

Article

Open Domain Chinese Triples Hierarchical Extraction Method

Chunhui He , Zhen Tan , Haoran Wang *, Chong Zhang, Yanli Hu and Bin Ge

Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China; xtuhch@163.com (C.H.); tanzhen@nudt.edu.cn (Z.T.); chongzhang@nudt.edu.cn (C.Z.); huyanli@nudt.edu.cn (Y.H.); gebin@nudt.edu.cn (B.G.)

* Correspondence: 15697312525@163.com; Tel.: +86-13786206971

Received: 15 June 2020; Accepted: 9 July 2020; Published: 14 July 2020



Featured Application: This method can be applied to the task of automatic extraction of triples in unstructured text.

Abstract: Open domain relation prediction is an important task in triples extraction. When faced with the task of constructing large-scale knowledge graph systems, with the exception of structured data, it is necessary to automatically extract triples from a large amount of unstructured text to expand entities and relations. Although a large number of English open relation prediction methods have achieved good performance, the high-performance system for open domain Chinese triples extraction remains undeveloped due to the lack of large-scale Chinese annotation corpora and the difficulty of Chinese language processing. In this paper, we propose an integrated open domain Chinese triples hierarchical extraction method (CTHE) to solve this problem, considering the advantages of Bi-LSTM-CRF and Att-Bi-GRU models based on the pre-trained BERT encoding model. This method can recognize the named entities from Chinese sentences to establish entity pairs, and implement hierarchical extraction of specific and open relations based on the user-defined schema library and attention mechanism. The experimental results demonstrate the effectiveness of this method, which achieved stable performance on the test dataset, and better precision and F1-score in comparison with state-of-the-art Chinese open domain triples extraction methods. Furthermore, a large-scale annotated dataset for a Chinese named entity recognition (NER) task is established, which provides support for research on Chinese NER tasks.

Keywords: named entity recognition; open relation prediction; information extraction; CTHE

1. Introduction

To automatically expand new knowledge, obtaining new structured knowledge from massive amounts of unstructured data has become a popular research issue. Knowledge extraction technology represented by entity relation extraction has been successful. In recent years, in particular, supervised learning models have greatly promoted the development of specific relation triples extraction. However, compared with the complex challenges of extracting open relation triples in actual scenarios, some limitations remain with the existing methods. It is thus necessary to develop effective methods to resolve the problems of open domain triples extraction arising from actual scenario requirements.

Traditional entity-relation triples extraction methods usually have a pre-defined closed relation set and, in previous research, tasks have been converted into a relation classification problem with good results. However, under the open relation triples extraction scenario, the text contains a large number of open entity relations, which far exceeds the number of pre-defined relation types. In this case, the traditional relation classification models cannot directly and effectively obtain the new type of

relations between entities in the text. Determining a means to automatically discover new relations between entity pairs and implement open relation triples extraction remains a challenge.

To achieve the extraction of open relation triples, some scholars have proposed the task of open relation extraction (ORE), which is dedicated to extract the relation facts between entities from unstructured text. Recently, Ruidong Wu [1] proposed a supervised open relation triples extraction framework that implements the free switching of supervised and weakly supervised modes through the relation siamese network (RSN), which can simultaneously use supervised data and unsupervised data of new relations in unstructured text to jointly learn the semantic similarity of different relation facts. Specifically, the RSN uses a siamese network structure to learn the deep semantic features of the relation samples and the semantic similarity between them from the labeled data of the pre-defined relation, which can be used to calculate the semantic similarity of the text containing the open relation. However, Tianyu Gao [2] proposed that for a new type of open relation, a few precise examples are needed as seeds, and a pre-trained relation siamese network method can be used to train an extraction model suitable for the new type of relation.

This paper proposes an integrated open domain Chinese triples hierarchical extraction method to combine the advantages of deep learning with unsupervised algorithms and effectively expand the generalization ability of the open relation triples extraction model. The main contributions of this paper are as follows:

- a) We propose an integrated open domain Chinese triples hierarchical extraction method with BERT-based deep learning and unsupervised learning algorithms.
- b) Based on the segmentation strategy of dynamically adding a user-defined entities dictionary, we implement a simple and efficient open relation triples automatic extraction algorithm based on attention mechanism.
- c) A large-scale Chinese named entity recognition dataset is constructed, which provides a support for carrying out Chinese named entity recognition tasks.

2. Related Work

Triples extraction is one of the core tasks in the field of information extraction. It has been continuously researched for more than 20 years. Feature engineering [3], kernel methods [4–6], and graph models [7] have been widely used, and some staged results have been achieved. With the development of deep learning, neural network models have achieved breakthroughs in triples extraction.

Named entity recognition is a prerequisite for relation prediction. A series of results have been achieved using machine learning algorithms to automatically recognize named entities in sentences. For named entity recognition tasks, CRF [8] and Bi-LSTM-CRF [9] are two well-known methods.

For traditional relation classification tasks, it is effective to use feature engineering and machine learning algorithms to achieve relation classification, such as bootstrapping [10] and SVM [11]. In recent years, with the development of deep learning, many studies based on neural networks [12] have achieved good performance. Among these, on public datasets for specific relation extraction tasks, the popular methods of recent years have been Att-based LSTM [13] and Att-based CNN [14–16].

With further research, some scholars have combined named entity recognition and relation extraction to put forward some joint entity-relation extraction models [17]. Related research shows that the better-performing method uses a novel labeling strategy for jointly modeling of entity-relations [18,19].

In addition to the triple extraction of a specific relation, extraction of triples for open relations from unstructured text has also been recognized as a highly important task. In recent years, preliminary results have been obtained for the automatic extraction of triples within unstructured Chinese text. The more well-known methods are CORE [20], UnCORE [21], and DSNFs-based ORE [22]. In addition, with the advent of the attention mechanism [23] and large-scale pre-training language model BERT [24], many new records have been set for natural language processing tasks. Some studies have used these approaches to complete relation classification [25–27] and have achieved a higher accuracy.

In addition, Cui [28] proposed a multi-layered encoder-decoder framework to generate relation tuples. Inspired by the above studies, in this paper we propose an open domain Chinese triples hierarchical extraction method.

3. The Proposed CTHE Method

3.1. The Overall Method

To overcome the shortcomings of the existing triples extraction methods, we proposed an open domain Chinese triples extraction method. This method is a hierarchical framework, which combines five modules to achieve the triples extraction task with open relations.

- ⊙ Named Entity Recognition Module: this module aims to extract all named entities from the input original sentence, and combine some simple rules to form all the entity pairs that may have a relation.
- ⊙ Schema Match Module: this module aims to use a specific schema library to filter entity pairs for a specific pattern.
- ⊙ BERT-based Att-Bi-GRU Specific Relation Prediction Module: this module aims to apply the BERT-based Att-Bi-GRU model to finish the prediction of all the specific relations.
- ⊙ Relation Check and Confidence Discrimination Module: this module aims to check unknown relation and automatically determine the confidence of the predicted result by the BERT-based Att-Bi-GRU model.
- ⊙ Open Relation Prediction Module: this module aims to achieve entity relation prediction under non-specific, unknown classification, and low-confidence specific relation based on traditional sentence semantic dependency parse.

By combining named entity recognition, supervised learning methods with relation classification, and unsupervised learning methods with relation extraction, we built an integrated open domain Chinese triples hierarchical extraction framework. The overall framework of our model is shown in Figure 1. It is worth noting that all modules in the entire system are not independent, and they have a strict logical flow between input and output.

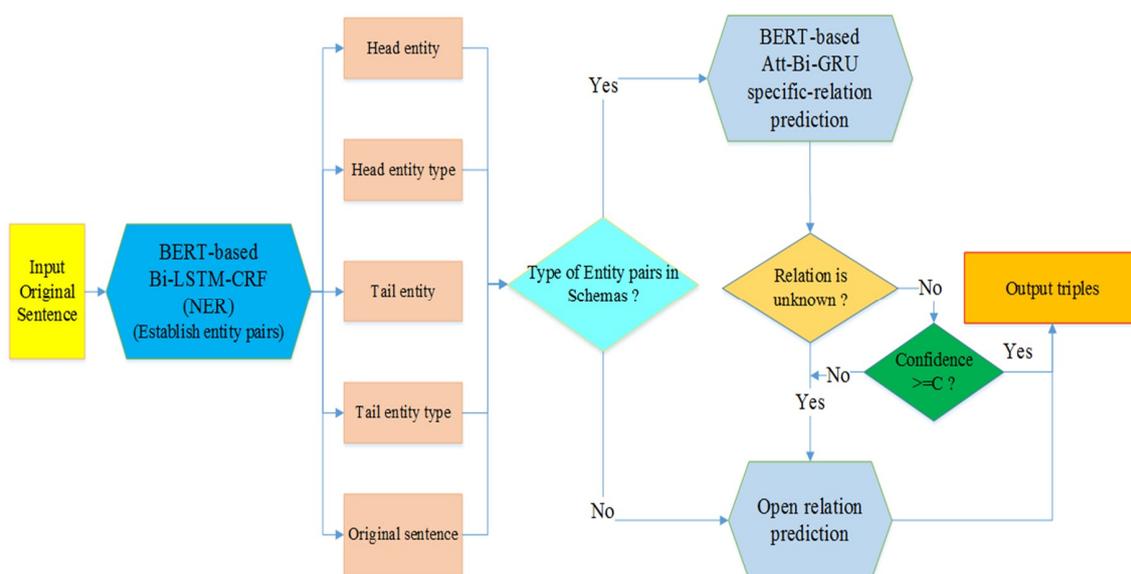


Figure 1. The framework of open domain Chinese triples hierarchical extraction method.

3.2. Named Entity Recognition Module

According to the aims of the named entity recognition module, the module implements two main functions. First, it accurately recognizes all named entities from the original sentence, including the entity name and the corresponding entity type. Second, it combines some simple rules to establish the entity pairs that may have a certain relation. Each entity pair includes five main parts: the head entity, the tail entity, their corresponding types, and the original sentence content. For the named entity recognition part, the BERT-based Bi-LSTM-CRF model is constructed. The framework of this model was shown in Figure 2.

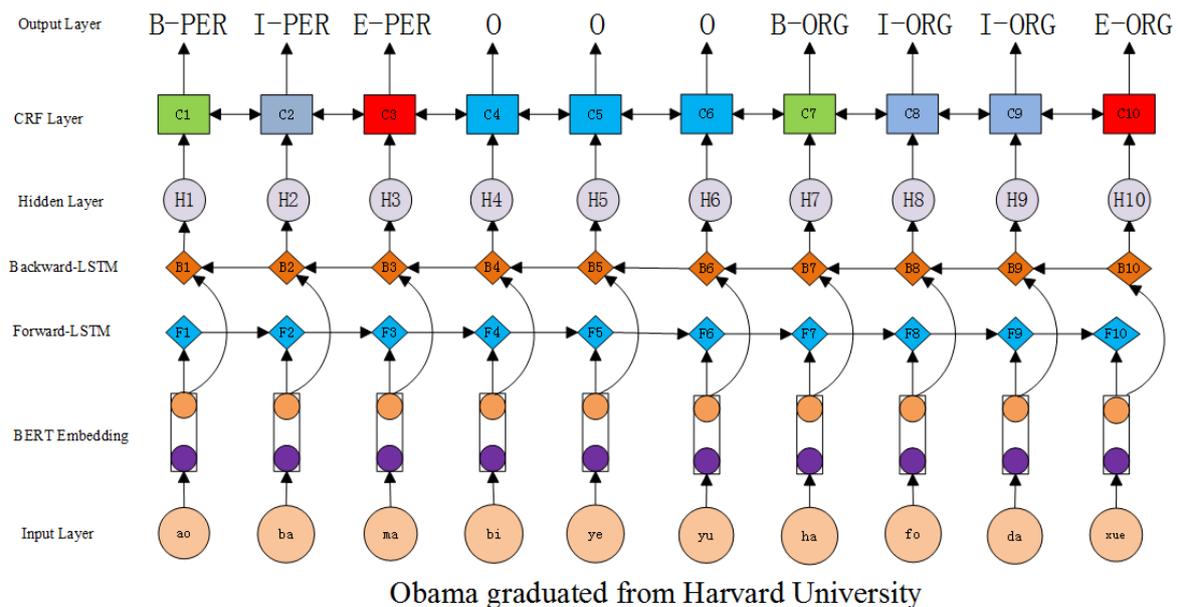


Figure 2. The framework of the BERT-based Bi-LSTM-CRF model.

In this model, a five-level BIOES tagging system is used to complete the tagging tasks for all corpus. The embedding layer of the model implements the conversion of character vector by the Chinese-BERT-Base. This is then followed by a Bi-LSTM network layer, a hidden layer, and a CRF layer (for the detailed calculation process of these layers, please see reference [9]). Finally, the sequence labeled results are output. According to the results of the entity list, we can quickly obtain all entity pairs. First, these entities are sorted based on the position in which they first appeared in the original sentence. Then, from the beginning to end, they are combined to establish the entity pair with one-to-one matching. The entity that appears at the front is determined as the head entity, and the entity that appears last is the tail entity. For example, we input an original sentence “Obama graduated from Harvard University”. The named entity recognition model output the entity list as [Obama-PER, Harvard University-ORG], and we can establish the entity pair as (Obama, PER, Harvard University, ORG, Obama graduated from Harvard University). Finally, all reasonable entity pairs are selected based on the filtering rules of the entity pair type. For example, if the type of the head entity and the tail entity is DATE-to-DATE, there may be no meaningful relation between the two entities, and it should be removed from the entity pairs list. However, if the type of entity pair is PER-to-ORG, it will be retained. Thus, we can identify all potential entity pairs.

3.3. Schema Match Module

The aim of the schema match module is to select certain entity pairs with a specific pattern based on the type of head entity and tail entity. These patterns can be designed according to the user’s requirements. For example, the entity pair pattern that the user mostly cares about is the head entity type as person and the tail entity type as local. Then, we add a pattern (subject_type: PER, object_type: LOC)

to the schema. It is worth noting that each pattern may map to multiple relations. For the case pattern above, the possible relations are birthplace or nationality. For any entity pair, if there is a corresponding pattern in the schemas, the BERT-based Att-Bi-GRU model will be used to obtain the specific relation. Otherwise, the open relation prediction module will be used to obtain the open relation.

3.4. BERT-Based Att-Bi-GRU Specific Relation Prediction Module

The main goal of this module is to accurately predict the entity pair relation with a specific pattern. If the entity pair meets this pattern, the trained BERT-based Att-Bi-GRU model is used to predict the relation. For example, we input the entity pair (Obama, PER, Harvard University, ORG, Obama graduated from Harvard University), and output the possible prediction result as Graduated_University (Obama, Harvard University). The framework of the BERT-based Att-Bi-GRU is shown in Figure 3. The model contains Input, Word segmentation, BERT embedding, Bi-GRU, Attention, Concat, Normalized, and Output layers. As a supervised method, it requires annotation corpora to complete the training of the model. The input sample includes a head entity, a tail entity, an original sentence, and the relation. For the word segmentation layer, we used the pyhanlp (pyhanlp-0.1.66-cp35) module. It is important to note that to improve the accuracy of segmentation, all of the entities from named entity recognition results are added into the user-defined dictionary in the segmentation stage.

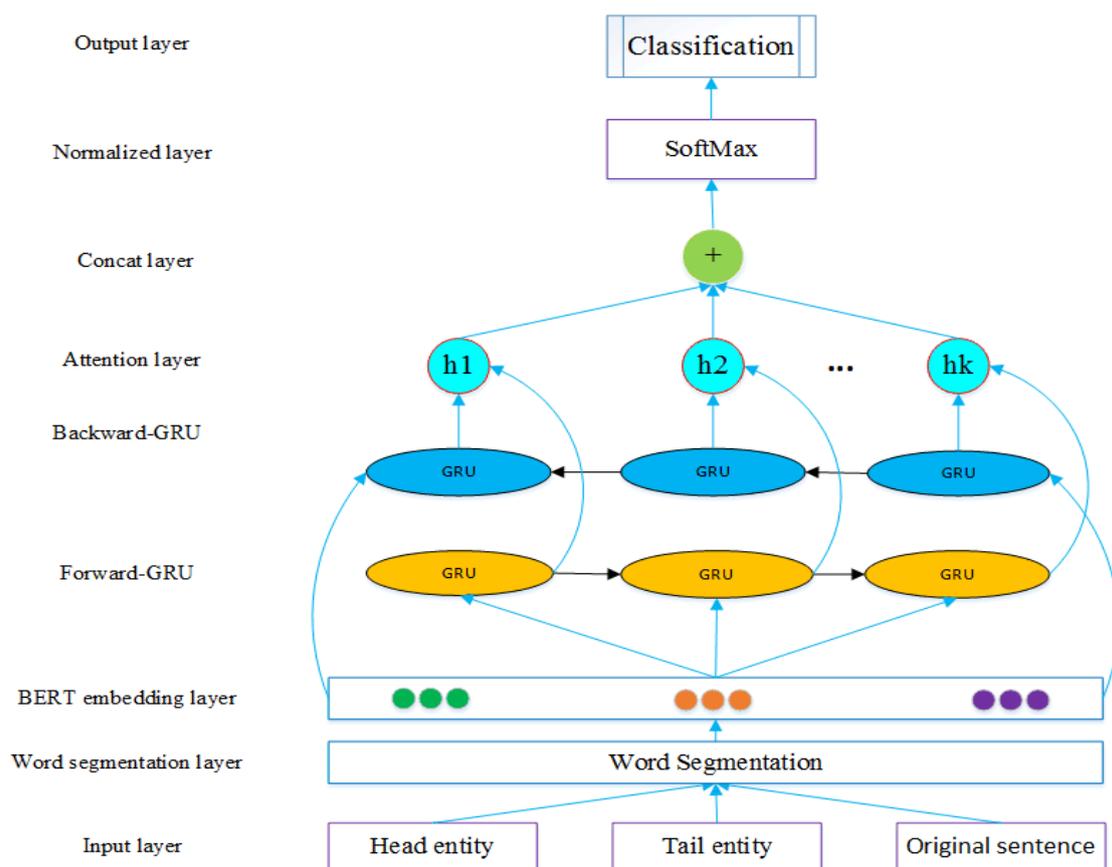


Figure 3. The framework of the BERT-based Att-Bi-GRU model.

The BERT embedding layer is based on the fine-tuned pre-trained Chinese-BERT-wwm-ext vector embedding model to obtain the vector transform results $[x_1, x_2, \dots, x_k]$ of all words. The remaining layers are implemented using the functions provided by Tensorflow (Tensorflow-1.12.0-cp35).

The calculation formulas of the Bi-GRU layer and attention layer are as follows. For each word t , the GRU cell unit computes h_t with input x_t and previous state h_{t-1} , as:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{1}$$

$$u_t = \sigma(W_u x_t + U_u h_{t-1}) \tag{2}$$

$$\tilde{h}_t = \tanh(W_c x_t + U(r_t \odot h_{t-1})) \tag{3}$$

$$h_t = (1 - u_t) \odot h_{t-1} + u_t \odot \tilde{h}_t \tag{4}$$

where h_t is hidden state, r_t is reset gate, and u_t is update gate. W_r, W_u, W_c , and U_r, U_u, U represent the parameters of GRU. σ is sigmoid function, and \odot refers to the production with element-wise.

For the word t , we use the hidden state $\xrightarrow{h_t}$ and $\xleftarrow{h_t}$ to represent the encode results from the forward and backward GRU. Then, we use the concatenation $h_t = [\xrightarrow{h_t}; \xleftarrow{h_t}]$ as the output of the Bi-GRU layer of word t .

For the task of relation classification of the entity pair, to find the hidden features between the head entity and tail entity, we introduced a position detection attention mechanism to calculate the weight of each word. The feature vector h_T of the word t is defined as a weighted sum, which is computed as follows:

$$h_T = \sum_{i=1}^k \alpha_i h_i \tag{5}$$

where h_i is the i -th cell unit output of the Bi-GRU layer. In addition, the α_i was calculated as followed:

$$\alpha_i = Aw_i + B \tag{6}$$

$$w_i = \begin{cases} 0, & t \notin \text{Set}[\text{head entity}, \text{tail entity}] \\ 1, & t \in \text{Set}[\text{head entity}, \text{tail entity}] \end{cases} \tag{7}$$

where, $\text{Set}[\text{head entity}, \text{tail entity}]$ is the word set between the head entity and tail entity in the original sentence, k is the total number of words in the original sentence, t is the t -th word in the original sentence, and A, B are the parameters of the network. In our method, we pay more attention to the words between the head entity and tail entity.

3.5. Relation Check and Confidence Discrimination Module

In this module, the main goal is to automatically check the relation. First, it will check the unknown relation. Unknown relation means that the current model cannot predict its relation and proceed directly to the open relation prediction module. Second, if the relation is known, then the model will give the relation R with the largest probability value. At this time, the probability value p (calculated by the SoftMax function) will be mapped to the confidence level CL . Next, it is determined if the confidence level is lower than a certain threshold value C . If $CL \geq C$, the relation R will be output, and the prediction ends; otherwise, it means that the confidence of the predicted relation is low, and it will directly go to the open relation prediction module.

3.6. Open Relation Prediction Module

This module aims to achieve entity relation prediction results under non-specific, unknown classification, and low-confidence specific patterns via sentence semantic dependency parse. In this module, we designed the extraction function to realize the prediction of open relations. First, we used the pyhanlp module to derive the CONLL format of the sentence. Second, we use an open relation extraction algorithm (#ORE Algorithm) to extract the open relation.

Algorithm 1 #ORE Algorithm**Input:** head entity, tail entity, sentence dependency parse result (CONLL format).**Parameter:** ParentID, Subject, Predicate, Object**Output:** predicted relation between head entity and tail entity.

```

1:   Let ParentID = -1, Subject = [], Predicate = [], Object = []
2:   For W in CONLL.list:
3:     if W.rel == 'HED':
4:       ParentID=W.ID
5:       Predicate.append(W.word)
6:     if W.parentID==ParentID:
7:       if W.rel == 'CMP':
8:         Predicate.append(W.word)
9:         break
10:    else:
11:      break
12:  For W in CONLL [0:ParentID]:
13:    if W.rel in ['SBV'] and W.parentID <= ParentID:
14:      Subject.append(W.word)
15:      break
16:  For W in CONLL[ParentID:len(CONLL.list)]:
17:    if W.rel in ['VOB', 'IOB', 'FOB', 'POB'] and W.parentID >= ParentID:
18:      Object.append(W.word)
19:      break
20:  if (Subject[0] in head entity) and (Object[0] in tail entity) and (len(Predicate[0])>0):
21:    Return "".join(Predicate)
22:  else:
23:    Return 'unknown'

```

To improve the accuracy of the sentence dependency parse, we also add all of the entities from named entity recognition results into the user-defined dictionary in the segmentation stage. Figure 4 shows the result of the sample sentence dependency parse of Section 3.2.

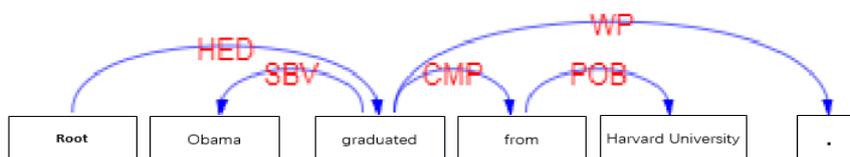


Figure 4. The result of the sample sentence dependency parse.

From Figure 4, it is easy to find that the relation of the entity pair (奥巴马: Obama, 哈佛大学: Harvard University) is (毕业于: graduated), because the result of sentence dependency parse with (毕业于: graduated) is Root->(HED+CMP).

4. Experiments

Dataset. We finished the experiments based on the public dataset of the Chinese Language and Smart Technology Information Extraction (CLSTIE). The framework of our method consists of the named entity recognition and specific relation prediction. First, we constructed a dataset #CLSTIE-NER (named entity recognition based on CLSTIE). For the specific relation prediction task, we used a sub-dataset #RC from CLSTIE. Our train-val-test dataset used a division ratio of 0.8:0.15:0.05. The statistical information of the two datasets is shown in Tables 1 and 2.

Table 1. The statistical info of the Chinese Language and Smart Technology Information Extraction named entity recognition (#CLSTIE-NER) dataset.

Corpus Language	Samples Size	Entity Categories	Entity Size
Chinese	194,747	25	860,000+

Table 2. The statistical info of #RC dataset.

Corpus Language	Samples Size	Schema Size	Specific Relation Size	Triples Size
Chinese	194,747	22	33	430,000+

We also constructed a small-scale independent test dataset, #ORP-Test, from Chinese Wikipedia to evaluate our proposed entire open relation hierarchical prediction method. It covered a total of 500 representative sentences, and all of the entity pairs and relations were labeled by humans.

Metrics. We measured all of the results in terms of Precision (P), Recall (R), and F_1 -score (F_1) in our experiments. It should be noted that F_1 -score is a comprehensive indicator, and its calculation is as follows:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (8)$$

This metric combines the effects of Precision and Recall. When the F_1 -score is higher, it shows that the method has a better performance. In our experiment, the micro-average method was used for the above indicators.

4.1. Experimental Design

The experiment fully evaluates the performance of our models in named entity recognition, specific relation prediction, and open relation prediction tasks. We designed three different comparison experiments. The first used the #CLSTIE-NER dataset to measure the performance of the BERT-based Bi-LSTM-CRF model. Secondly, the CLSTIE-train and CLSTIE-dev sub-dataset #RC was used to verify the performance of the BERT-based Att-Bi-GRU model. Finally, we used the #ORP-Test dataset to measure the entire open relation hierarchical prediction method.

4.2. Experimental Results and Analysis

Experimental parameter settings. In our experiments, there were some differences in core parameter settings between different models. The detailed description is shown in Table 3.

Table 3. The core parameters settings of fine-tuned BERT-based encoding model.

Model	Name	Value	Name	Value
BERT-based Bi-LSTM-CRF	max_seq_length	128	batch_size	4
	learning_rate	5×10^{-5}	num_labels	25
	dropout_keep_prob	0.5	train_epoch	10
	task_name	NER	bert_model_dir	chinese_L-12_H-768_A-12
BERT-based Att-Bi-GRU	max_seq_length	128	batch_size	8
	learning_rate	2×10^{-5}	num_relations	33
	dropout_keep_prob	0.5	train_epoch	12
	task_name	RC	bert_model_dir	chinese_wwm_ext_L-12_H-768_A-12

In the experiment, considering that our machine configuration was limited, we set the maximum batch_size to 8. It is strongly recommended that that is updated according to the actual configuration of the machine being used.

Results and Analysis. According to the introduction of the experimental design section, to evaluate different tasks, three groups of comparative experiments were carried out. For the NER task, we conducted

relevant experiments on the #CLSTIE-NER dataset, and selected CRF [8] and Bi-LSTM-CRF [9] as the baseline models. The experimental results are shown in Table 4.

Table 4. The optimal results of different models on #CLSTIE-NER dataset.

Model Name	Ave Precision	Ave Recall	Ave F1-Score (%)
CRF	85.23	77.64	81.26
Bi-LSTM-CRF	87.34	82.15	84.67
BERT-based Bi-LSTM-CRF	90.51	85.27	87.81

Compared with the baseline models, the results in Table 4 show that our model was superior among all indicators, and the average Precision, Recall, and F₁-score reached 90.51%, 85.27%, and 87.81%, respectively, on the experimental dataset. The performance of the CRF model was slightly worse, with an F₁-score of 81.26%.

For the specific relation prediction task, we conducted relevant experiments on the #RC dataset, and selected DepNN [12], Att-based Bi-LSTM [13], Att-based CNN [16], and BERT-based Att-Bi-LSTM as the baseline models. The experimental results are shown in Table 5.

Table 5. The optimal results of different models on #RC dataset.

Model Name	Ave Precision	Ave Recall	Ave F1-Score (%)
DepNN	83.19	81.04	82.1
Att-based Bi-LSTM	88.49	86.62	87.55
Att-based CNN	89.57	87.03	88.28
BERT-based Att-Bi-LSTM	89.37	86.95	88.14
BERT-based Att-Bi-GRU	89.81	87.88	88.83

For the specific relation prediction task, the results in Table 5 show that our model was better than all baseline models. Relative to the DepNN method, our model indicators increased by more than 6%. After adding BERT encoding, the classification performance of the model was effectively improved.

For the entire open relation prediction task, we conducted the experiment on the #ORP-Test dataset and selected ZORE [20], UnCORE [21], and DSNFs-based ORE [22] as the baseline models. The results are shown in Table 6.

Table 6. The optimal results of CTHE method on #ORP-Test.

Model Name	Ave Precision	Ave Recall	Ave F1-Score (%)
ZORE	62.39	69.51	65.76
UnCORE	80.43	49.76	61.48
DSNFs-based ORE	83.67	59.12	69.28
CTHE (Ours)	84.74	61.39	71.2

According to the results of Table 6, our model had a higher comprehensive performance than the other three baseline models. Precision and F₁-score reached 84.74% and 71.2%, and were at least 1.07% higher than those of the baseline models.

Case Analysis. Here, two simple cases are shown in Table 7 to illustrate our method.

In Table 7, the predict entity list refers to the named entity recognition results obtained by the algorithm, and the target entity list refers to the results of human annotation. In addition, the predicted triples represent all triples predicted by the algorithm, and target triples represent the results of

all triples labeled by humans. According to the results, we know that named entity recognition performance is reliable. However, in open relation prediction, the effect is better for short sentences, and for longer compound sentences, some triples may be lost. For example, the triple President (Donald Trump, US) was lost in case II.

Table 7. Case analysis of CTHE method.

Case I	奥巴马毕业于哈佛大学。 Obama Graduated from Harvard University.
Predict Entity list	[奥巴马, 哈佛大学] [Obama, Harvard University].
Target Entity list	[奥巴马, 哈佛大学] [Obama, Harvard University].
Predicted triples	毕业于(奥巴马,哈佛大学) [Graduated (Obama, Harvard University)].
Target triples	毕业于(奥巴马,哈佛大学) [Graduated (Obama, Harvard University)].
Case II	唐纳德·特朗普生于纽约,是美国第45任总统。 Born in New York, Donald Trump is the 45th president of the US.
Predict Entity list	[美国, 唐纳德·特朗普, 纽约] [US, Donald Trump, New York].
Target Entity list	[美国, 唐纳德·特朗普, 纽约] [US, Donald Trump, New York].
Predicted triples	国籍(美国,唐纳德·特朗普) [Nationality(US, Donald Trump)], 出生地(纽约,唐纳德·特朗普) [Birthplace(New York, Donald Trump)].
Target triples	总统(唐纳德·特朗普, 美国) [President(Donald Trump, US)], 国籍(美国, 唐纳德·特朗普) [Nationality(US, Donald Trump)], 出生地(纽约, 唐纳德·特朗普) [Birthplace(New York, Donald Trump)].

5. Discussion

Here we provide a discussion based on the two dimensions of parameter sensitivity analysis and shortages summary.

Sensitivity Analysis. Here we conduct core parameter sensitivity analysis for BERT-based Bi-LSTM-CRF, BERT-based Att-Bi-GRU, and CTHE methods. For the first two methods, we analyzed the number of training epochs, and during the open relation prediction, we analyzed the value of confidence threshold C; the analysis results are illustrated in Figure 5.

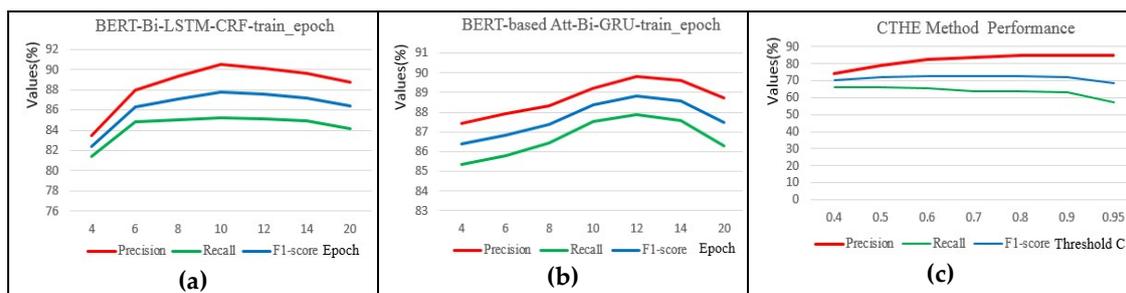


Figure 5. The analysis results of parameter sensitivity.

From Figure 5a,b we can observe that the number of epochs has a large effect on experimental performance. When the number of epochs was 10 and 12, the performance of BERT-based Bi-LSTM-CRF and BERT-based Att-Bi-GRU models is optimal. From Figure 5c, we find that the CTHE method achieved a higher F₁-score when the confidence threshold C-value was between 0.6 and 0.9. In our experiment, we set it to 0.85 because it took some time to train the model. During the tuning process, we only conducted 12 different experiments. Therefore, we expect future researchers can further optimize the parameter settings of the model.

Shortages Summary. First, our hierarchical prediction method is dependent on the accuracy of named entity recognition. If the performance of the entity recognition model is too low, some entities will be lost. Second, in the open relation prediction module, relation losses may occur even if

we dynamically add entities to a custom dictionary to improve the accuracy of segmentation and dependency parse.

Second, the experimental results show that our method is effective for entity-relation prediction of non-composite sentences, and it may miss triples for entity-relation extraction from some long compound sentences. Finally, it is worth noting that our model can only predict one kind of relation between entity pairs, and it cannot predict multiple relations of entity pairs. Our method only supports multiple relation prediction for the specific relation prediction between entity pairs. For the prediction of open relations, it cannot support multiple relation prediction. Therefore, as an integrated knowledge extraction method, it does not support the prediction of multiple relations between entity pairs.

6. Conclusions and Future Work

This article proposes an open domain Chinese triples hierarchical extraction method that resolves open relation triples extraction problems. It builds an integrated method for open relation triples extraction by combining deep learning and unsupervised learning. Experiments on the human annotated test dataset showed good performance.

The biggest advantage is that our method has the ability of both supervised and unsupervised learning methods. For the triples extraction task, our method completes not only specific relation prediction with Chinese entity pairs but also open relation prediction for Chinese entity pairs.

Important future research directions are improving the performance of open relation prediction by combining external knowledge graphs and enhancing the architecture of the method to support multiple relation prediction between Chinese entity pairs.

Author Contributions: Conceptualization, C.H. and H.W.; methodology, C.H.; software, C.H.; validation, Z.T., C.Z. and B.G.; formal analysis, C.H. and Y.H.; investigation, H.W. and Z.T.; resources, C.H. and B.G.; data curation, H.W. and C.H.; writing—original draft preparation, C.H.; writing—review and editing, Z.T. and C.H.; visualization, C.Z. and Y.H.; supervision, H.W. and B.G.; project administration, Z.T. and B.G.; funding acquisition, Z.T. and B.G.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received the support of the National Natural Science Foundation of China (NSFC) via grant No. 61902417 and No. 61872446, and it is also funded by Basic Foundation via grant No. 2019-JCJQ-JJ-231.

Acknowledgments: The authors thank all of the reviewers for their comments on this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Figure A1 shows the results of case II (Table 7) in a local triples extraction display system.

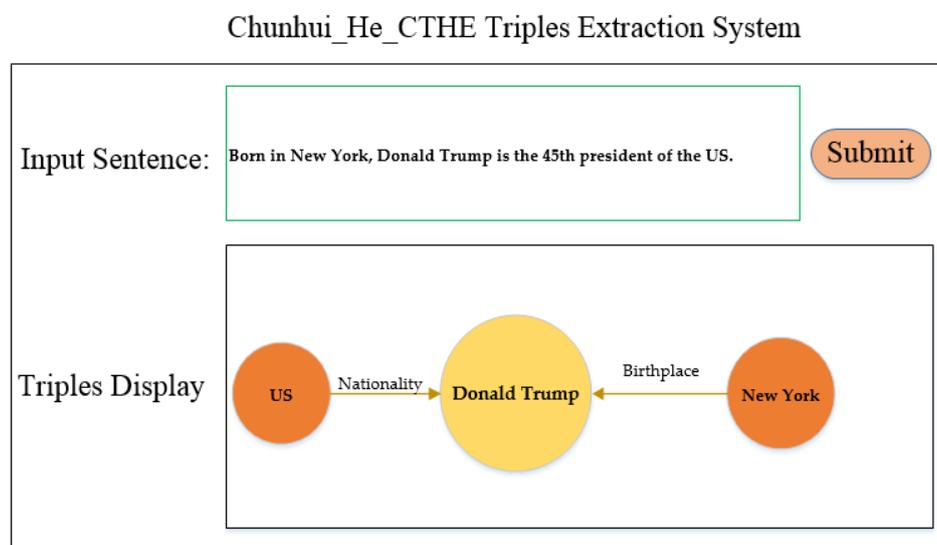


Figure A1. The extraction results of case II (Table 7) in our local display system.

References

1. Wu, R.; Yao, Y.; Han, X.; Xie, R.; Liu, Z.; Lin, F.; Lin, L.; Sun, M. Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019; pp. 219–228.
2. Gao, T.; Han, X.; Xie, R.; Liu, Z.; Lin, F.; Lin, L.; Sun, M. Neural snowball for few-shot relation learning. *arXiv* **2019**, arXiv:1908.11007. [[CrossRef](#)]
3. Zhang, M.; Zhang, J.; Su, J.; Zhou, G. A composite kernel to extract relations between entities with both flat and structured features. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, July 2006; pp. 825–832. [[CrossRef](#)]
4. Zelenko, D.; Aone, C.; Richardella, A. Kernel methods for relation extraction. *J. Mach. Learn. Res.* **2002**, *10*, 71–78. [[CrossRef](#)]
5. Mooney, R.J.; Bunescu, R. Subsequence kernels for relation extraction. In Proceedings of the 18th International Conference on Neural Information Processing System, Vancouver, BC, Canada, 5–8 December 2005; pp. 171–178.
6. Zhou, G.; Zhang, M.; Ji, D.; Zhu, Q. Tree kernel-based relation extraction with context-sensitive structured parse tree information. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CONLL), Prague, Czech Republic, 28–30 June 2007; pp. 728–736.
7. Peng, N.; Poon, H.; Quirk, C.; Toutanova, K.; Yih, W.T. Cross-sentence n-ary relation extraction with graph LSTMs. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 101–115. [[CrossRef](#)]
8. Wu, C.W.; Jan, S.Y.; Tsai, R.T.H.; Hsu, W.L. On using ensemble methods for Chinese named entity recognition. In Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, Sydney, Australia, July 2006; pp. 142–145.
9. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.
10. Wang, G.; Yu, Y.; Zhu, H. PORE: Positive-only relation extraction from wikipedia text. *Comput. Vision* **2007**, *4825*, 580–594.
11. Minard, A.L.; Ligozat, A.L.; Grau, B. Multi-class SVM for relation extraction from clinical reports. In Proceedings of the International Conference Recent Advances in NLP 2011, Hissar, Bulgaria, 12–14 September 2011; pp. 604–609.
12. Liu, Y.; Wei, F.; Li, S.; Ji, H.; Zhou, M.; Wang, H. A Dependency-based neural network for relation classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Beijing, China, 26–31 July 2015; pp. 285–290.
13. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.-W.; Xu, B. 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), Berlin, Germany, 7–12 August 2016; pp. 207–212.
14. Dos Santos, C.; Xiang, B.; Zhou, B. Classifying relations by ranking with convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 626–634.
15. Wang, L.; Cao, Z.; De Melo, G.; Liu, Z.; Erk, K.; Smith, N.A. Relation Classification via Multi-Level Attention CNNs. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 1298–1307.
16. Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; Sun, M.; Erk, K.; Smith, N.A. Neural relation extraction with selective attention over instances. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 2124–2133.
17. Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; Xu, B.; Barzilay, R.; Kan, M.Y. Joint extraction of entities and relations based on a novel tagging scheme. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017.

18. Tan, Z.; Zhao, X.; Wang, W.; Xiao, W. Jointly extracting multiple triplets with multilayer translation constraints. In Proceedings of the AAAI Conference on Artificial Intelligence; Association for the Advancement of Artificial Intelligence (AAAI), Honolulu, HI, USA, 27 January–1 February 2019; pp. 7080–7087.
19. Yu, B.; Zhang, Z.; Su, J. Joint extraction of entities and relations based on a novel decomposition strategy. *arXiv* **2019**, arXiv:1909.04273.
20. Qiu, L.; Zhang, Y. ZORE: A Syntax-based System for Chinese Open Relation Extraction. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1870–1880.
21. Bing, Q.; Anan, L.; Ting, L. Unsupervised Chinese open entity relation extraction. *J. Res. Dev.* **2015**, *52*, 1029–1035.
22. Jia, S.; E, S.; Li, M.; Xiang, Y. Chinese open relation extraction and knowledge base establishment. *ACM Trans. Asian Low Resour. Lang. Inf. Process.* **2018**, *17*, 1–22. [[CrossRef](#)]
23. Vaswani, A.; Shazeer, N.; Parmar, N. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
24. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
25. Peng, C. *Bert Based Relation Classification Network for Inter-Personal Relationship Extraction*; CCKS2019-Shared Task; GridSum Tech.: Shenzhen, China, 2019.
26. Zhu, Z.; Su, J.; Zhou, Y.; Yu, S. Improving distantly supervised relation classification with attention and semantic weight. *IEEE Access* **2019**, *7*, 91160–91168. [[CrossRef](#)]
27. Park, S.S.; Kim, H. Relation extraction among multiple entities using a dual pointer network with a multi-head attention mechanism. In Proceedings of the Second Workshop on Fact Extraction and Verification (FEVER); Association for Computational Linguistics (ACL), Hong Kong, China, 3–7 November 2019; pp. 47–51.
28. Cui, L.; Wei, F.; Zhou, M. Neural open information extraction. *arXiv* **2018**, arXiv:1805.04270.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).