidirectional GRU network and attention mechanism to train the upstream word vector. The GRU network is a variant of the LSTM network, which can solve the problems of long-term dependence and gradient dispersion. Compared with the LSTM network, the GRU network has higher computational efficiency and occupies less memory. Therefore, considering the heavy workload of EPC and RE task of computation, the bidirectional GRU network is used. The entire model is accompanied by a final attention layer, as shown in Figure 5.

A self-attention mechanism is used in Figure 5. For a text input, the model can generate different attention scores on different words, so that the model can pay more attention to certain words and improve the overall performance of the model. More specifically, given a sentence representation where $s = \{X_1, X_2 \ldots X_n\}$, and $X_n$ represents a word embedding, the model first randomly initializes the three

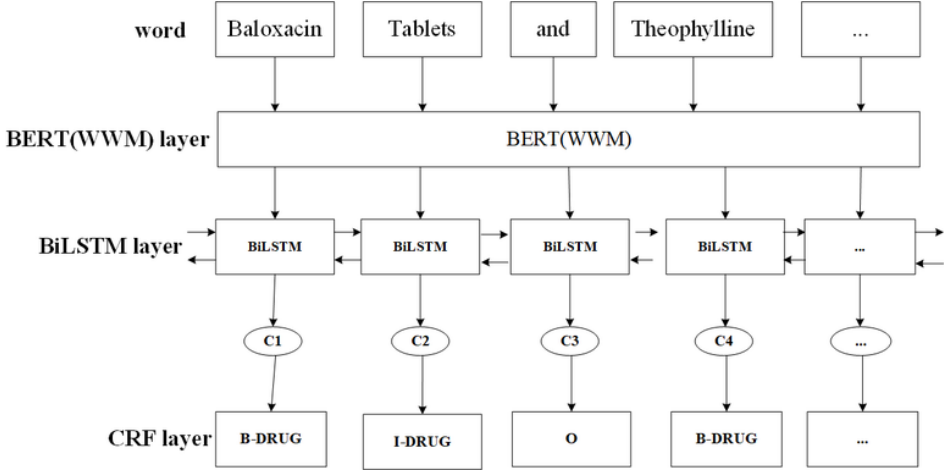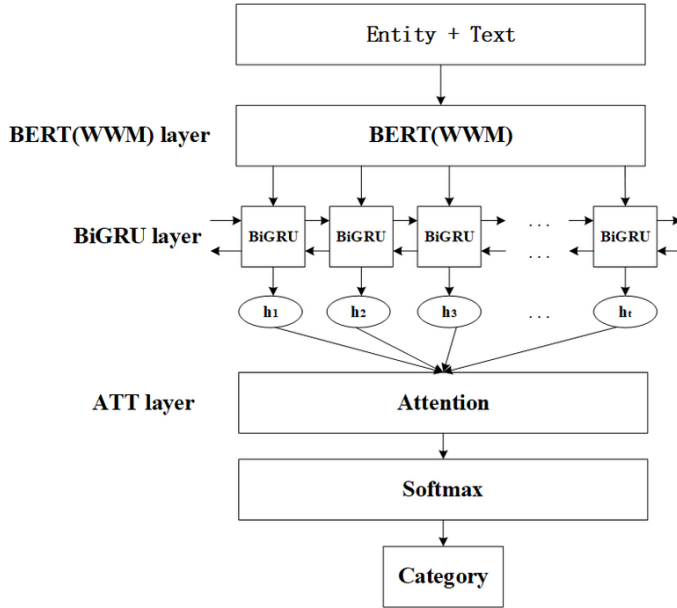**Figure 4. BERT(wwm)-BiLSTM-CRF model architecture**



**Figure 5. BERT(wwm)-BiGRU-ATT model architecture**



vectors $Q$, $K$, $V$ for each word, representing the query, key and value vectors. Then, it trains three weight matrices $W_Q$, $W_K$, $W_V$ and adjusts the word vectors using $q_n = X_n * W_Q$, $k_n = X_n * W_K$, $v_n = X_n * W_V$. Then, the preliminary word score is given by $f_1 = q_1 * \sum_{i=1}^{n} k_i$, and the final score is given by $F = \sum_{j=1}^{n} v_j * Softmax\left(f_j\right)$. This score determines the importance of other parts of the input sentence when encoding the word at the position. As shown in the example in Figure 6.

**Figure 6. Text example of an attention mechanism (translated from Chinese)**



When combined with **voriconazole**, the blood concentration of sirolimus may **increase** significantly, so the two drugs **can not** be used at the same time

For the human brain, the two words "升高" (increase) and "不可" (cannot) in such a text determines the relation between the two entities. Similarly, through attention calculations, these two words are assigned with a greater attention score, thus the appearance of these words can give hints to the model for identifying relevant relations between entities.

## ENTITY PAIR CALIBRATION

As described in Problem Definition, there are often multiple expressions of the same entity name in a drug instruction, and there are inclusion relationships (for example, one entity is a component of another entity), sub-class, and type-of relationships among different entities. There should be no drug-drug interaction between such entity pairs. When this kind of entity pair is directly used as the input of the subsequent relation extraction task, it will cause a lot of noise. The objective of $T_{epc}$ is to identify entity roles and use this information to reduce noise for $T_{re}$.

For example, in the two texts in Figure 7, the first text involves three entity names: quinolones, theophylline, and balofloxacin tablets. Among the three entity names, quinolones is a drug type for balofloxacin tablets. Similarly, in the second text, vitamin-A is a component of multivitamin chewable tablets. In both cases, the entity pairs are incorrect, i.e., noise for relation extraction. A naive approach will simply extract all entities from text and apply a Cartesian product to produce all entity pairs, which will result in very noisy input for the RE module.

A natural improvement will be using a filter based on a priori knowledge after the Cartesian product to eliminate illegal entity pairs, such as those in Table 4. However, this approach imposes a strong assumption that an accurate and complete knowledge base that describes the type-of, component-of, and subclass-of relations between drugs, exists. In practice, such a priori knowledge may be incomplete or even non-existence.

In order to avoid using such wrong entity pairs, this paper proposes to check the entity pair after the named entity recognition task and before the entity relationship extraction, so as to determine whether the entity pair is valid. The basic task of entity pair verification is to label each entity name extracted by the named entity recognition task into the primary entity or the secondary entity using a deep learning based algorithm. The primary entity set contains only the drug entity that the instruction

**Figure 7. Examples of entity type-of and component-of relations in drug instructions (translated from Chinese)**



The combination of **quinolones** and **theophylline** may increase the serum theophylline level, therefore, the dosage of theophylline should be reduced when **balofloxacin** tablets are combined with theophylline

Antacids can affect the absorption of vitamin A in children's vitamin chewable tablets, so they should not be taken together

**Table 4. Traditional rule-based entity pair extraction**

| |
|---|
| Input: set of entities: ENs<br>Output: set of entity pairs: EN_tuple |
| begin<br>  $ENre_1$, $ENre_2$, EN_tuple = [], [],[]<br>  for (EN in ENs):<br>  if EN in compositions:<br>    $ENre_1$.append(EN)<br>  else:<br>    $ENre_2$.append(EN)<br>EN_tuple = Cartesian_product($ENre_1$, $ENre_2$)<br>EN_tuple = FilterWithAprioriKB(EN_tuple)<br>return EN_tuple<br>end |

is intended for, or its type, super-class, or drug-composition. The secondary entity contains only the entity associated with the primary entity in the instruction. This way, one can produce correct entity pairs for relation extraction.

To successfully recognize the entity category from the text, this research uses the "WWM" mechanism similar to the BERT(wwm) model. Here, the primary /secondary entities are masked (with m '#' symbols, where m is the length of the masked entity) when training the algorithm to identify them. Meanwhile, to connect the information in the entity name, the same splicing method is used in BERT and linked to the entity name (with a '$' symbol) in front of the masked sentence. This not only ensures a lossless text embedding, but also obtains flexibility in the embedding method. Table 5 depicts an example of EPC training data.

Table 6 details the procedure of the EPC module. It first transforms the named entity set ENs and the relevant sentence text with a word embedding function trans_en_ text(). The result of this function is the encoded text: text_new. The label of the entity can be generated when text_ new is fed into the neural network model, and the category of the entity can be determined according to the label. Finally, all the primary and secondary entities are processed by function: Cartesian_product(), generating the output entity pairs: EN_ tuple.

## EXPERIMENT AND RESULT ANALYSIS

In order to evaluate the approach chosen, the modules in this framework were experimented over realistic datasets[1]. In this section, we first present the experiment dataset and configuration, then, we demonstrate our experiment results and analysis.

**Table 5. Example of the training data for entity pair checking algorithm (translated from Chinese text)**

| Label | Text (masked) |
|---|---|
| Primary Entity | quinolones$The combination of ########### and theophylline may increase the level of theophylline in serum, so the dose of theophylline should be reduced when balofloxacin tablets are combined with theophylline |
| Primary Entity | balofloxacin tablets$The combination of quinolones and theophylline may increase the level of theophylline in serum, so the dose of theophylline should be reduced when ########### are combined with theophylline |
| Secondary Entity | theophylline$The combination of quinolones and ######## may increase the level of ######## in serum, so the dose of theophylline should be reduced when balofloxacin tablets are combined with theophylline |

**Table 6. Entity pair calibration extraction using pre-training**

---

**Input: set of entities: ENs, text containing entities: text**
**Output: set of entity pairs: *EN_tuple***

**begin**
*ENnn₁, ENnn₂ EN_tuple = [], [],[]*
**for** (*EN* **in** *ENs*):
*text_new* = **trans_en_text**(*EN, text*)
*label* = **NN**(*text_new*)
**if** label **==** *en_master***:**
*ENnn₁.***append**(*EN*)
**else:**
*ENnn₂.***append**(*EN*)
*EN_tuple* = **Cartesian_product**(*ENnn₁, ENnn₂*)
**return** *EN_tuple*
**end**

---

## Dataset and Parameter Settings

The experimental data of this paper mainly comes from authoritative data sources on the Internet. Starting from a medical entity dictionary, the medical domain sequence annotation dataset with three million Chinese characters for named entity recognition is constructed, which is used to train and verify the NER module. There are four types of entity names: drug name, disease name, body part name, and symptom name, as shown in Table 7.

For the EPC RE module, 60,000 Chinese sentences were prepared from more than 4,000 drug instructions. EPC labels entities with two categories: primary or secondary. RE labels relations among entities with 12 fine-grained drug-drug and drug-disease relations (as introduced in the third section). The proportion for training and test dataset is 4:1.

In terms of parameter setting, this paper uses the base version of the Chinese BERT(wwm) pre-training model, and the network structure consists of 12-layer, 768-hidden, 12-heads, and 110 million parameters. The maximum sentence length set by the named entity recognition part is 128. The number of sentences processed each time is 64. The learning rate is $5 * 10-5$. The number of LSTM network nodes is set to 128. Dropout is set to 0.5 to prevent overfitting, and the training times are three. The maximum sentence length of the entity pair calibration and relation extraction part is 80. The number of sentences processed each time is 16. The learning rate was 0.001. The number of GRU network nodes is set to 128. Dropout is set to 0.2 to prevent overfitting, and the training times are 10. The words embedded in the BiGRU-ATT model adopt the character vector obtained by word2vec training. The number of characters in the table is 16116 and the vector dimension is set to 100. All the other parameters are initialized randomly.

**Table 7. Entity name categories and examples**

| Entity categories | Examples |
| --- | --- |
| DRUG | oxazepam, atorvastatin, tadalafil |
| DISEASE | diabetes, AIDS, nasopharyngeal cancer |
| BODY | tongue, stomach, spleen |
| SYMPTOM | wheezing, fatigue, ventosity |

## Result Analysis

### Named Entity Recognition (NER)

The performance of the algorithm model in this paper is mainly evaluated by P (precision), R (recall), and F1. The comparative experiment of the named entity recognition algorithm model in the information extraction framework is shown in Figure 8.

As can be seen from Figure 8, pre-training models significantly improve the performance of NER. In addition, the BERT(wwm)-BiLSTM-CRF model generally performs the best in the recognition of drug, symptom, and body names, except for disease names. In the prediction results of the entity recognition model, the BERT(wwm)-BiLSTM-CRF model can also successfully recognize the names that are not matched by dictionary matching, as shown in Table 8.

In the graph, the first column is the text data labeled by sequence, the second column is the label labeled by the dictionary, and the third column is the label predicted by the model. It can be seen that the BERT(wwm)-BiLSTM-CRF model is more successful in identifying the drug name "双氯芬酸钠" (Diclofenac Sodium), which does not appear in the manual summary dictionary. The model realizes the generalization of the name recognition, which can further supplement the names in the dictionary.

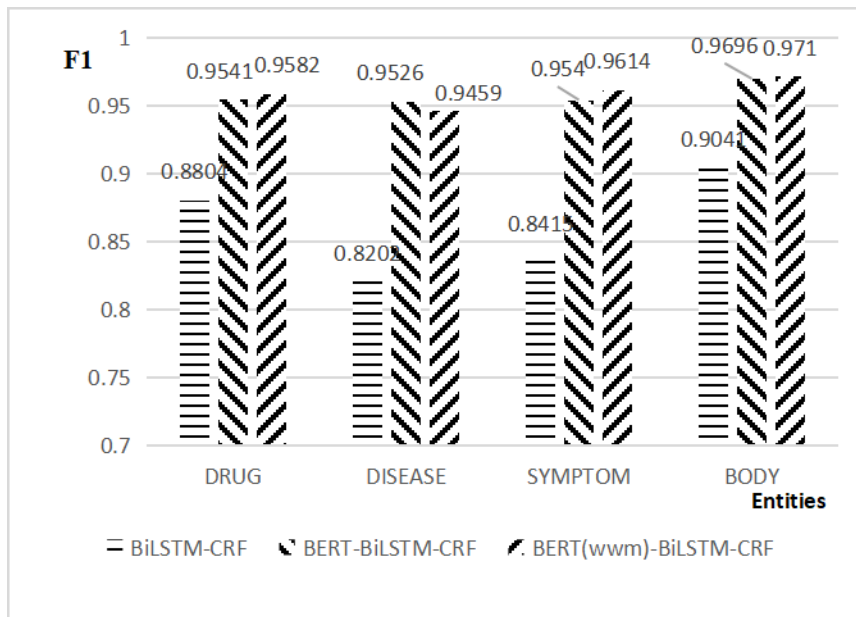**Figure 8. F1 score for named entity recognition algorithm model**



**Table 8. BERT(wwm)-BiLSTM-CRF model predicts result**

| Token | Diclofenac | sodium | used | in | the | treatment | of | osteoarthritis |
|---|---|---|---|---|---|---|---|---|
| Dictionary-based matching | O | O | O | O | O | O | O | B-DISEASES |
| BERTwwm-BiLSTM-CRF | B-DRUG | I-DRUG | O | O | O | O | O | B-DISEASES |

## Entity Pair Calibration (EPC)

The experimental results of the entity pair calibrator are shown in Table 9, where En1 represents the primary entity, and EN2 represents the secondary entity.

It can be seen from the results in Table 6 that the experimental performance of DPCNN (Johnson, & Zhang, 2017) without pre-training neural network, is quite good. Additionally, it can be seen that without increasing too much computing cost, better experimental results are achieved by increasing the depth of the network, but there is still a performance gap compared to the model used in this paper. The results show that the F1 values of the main entity and the sub-entity recognition effect of the BERT(wwm)-BiGRU-ATT model are all above 0.98, which can successfully distinguish the entity categories in the text.

## Relation Extraction (RE)

The experimental results of RE module w.r.t. coarse-grained relations are shown in Figures 9 through 12. Nine different models were tested:

1.  Conventional deep learning approach (BiGRU-ATT).
2.  Deep learning with pre-training model (BERT-BiGRU-ATT).
3.  On top of 2), add an entity pair filter with a priori knowledge matching (BERT-BiGRU-ATT(RE)).
4.  On top of 2), adding the EPC module as described in Section "Entity Pair Calibration" (BERT-BiGRU-ATT (EPC)).
5.  R-BERT as in (Wu & He 2019).
6.  Roberta as in (Liu et al., 2019).
7.  BERT-CLS as in (Devlin et al., 2018).
8.  CasRel, as in (Wei et al., 2020).
9.  PRGC as in (Zheng et al., 2021).

Models #1 and #2 are conventional DL models. Model #3 adds an entity calibration with predefined rules. model #4 is the approach proposed in this paper. Models #5 to #7 are SOTA pre-training models. Models #8 and #9 are SOTA joint models.

As can be seen from Figures 9 through 12, pre-training-based pipeline models (Models #3 through #7) as well as joint models (Models #8 and #9) outperform conventional deep learning models (Models #1 and #2). More interestingly, although adding the EPC module performs the best as was expected; using a priori knowledge to eliminate wrong entity pairs performs similarly to not using filters. This demonstrates that the use of a priori knowledge heavily depends on the quality and completeness of the knowledge.

**Table 9. Experimental results of entity pair calibration (EPC) algorithm model**

| Model | Entity Categories | Precision | Recall | $F_1$ |
|---|---|---|---|---|
| DPCNN | EN 1<br>EN 2 | 0.9256<br>0.9694 | 0.9387<br>0.9626 | 0.9321<br>0.9660 |
| BiLSTM-ATT | EN 1<br>EN 2 | 0.8935<br>0.9552 | 0.9104<br>0.9463 | 0.9019<br>0.9507 |
| BERT-BiGRU-ATT | EN 1<br>EN 2 | 0.9717<br>0.9920 | 0.9836<br>0.9862 | 0.9776<br>0.9891 |
| BERT(wwm)-BiGRU-ATT | EN 1<br>EN 2 | **0.9719**<br>**0.9960** | **0.9918**<br>**0.9862** | **0.9817**<br>**0.9911** |

**Figure 9. F1 score for positive and negative correlation extraction between drugs and diseases**
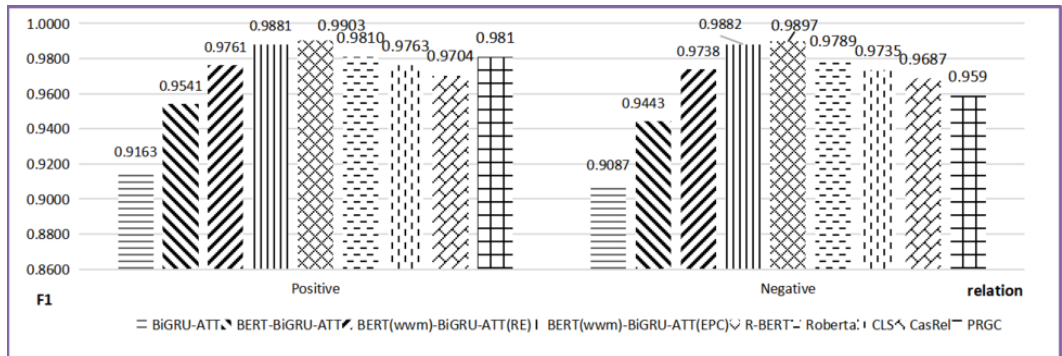


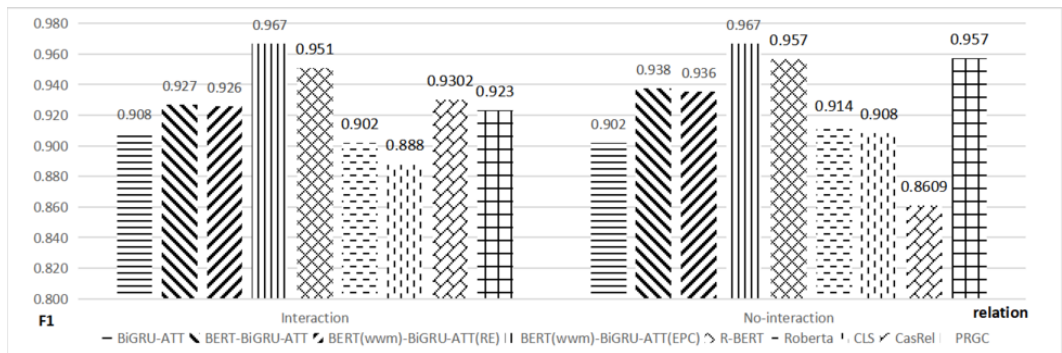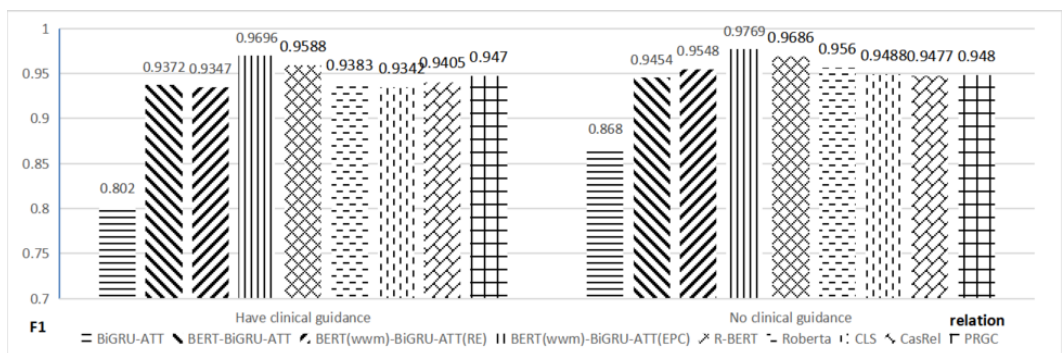**Figure 10. F1 score for extracting the existence of drug-drug interaction**



**Figure 11. F1 score for extracting the existence of "clinical guidance" in drug-drug interaction**



In addition, the authors' framework slightly outperformed R-BERT (Model #5) in 3 out of 4 coarse relations, and was better than Models #6 through #9 in all four coarse-grained relations.

Figures 13 and 14 show the experiment results for two more fine-grained RE tasks w.r.t., the "clinical guidance," and "interaction result," using the same nine models.

Results from Figures 13 and 14 suggest that the performance of fine-grained relation extraction is improved significantly by EPC. More specifically, for the relation of "use with caution (UC)," the

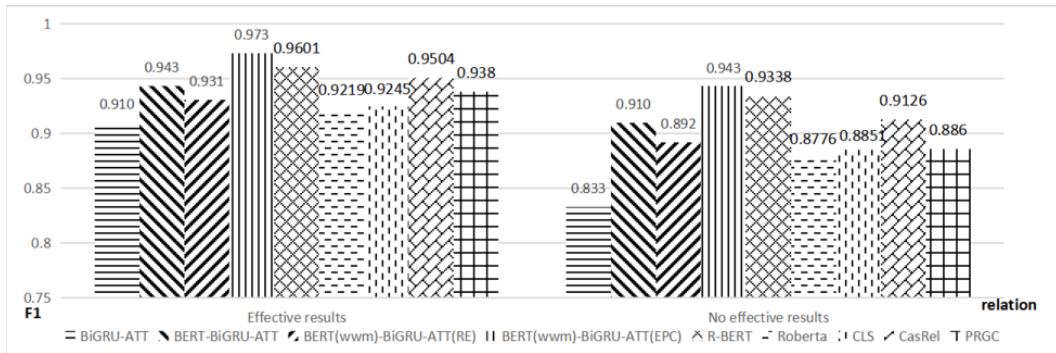**Figure 12. F1 score for extracting the existence of "interaction result" in drug-drug interaction**



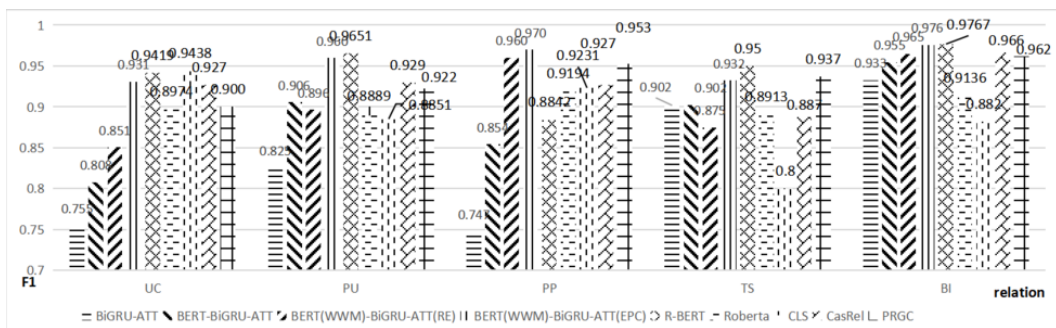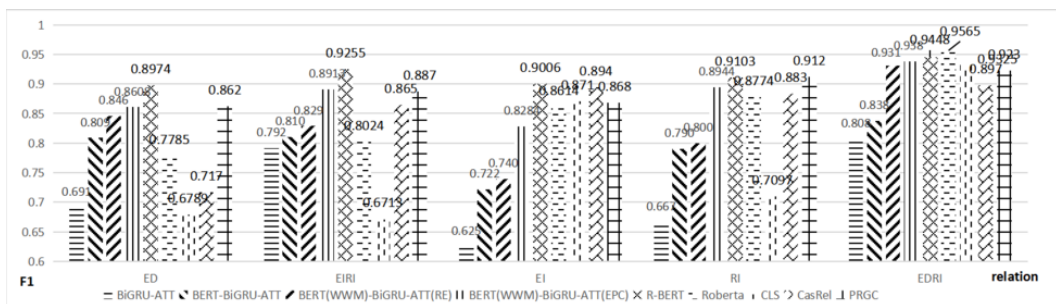**Figure 13. F1 score for fine-grained "clinical guidance" relation extraction**



**Figure 14. F1 score for fine-grained "interaction result" relation extraction**



F1 value is increased by 8.04%, and for the relation of "adverse reaction increase (RI)," the F1 value is increased by 9.44%. Compared to the coarse-grain relation extraction result (3.2% F1 improvement on average) in Figures 9 through 12, one can see that the EPC module offers a bigger contribution (4.9% F1 improvement on average), while extracting more fine-grained and complex relations.

In addition, when dealing with fine-grained relations, the authors' framework is slightly outperformed by R-BERT, and the performance is on par with PRGC (+0.5% F1), but is still better than Models #6 through #8. More specifically, the authors' framework was outperformed by R-BERT in 8 out of 10 fine-grained relations, but their framework was better than Roberta, BERT-CLS, and

CasRel in 9 out of 10 fine-grained relations, and they were better than PRGC in 6 out of 10 relations. A more comprehensive comparison on the average performance is depicted in Table 10.

The results from Table 10 demonstrates that the authors' model is 0.08% better than R-BERT (SOTA 2020) over coarse-grained relations, and is 2.2% worse than R-BERT over fine-grained relations. The results shows that R-BERT gains a lot of useful information from entity type annotation, and is indeed better than using the entity role information, especially in fine-grained relations. The main reason for this is that fine-grained relation extraction relies more on the semantic type of entities rather than on their syntactic roles. However, the authors' argue that entity type enhancement requires a priori knowledge and requires more manual labeling effort, where labeling the entity roles is much easier and does not rely on complete and accurate domain taxonomy.

It is also observable from Table 10 that PRGC, R-BERT and CasRel have the top three smallest performance drop when dealing with fine-grained relations (-3.0%, -3.4% and -4.6%), and the authors' approach ranked fourth with a -5.3% performance drop. These results indicate that: 1) both entity type enhancements and joint models are naturally more stable and, 2) without introducing much labeling effort (for annotating entity types or entity-relation correlations), the authors' approach is also quite resilient over fine-grained relation extractions.

## CONCLUSION

Extracting fine-grained drug knowledge from professional drug instructions is a complicated task for both humans and machines. In order to address this problem, the authors' proposed an end-to-end information extraction framework based on pre-training and deep learning models. The structure of the framework follows a basic design of pipeline information extraction methodology that includes data gathering, named entity recognition, and relation extraction, on top of which, the authors' designed a novel procedure called Entity Pair Calibration (EPC) and placed it between NER and RE tasks to reduce the input noise for drug-drug interaction extraction.

The framework was empirically evaluated over more than 60,000 sentences from Chinese drug instructions, covering 4,000 different drugs. Experiment results demonstrate that the proposed framework can achieve >=0.95 F1 score in NER and >=0.83 in RE tasks. Moreover, the proposed EPC module can significantly improve the performance for fine-grained relation extraction (F1 score improvement up to 9.44%; 5% on average). Compared to using entity type information (e.g., R-BERT), enhancing the pre-training models with entity roles can achieve similar performance without relying on a priori knowledge and complicated labeling. Using this framework, the authors have successfully built a drug interaction knowledge base with over 1,300,000 RDF triples, describing more than 180,000 drug-drug and drug-disease relations. It is worth mentioning that although the authors designed and evaluated the framework for Chinese drug instructions, it can be easily adapted for other languages by changing the pre-training model into BERT or other language-specific models. In addition, the fundamental idea of EPC can be reused for other languages.

The future work may be carried out in three aspects. First, the pre-training model BERT(wwm) used in this paper can be optimized with grammatical structure and semantic information (Liu et al.,

**Table 10. Overall F1 performance over coarse- and fine-grained relations**

|  | BiGRU-ATT | BERT-BiGRU-ATT | BERT(wwm)-BiGRU-ATT(RE) | BERT(wwm)-BiGRU-ATT(EPC) | R-BERT | Roberta | CLS | CasRel | PRGC |
|---|---|---|---|---|---|---|---|---|---|
| **Coarse-grained** | 0.881 | 0.937 | 0.940 | **0.972** | *0.964* | 0.934 | 0.930 | 0.935 | 0.942 |
| **Fine-grained** | 0.774 | 0.839 | 0.869 | *0.918* | **0.930** | 0.879 | 0.830 | 0.889 | 0.913 |
| **Δ** | -0.107 | -0.098 | -0.071 | *-0.053* | -0.034 | -0.055 | -0.100 | -0.046 | **-0.030** |

2019; Sun et al., 2020). Second, the architecture of the framework is configurable but not dynamically adaptable. The authors could improve on this aspect by monitoring the result and dynamically switching to better models. In addition, adding confrontation training (Goodfellow, Shlens, & Szegedy, 2014) in the word embedding stage of the model can improve the robustness of the model. Finally, there is an inherent problem in this domain, which is the imbalance of samples in drug-drug interactions. This problem is not easily solvable using pure learning-based approaches. However, the authors could explore using a hybrid approach of knowledge and learning-based approaches to address this issue (Li, Xu, & Lu, 2018).

## CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

## FUNDING AGENCY

## REFERENCES

Bao, T., Ren, N., Luo, R., Wang, B., Shen, G., & Guo, T. (2021). A BERT-Based Hybrid Short Text Classification Model Incorporating CNN and Attention-Based BiGRU. *Journal of Organizational and End User Computing*, *33*(6), 21. doi:10.4018/JOEUC.294580

Bao, Y., Wu, M., Chang, S., & Barzilay, R. (2019). *Few-shot text classification with distributional signatures.* arXiv preprint arXiv:1908.06039.

Giannis, B., Johannes, D., Thomas, D., & Chris, D. (2018). Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications*, *114*(DEC), 34–45.

Beliga, S., Meštrović, A., & Martinčić-Ipšić, S. (2016). Selectivity-based keyword extraction method. *International Journal on Semantic Web and Information Systems*, *12*(3), 1–26. doi:10.4018/IJSWIS.2016070101

Bose, P., Srinivasan, S., Sleeman, W. C. IV, Palta, J., Kapoor, R., & Ghosh, P. (2021). A survey on recent named entity recognition and relationship extraction techniques on clinical texts. *Applied Sciences (Basel, Switzerland)*, *11*(18), 8319. doi:10.3390/app11188319

Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3504–3514. doi:10.1109/TASLP.2021.3124365

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805.

Fang, X., & Sheng, H. (2002, October). Pattern acquisition for Chinese named entity recognition: A supervised learning approach. In *International Conference on Advances in Information Systems* (pp. 166-175). Springer Berlin Heidelberg. doi:10.1007/3-540-36077-8_16

Fatehifar, M., & Karshenas, H. (2021). Drug-drug interaction extraction using a position and similarity fusion-based attention mechanism. *Journal of Biomedical Informatics*, *115*, 103707. doi:10.1016/j.jbi.2021.103707 PMID:33571676

Feng-Lin, L. I., & Jia, K. E. (2018). Research progress of entity relation extraction base on deep learning framework. *Information Science*, *36*(3), 169–176.

Gao, F., Zhou, L., & Gu, J. (2021, October). Entity Pair Recognition using Semantic Enrichment and Adversarial Training for Chinese Drug Knowledge Extraction. In *Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences* (pp. 35-42). doi:10.1145/3500931.3500939

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and harnessing adversarial examples.* arXiv preprint arXiv:1412.6572.

Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., & Declerck, T. (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of Biomedical Informatics*, *46*(5), 914–920. doi:10.1016/j.jbi.2013.07.011 PMID:23906817

Jagan, B., Parthasarathi, R., & Geetha, T. V. (2019). Bootstrapping of semantic relation extraction for a morphologically rich language: Semi-supervised learning of semantic relations. *International Journal on Semantic Web and Information Systems*, *15*(1), 119–149. doi:10.4018/IJSWIS.2019010106

Viswanathan, J., Saranya, N., & Inbamani, A. (2021). Deep Learning Applications in Medical Imaging. *International Journal on Semantic Web and Information Systems*, 156–177. Advance online publication. doi:10.4018/978-1-7998-5071-7.ch007

Johnson, R., & Zhang, T. (2017, July). Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (*Volume 1*: Long Papers)* (pp. 562-570). doi:10.18653/v1/P17-1052

Lejun, G., Xiaolin, L., Zhihong, G., & Huakang, L. (2020). Extraction of drug interaction relationship based on bidirectional GRU and CNN [基于双向gru和cnn的药物相互作用关系抽取]. *Journal of Shaanxi Normal University (Natural Science Edition), 48*(6), 6. http://qikan.cqvip.com/Qikan/Article/Detail?id=710332435

Li, S., Xu, H., & Lu, Z. (2018). *Generalize symbolic knowledge with neural rule engine.* arXiv preprint arXiv:1808.10326.

Li, X., Yin, F., Sun, Z., Li, X., Yuan, A., Chai, D., Zhou, M., & Li, J. (2019). *Entity-relation extraction as multi-turn question answering.* arXiv preprint arXiv:1905.05529. 10.18653/v1/P19-1129

Lin, F., & Xie, D. (2020, October). Research on named entity recognition of traditional Chinese medicine electronic medical records. In *International Conference on Health Information Science* (pp. 61-67). Springer. doi:10.1007/978-3-030-61951-0_6

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *Roberta: A robustly optimized BERT pre-training approach.* arXiv preprint arXiv:1907.11692.

Lv, X., Hou, H., You, X., Zhang, X., & Han, J. (2020). Distant supervised relation extraction via DiSAN-2CNN on a feature level. *International Journal on Semantic Web and Information Systems*, *16*(2), 1–17. doi:10.4018/IJSWIS.2020040101

Mingliang, D., Jijun, T., & Fei, G. (2021, December). Document-level DDI relation extraction with document-entity embedding. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 392-397). IEEE. doi:10.1109/BIBM52615.2021.9669319

Miwa, M., & Bansal, M. (2016). *End-to-end relation extraction using LSTMs on sequences and tree structures meeting of the association for computational linguistics.* arXiv preprint arXiv:1601.00770.

Ningthoujam, D., Yadav, S., Bhattacharyya, P., & Ekbal, A. (2019). *Relation extraction between the clinical entities based on the shortest dependency path based lstm.*

Pawar, S., Palshikar, G. K., & Bhattacharyya, P. (2017). *Relation extraction: A survey.*

Shen, W., Wang, J., Luo, P., & Wang, M. (2015). A hybrid framework for semantic relation extraction over enterprise data. *International Journal on Semantic Web and Information Systems*, *11*(3), 1–24. doi:10.4018/IJSWIS.2015070101

Soares, L., FitzGerald, N., Ling, J., & Kwiatkowski, T. (2019). *Matching the Blanks: Distributional Similarity for Relation Learning meeting of the association for computational linguistics.* arXiv preprint arXiv:1906.03158.

Srinivasan, S., & L D, D. B. (2019). A parallel neural network approach for faster rumor identification in online social networks. *International Journal on Semantic Web and Information Systems, 15*(4), 69-89. http://DOI.org/10.4018/IJSWIS.2019100105

Sun, C., Yang, Z., Wang, L., Zhang, Y., Lin, H., & Wang, J. (2021). Biomedical named entity recognition using BERT in the machine reading comprehension framework. *Journal of Biomedical Informatics*, *118*, 103799. https://DOI.org/10.1016/j.jbi.2021.103799

Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H., & Wang, H. (2020, April). Ernie 2.0: A continual pre-training framework for language understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(5), 8968–8975.

Taewijit, S. (2021). Learning Pattern Relation-Based Hyperbolic Embedding for Adverse Drug Reaction Extraction. International Journal of Knowledge and Systems Science.

Tiktinsky, A., Viswanathan, V., Niezni, D., Azagury, D. M., Shamay, Y., Taub-Tabib, H., Hope, T., & Goldberg, Y. (2022). *A Dataset for N-ary Relation Extraction of Drug Combinations.* ArXiv, abs/2205.02289.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Majhi, V., Saikia, A., Datta, A., Sinha, A., & Paul, S. (2020). Comprehensive Review on Deep Learning for Neuronal Disorders: Applications of Deep Learning. *International Journal of Natural Computing Research*, *9*(1), 18.

Wang, Y. X., & Hebert, M. (2016, October). *Learning to learn: Model regression networks for easy small sample learning*. In *European Conference on Computer Vision* (pp. 616-634). Springer.

Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y. (2020). *A novel cascade binary tagging framework for relational triple extraction.* rXiv preprint arXiv:1909.03227.

Wu, S., & He, Y. (2019). Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2361-2364). ACM.

Yongchao, W., Hualing, M., Lingzhi, Z., & Wei, X. (2021). Joint extraction method of entity and relationship based on pointer network. *Computer Application Research*, *38*(4-5), 1004–1007.

Zeng, D., Liu, K., Lai, S., Zhou, G., & Zhao, J. (2014, August). Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 2335-2344). Academic Press.

Zhang, S. H., Du, A., & Jia, Z, & LI, T. R. (2021). Medical entity relation extraction based on deep neural network and self-attention mechanism [基于深度神经网络和自注意力机制的医学实体关系抽取]. *Computer Science*, *48*(10), 77–84.

Zheng, H., Wen, R., Chen, X., Yang, Y., Zhang, Y., Zhang, Z. Z., Zhang, N., Qin, B., Xu, M., & Zheng, Y. (2021). *PRGC: Potential relation and global correspondence based joint relational triple extraction.* arXiv preprint arXiv:2106.09895.

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (*Volume 2*: Short Papers)* (pp. 207-212). Academic Press.

Zhu, Z., Li, J., Zhao, Q., Wei, Y. C., & Jia, Y. (2021, July). Medical named entity recognition of Chinese electronic medical records based on stacked bidirectional long short-term memory. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 1930-1935). IEEE.

## ENDNOTE

[1]    Dataset and code available at: https://github.com/zhou-ls/Drug_Knowledge_Extraction_Framework