

Research Article

A Novel Chinese Entity Relationship Extraction Method Based on the Bidirectional Maximum Entropy Markov Model

Chengyao Lv ¹, Deng Pan,² Yaxiong Li ³, Jianxin Li,⁴ and Zong Wang⁵

¹School of Foreign Languages, China University of Geosciences, Wuhan 430074, China

²School of Foreign Languages, Hubei University of Science and Technology, Xianning 437000, China

³Information Centre, Hubei University of Science and Technology, Xianning 437000, China

⁴School of Information Technology, Deakin University, Geelong VIC3220, Australia

⁵School of Public Administration, China University of Geosciences, Wuhan 430074, China

Correspondence should be addressed to Yaxiong Li; liyaxiong@hbust.edu.cn

Received 9 October 2020; Revised 14 December 2020; Accepted 5 January 2021; Published 19 January 2021

Academic Editor: Jia Wu

Copyright © 2021 Chengyao Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To identify relationships among entities in natural language texts, extraction of entity relationships technically provides a fundamental support for knowledge graph, intelligent information retrieval, and semantic analysis, promotes the construction of knowledge bases, and improves efficiency of searching and semantic analysis. Traditional methods of relationship extraction, either those proposed at the earlier times or those based on traditional machine learning and deep learning, have focused on keeping relationships and entities in their own silos: extracting relationships and entities are conducted in steps before obtaining the mappings. To address this problem, a novel Chinese relationship extraction method is proposed in this paper. Firstly, the triple is treated as an entity relation chain and can identify the entity before the relationship and predict its corresponding relationship and the entity after the relationship. Secondly, the Joint Extraction of Entity Mentions and Relations model is based on the Bidirectional Long Short-Term Memory and Maximum Entropy Markov Model (Bi-MEMM). Experimental results indicate that the proposed model can achieve a precision of 79.2% which is much higher than that of traditional models.

1. Introduction

In the age of big data, techniques of extracting valuable information from enormous quantities of texts have drawn the attention of many researchers. The extraction of information includes entity extraction, relationship extraction, and event extraction. As the key step in information extraction, relationship extraction provides technical foundation for subsequent tasks such as knowledge graphs, intelligent information retrieval, and semantic analysis. Therefore, techniques of relationship extraction are beneficial not only for theoretical discussion but also for practical application.

Research on techniques to extract entities and their relationships can date back to the 1960s. Among the more prominent projects is the Linguistic String Project by New

York University, which took the route of constructing massive language (English) corpora and achieved very satisfactory results when the team used these corpora to extract information from medical texts. In addition, a systematic research at Yale University extracted events in domains such as “earthquake” and “strike” from news texts and promoted the research and development of entity relationship extraction. By the late 1980s, with the convening of the Message Understanding Conference, research on entity relationship extraction had started to boom. After decades’ development, theories and techniques of entity relationship extraction, from early models of manual design and rule extraction [1, 2] to late models based on machine learning [3] and deep learning [4, 5], are approaching maturity. With constant improvements in model accuracy and recall, extraction models are more adaptive than ever before.

However, the most existing extraction techniques either have been keeping relationships and entities in their own silos. Extracting relationships and entities was conducted in separate steps before obtaining the mappings, or tag triples as a whole used the “proximity principle” of reinforced learning to extract relationships. Existing extraction techniques fit into three categories. Firstly, the relationship can be predicted and identified by an entity pair. The premise of this idea is that the relationships are already predefined [6]. The task of relationship extraction then becomes the task of searching the predefined relationship space for the most probable relationship between a given entity pair based the context where the entity pair is located. Secondly texts can be explored by the relationship of entity pairs. This method aims at finding the maximum number of entity pairs matching the criteria of the given relationship. A common issue of the two methods mentioned above is the subtasks, entity identification and relationship identification, are completely independent of each other, resulting in extraneous information such as entities without relationship. This, in turn, increases error rates because the entities are paired up before their relationship is determined; when no relationship is found for an entity pair, this pair becomes extraneous. Such extraneous pairs increase error rates of the subtask and negatively impact the performance of subsequent relationship classification. Finally, some studies tag triples as a whole and use the “proximity principle” of reinforced learning to extract relationships [7]. This method integrates low-level features into more abstract high-level features to search for distributed feature representations and, thus, solves the problems of manual feature selection and the spread of feature extraction error haunting classical methods.

The conventional method has two drawbacks. Firstly, for most of the entity pairs do not hold relationships, numerous negative cases and imbalanced relationship classification occur. Secondly, overlapping triples become a critical issue. The shared entities or multiple relationships between two entities make learning more complicated or even impossible, since adequate training data cannot be obtained. For instance, “Mr. Zhang was born in Hubei, a province in Central China” could be interpreted into $\langle \text{Mr. Zhang, was born in, Hubei} \rangle$, $\langle \text{Mr. Zhang, was born in, China} \rangle$, and $\langle \text{Hubei, lies in, China} \rangle$. The conventional algorithm cannot identify and classify properly without sufficient data.

To address these problems, this paper proposes a new method, entity relation chain. The head entity before relationship should be identified firstly, and then, the corresponding relationship and the tail entity can be predicted. For instance, in the sentence “Mr. Zhang was born in Hubei province,” E_1 “Mr. Zhang” and E_2 “Hubei province” are usually identified firstly and the R “was born in” is recognized secondly. But, in the entity relation chain, E_1 “Mr. Zhang” is firstly identified, and every possible R generated from E_1 is the criterion for E_2 “Hubei province.” In this entity relation chain, E_1 can be taken as head entity, R as relation chain, and E_2 as tail entity.

Experiments on data sets from People’s Daily indicated that the proposed method can achieve a high performance.

We also evaluated the scalability of the method on English data sets of the English SemEval 2010 Task 8 which reveal that the Bi-MEMM also can obtain a better f-score.

This paper is organized as follows. Starting with the introduction of the research gap and our research purpose, we review and discuss the entity relationship extraction and the particularity of Chinese relation extraction. Then, we develop the Bi-MiEM method for the entity relation extraction. The detailed experimental evaluation is illustrated in Section 4, and Section 5 concludes this work and provides the future direction for further research.

2. Related Work

2.1. Definition of Entity Relationship Extraction. Entity relationship extraction is usually described as entity relationship triples $\langle E_1, R, E_2 \rangle$, in which E_1 and E_2 refer to the entity type and R refers to the relation description type text. After the preprocessing process of named entity recognition relation trigger word recognition, the determined triples $\langle E_1, R, E_2 \rangle$ are stored for further analysis or query.

According to the definition, we can divide the entity relationship extraction tasks into three key parts, name entity recognition, relation trigger word identification, and relation extraction. Name entity recognition refers to the identification of text having a specific meaning of the entity, mainly including the names of people and places, institutions, and proper nouns. Relation trigger word identification is to classify the words that trigger entity relationship, identify whether they are trigger words, and determine whether the extracted relations are positive. Relation extraction is the extraction of semantic relationships between entities from identified entities, such as location employee products.

2.2. Features of Entity Relationship Extraction. Compared with NLP tasks such as sentiment analysis and news classification, the extraction of relationship is unique in three aspects.

Firstly, Entity Relationship Extraction covers diverse domains. Researchers usually focus on one domain or a limited number of domains. With limited relationship categories, traditional techniques are mostly based upon rules [2, 8–10], dictionaries [1, 11], and ontologies [3, 12]. Machine learning-based techniques include supervised [6, 13], semisupervised [14, 15], and unsupervised [16, 17] models. Lately, deep learning-based techniques include supervised [18, 19] and distant supervised [20] models. All these models are relatively easy to build, but with poor portability and extensibility.

Secondly, Entity Relationship Extraction involves heterogeneous data. Data can come from different sources, and they can be structured, semistructured, or nonstructured. Deep learning [21] is usually applied in structured data; nonsupervised aggregation methods [4] are usually applied in nonstructured textual data due to unpredictable relationship categories; semisupervised [17] or distant

supervised [22] methods are usually applied in semi-structured data such as Wikipedia.

Lastly, Entity Relationship Extraction needs to handle various relationships, which easily leads to data noise. Relationships between entities are various, but early research often ignored such multiple relationships and failed to handle latent relationships. The adoption of graph structures [18] in relationship extraction in recent years ushered in a new technique for tackling overlaps of entities and relationships. To tackle data noise [23], it has been discovered that using a small number of adversarial examples can avoid model overfitting and proposed to use adversarial training to improve model performance.

2.3. Particularity of Chinese Relation Extraction. Relationship extraction of Chinese texts falls behind the extraction of English because of its complexity and difficulty. The following two characteristics of Chinese make it more challenging for Chinese than English in terms of relationship extraction.

Chinese trigger words are hard to extract and are in abundance. This makes the recall rate of relationship extraction low. In the ACE corpus, Chinese trigger words are 30% more than those in English [24].

For the Chinese language, words are often polysemous, sentence structures are complex and flexible, and omissions appear frequently. The fact that the same word can express completely different meanings in different contexts or the same meaning can be represented with many different expressions makes the identification of relationship types particularly difficult.

In view of these problems, this paper proposes the following possible solutions. Firstly, the Joint Extraction of Entity Mentions and Relations model similar to Seq2Seq is proposed and the Bidirectional Maximum Entropy Markov is integrated into the model. Secondly, different from the existing relationship extraction techniques, relationship triples are treated as an entity relationship chain, entity E_1 is identified first, and then, the corresponding relationship R and entity E_2 based on E_1 are predicted. Thirdly, the validity of the proposed model is verified in Chinese data sets and the scalability is evaluated in English data sets.

3. Extraction Method Based on the Bi-MEMM Model

The previous solutions cannot efficiently deal with the entity relationship extraction entity overlap, relationship cross-over, and so on. In this paper, a Bi-MEMM model similar to seq2seq simulated probability graph is proposed to solve such problems. The seq2seq decoder is modeled in the following way:

$$P(y_1, y_2, \dots, y_n | x) = P(y_1 | x) P(y_2 | x, y_1) \dots P(y_n | x, y_1, y_2, \dots, y_{n-1}). \quad (1)$$

In formula (1), the first word is predicted by x and the second word is predicted if the first word is known and repeated until the end mark appears. Similarly, the extraction of triples can be modeled in the following way:

$$P(E_1, R, E_2) = P(E_1) P(E_2 | E_1) P(R | E_1, E_2). \quad (2)$$

In formula (2), " E_1 " can be predicted first, and " E_2 " corresponding to " E_1 " can be predicted by passing in " E_1 ". Then, E_1 and E_2 can be introduced to predict the relationship R between E_1 and " E_2 ." In actual processing, we can also combine the predictions of E_2 and R into one step, so the total step only needs two steps; the first step is to predict E_1 , and then, E_1 is introduced to predict E_2 and R corresponding to " E_1 ."

3.1. Bi-MEMM Model. Figure 1 demonstrates the overall structure of our Bi-MEMM model. It can be detailed as follows.

When it comes to techniques for extracting relationships and entities, character-word embedding is necessary only for Chinese, as word embedding is sufficient for English. By means of word segmentation with Chinese texts, we obtain character embedding and word embedding. Then, we perform matrix transformation of word embedding and concatenate the transformed word embedding with character embedding of the word's constituent characters. The result of such concatenation is character-word embedding. For instance, "中國" has two character-word embedding: one is the concatenation of the matrix-transformed word embedding with character "中", and the other is the concatenation of the matrix-transformed word embedding with character "國".

Firstly, character-word-position embedding is transformed into coding matrix M through the Bi-LSTM Layer and Tanh Layer/Attention Layer.

Secondly, matrix M is copied into the Bi-MEMM Layer and Dense Layer. Sigmoid can be used as activation function for the Dense Layer. Then, a two-dimension vector generated by each character can be used to predict the head and tail position of E_1 .

Thirdly, a labelled E_1 is randomly picked (randomly pick E_1 when training, and traverse all E_1 's when predicting), the subsequence corresponding to E_1 is fed in M into the first Self Attention Layer, together with the Position Embedding at corresponding position, and it transformed into a vector with the same length as the input sequence.

Lastly, matrix M is sent into the Bi-MEMM Layer and Dense Layer again. For each R corresponding to E_1 , the head and tail positions of E_2 can be also predicted by the Dense Layer with the activation function of sigmoid.

From the model structure of Figure 1, we can figure out it is similar to the copy mechanism, joint extraction model. In entity 'E1' identification, $\langle E_1, R, E_2 \rangle$, Bi-MEMM plays the same role as CRF. In E_2 recognition, Bi-MEMM predicts E_2 by every possible R with E_1 . If there is E_2 , the corresponding triples are regarded as an option or the triples will be discarded.

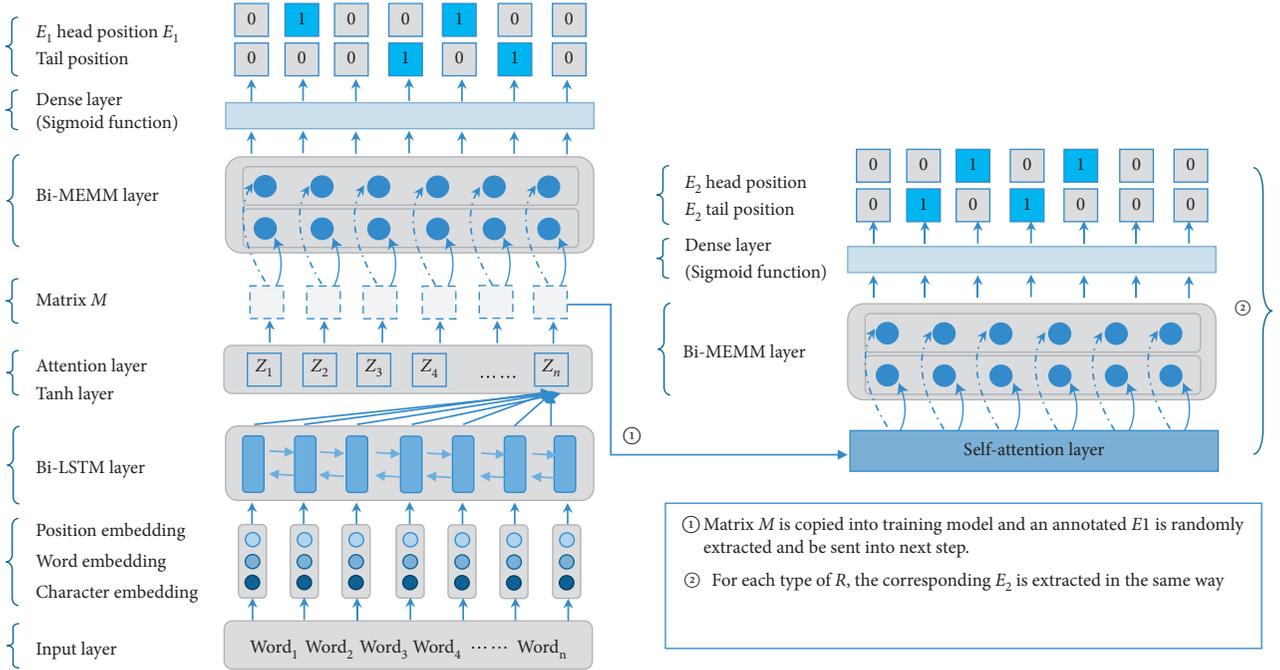


FIGURE 1: The structure of Bi-MEMM.

3.2. *Bi-MEMM Construction and the Loss Function.* In formula (1), we assume that the dependency occurs only in adjacent locations, and the following formula is obtained:

$$P(y_1, y_2, \dots, y_n | x) = P(y_1 | x) P(y_2 | x, y_1) P(y_3 | x, y_2) \dots P(y_n | x, \dots, y_{n-1}). \quad (3)$$

In formula (3), $X = (x_1, x_2, \dots, x_n)$ is the input and $Y = (y_1, y_2, \dots, y_n)$ is the tag sequence with the same size of X . According to the design of Linear CRF (Linear Chain Conditional Random Field), the following formula is obtained from formula (3):

$$P(y_1 | x) = \frac{e^{f(y_1; x)}}{\sum_{y_1} e^{f(y_1; x)}}, \quad (4)$$

$$P(y_k | x, y_{k-1}) = \frac{e^{g(y_{k-1}, y_k) + f(y_k; x)}}{\sum_{y_k} e^{g(y_{k-1}, y_k) + f(y_k; x)}}$$

$$P \leftarrow y | x \leftarrow = \frac{e^{f(y_1; x) + g(y_1, y_2) + \dots + g(y_{n-1}, y_n) + f(y_n; x)}}{\left(\sum_{y_1} e^{f(y_1; x)} \right) \left(\sum_{y_2} e^{g(y_1, y_2) + f(y_2; x)} \right) \dots \left(\sum_{y_n} e^{g(y_{n-1}, y_n) + f(y_n; x)} \right)}. \quad (5)$$

So far, we can see that MEMM, like seq2seq, has one significant defect: exposure bias [25]. When the model is trained, the prediction of the current step assumes that the labels of the previous step are correct and acquired. However, in the prediction stage, the actual labels of the previous

where $g(y_{k-1}, y_k)$ is called the transition matrix. At this point, this is the MEMM. From equation (4), we can see that the solution of the MEMM is to decompose the overall probability distribution into the product of a stepwise distribution, so to calculate the loss, you only need to sum the cross entropy of each step.

Substituting equation (4) into equation (3), we can get the loss of MEMM as follows:

step are unknown. If the current step is not strengthened during training, the reliability of the entire data chain will be greatly reduced.

The way to calculate the probability of equation (5) is from left to right. Experiments show that adding a right-to-

left MEMM during modelling with reference to the LSTM and Bi-LSTM modes can improve its effect. Then, we can get the following loss function.

Finally, the average cross entropy of formulae (5) and (6) are taken as the final loss. This can make up for the shortcomings of its asymmetric behaviour without increasing the parameters, and it can also strengthen the current training.

$$P_{y|x} = \frac{e^{f(y_1;x)+g(y_1,y_2)+\dots+g(y_{n-1},y_n)+f(y_n;x)}}{\left(\sum_{y_n} e^{f(y_n;x)}\right)\left(\sum_{y_{n-1}} e^{g(y_n,y_{n-1})+f(y_{n-1};x)}\right)\dots\left(\sum_{y_1} e^{g(y_2,y_1)+f(y_1;x)}\right)} \quad (6)$$

4. Experimental Design

Experiments are carried out to evaluate the efficiency of proposed method on Chinese data sets and the scalability on English data sets. For the Chinese data set, corpus data from People’s Daily in January in the news field are collected, and the English data set adopted SemEval 2010 Task 8.

Several similar methods such as Bi-LSTM + CRF [5], Att-Bi-LSTM + CRF [26], and bert-based [27] were taken as the base line on the Chinese entity relationship extraction test. The proposed joint extraction model is applied to Chinese data sets to verify its validity. BRNN [28], SDP-BLSTM [29], CNN [30], Att-RCNN [31], and Hybrid Bi-LSTM-Siamese [32] are also carried out as the base line for the scalability evaluation.

4.1. Data Sets. SemEval 2010 Task 8 marks the semantic relationship between noun pairs in a sentence rather than entity pairs. There are 10 classes (cause-effect, component-whole, entity-destination, product-producer, entity-origin, member-collection, message-topic, content-container, instrument-agency, and others) in total, among which one type does not distinguish the sequence of relationship arguments.

The corpus of People’s Daily mainly includes three kinds of entity relations, personal name, place name, and organization name. In this paper, Spacy [33], PyhanLP [34], and other natural language processing auxiliary tools [35] are used in experiments.

4.2. Hyperparameters. Due to differences in the data set of Chinese and English, for example, factors Embedding of China Character and Word Embedding of English are not consistent with some superparameters. In this paper, the average cross entropy of formula (6) is used as the loss function to train deep learning network with an Adam optimizer. The superparameters are shown in Table 1.

4.3. Evaluation Criteria. Precision, recall, and F-measure are adopted as the basic evaluation criteria, in which precision and recall are contradictory and F-measure is taken to

TABLE 1: Hyperparameters settings.

Hyperparameters	Chinese	English
Learning rate	0.0005	0.0005
Hidden neurons size	250	200
Dropout	0.6	0.5
Character embeddings/word embeddings	300	300
Batch size	16	32

evaluate comprehensively and globally. Their calculation formulae are listed, respectively, as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}, \quad (7)$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}, \quad (8)$$

$$F_1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (9)$$

4.4. Experimental Results and Analysis. For the Chinese entity relationship extraction dataset, Bi-LSTM-CRF, Att-Bi-LSTM-CRF, and bert-based are applied as benchmark for performance testing. Precision, recall, and F-score are used as the evaluation criteria. The precision of different methods is shown in Table 2. Their recall and F-score are displayed in Figures 2 and 3. For the English entity relationship extraction data set, the F-score of six models is listed in Table 3 for the scalability evaluation of the proposed model.

Table 2 displays the precision of Bi-LSTM-CRF, Att-Bi-LSTM-CRF, bert-based, and the F-scores of our methods varying from 72.5% to 79.2%. The proposed Bi-MEMM method enjoys the highest precision of 79.2%, while precision values of the other methods are 72.5%, 73.6% and 75.1%. In terms of recall and F-score shown in Figures 2 and 3, it can be concluded that our model performs efficiently with a highest recall of 80.4% and a outstanding F-score of 79.8%, while those of other methods are 71.6% and 72.05% (Bi-LSTM-CRF), 74.3% and 73.94% (Att-Bi-LSTM-CRF), and 76.3% and 75.69% (bert-based).

Bi-MEMM has some features which can overcome the pitiful of traditional methods while dealing with Chinese entity relationship extraction. Firstly, the MEMM model, like the CRF model, has an attractive feature with the convexity of its loss function. The Bi-MEMM model fundamentally solves the label bias problem of the MEMM model and can make full use of context information. It can use complex, overlapping, and nonindependent information for its training and inference. Compared with the CRF model, the performance of feature selection in the Bi-

TABLE 2: Chinese entity relationship extraction precision.

Methods	Precision
Bi-LSTM-CRF [5]	72.5
Att-Bi-LSTM-CRF [26]	73.6
Bert-based [27]	75.1
Our method	79.2

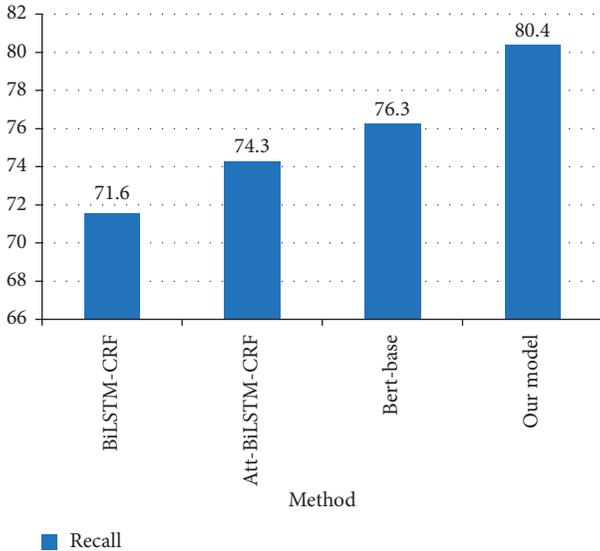


FIGURE 2: Recall of different methods.

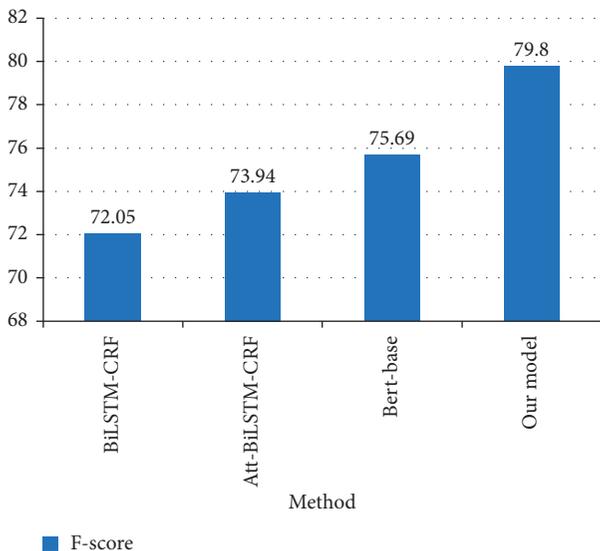


FIGURE 3: F-score of different methods.

MEMM model is no longer directly determining the level of system performance. Secondly, the entity relationship chain we proposed can efficiently tackle the problems as entity overlap and relationship intersection without the following two shortcomings. The first is error accumulation and entity redundancy caused by the mutual influence of entity recognition and relationship extraction which can lead to the computational complexity; the second is the lack of interaction information caused by ignoring the internal

TABLE 3: English entity relationship extraction results.

Methods	F-score
BRNN [28]	82.5
SDP-BLSTM [29]	83.7
CNN [30]	77.5
Att-RCNN [31]	83.7
Hybrid Bi-LSTM-Siamese [32]	81.8
Our method	84.60

connection and dependency between the entity recognition and relationship extraction.

Table 3 reveals the scalability of our proposed method, which can handle the English entity relationship extraction. Moreover, our method can reach an outstanding F-score of 84.6% which is overall higher than that of the other five methods. The results indicate that the proposed method not only performs well in dealing with Chinese entity relationship extraction but also has a superior scalability while dealing with English.

5. Summary and Future Work

In this paper, a joint extraction model based on joint coding is proposed, and Bi-MEMM is introduced into the joint extraction model and applied to entity relationship extraction tasks. Experiments show that the model performs well in Chinese data sets and has a strong scalability in English data sets. It owns the ability to learn the internal structure of a sentence without considering the complexity of named entities and relationships in the sentence. At the same time, we also notice that the model is still inadequate in dealing with the long-distance constraint of sample sentences, the implicit relation in entities, the reasoning of the same relation such as referential relation, subordination relation, and date writing format problem. Of course, annotation set data is also an important factor that cannot be ignored. We expect that future work could be carried out from the following aspects, such as integrating natural language algorithms (e.g., anaphora resolution into Deep Learning algorithms) and external knowledge bases (e.g., thesaurus, WordNet, HowNet, and knowledge map prior validation) waiting to be introduced into the model. We believe that the introduction of these methods in future modelling will greatly improve their accuracy.

Data Availability

The data used to support the findings of this study are available from the following website list: People’s Daily: <https://github.com/buppt/ChineseNER>; English SemEval 2010 Task 8: <https://www.kaggle.com/drtoshi/semEval2010-task-8-dataset>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This paper was partially supported by the National Natural Science Foundation of China (Nos. U1711266, 62076224, and 41925007), Department of Education Neo-Generation Information Technology Innovation Project (Nos. 2018A03006 and 2018A02021), and General Project of Education Humanities and Social Sciences in Hubei Province (No. 18Y38).

References

- [1] C. Aone and M. Ramos-Santacruz, "REES: a large-scale relation and event extraction system," in *Proceedings of the Sixth Applied Natural Language Processing Conference*, pp. 76–83, Seattle, WA, USA, April 2000.
- [2] J. S. Aitken, "Learning information extraction rules: an inductive logic programming approach," in *Proceedings of the ECAI 2002*, pp. 355–359, Lyon, France, July 2002.
- [3] A. Schutz and P. Buitelaar, "Relext: a tool for relation extraction from text in ontology extension," in *Proceedings of the International Semantic Web Conference 2005*, pp. 593–606, Galway, Ireland, November 2005.
- [4] B. Rink and S. Harabagiu, "A generative model for unsupervised discovery of relations and argument classes from clinical texts," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 519–528, Edinburgh, Scotland, July 2011.
- [5] S. Thattinaphanich and S. Prom-on, "Thai named entity recognition using Bi-LSTM-CRF with word and character representation," in *Proceedings of the 2019 4th International Conference on Information Technology (InCIT)*, pp. 149–154, Bangkok, Thailand, October 2019.
- [6] A. Culotta and J. Sorensen, "Dependency tree kernels for relation extraction," in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04) 2004*, pp. 423–429, Barcelona, Spain, July.
- [7] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," 2017, <https://arxiv.org/abs/1706.05075>.
- [8] B. Deng, X. Fan, and L. Yang, "Entity relation extraction method using semantic pattern," *Jisuanji Gongcheng/Computer Engineering*, vol. 33, no. 10, pp. 212–214, 2007.
- [9] R. McDonald, F. Pereira, S. Kulick, S. Winters, Y. Jin, and P. White, "Simple algorithms for complex relation extraction with applications to biomedical IE," in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05) 2005*, pp. 491–498, Ann Arbor, MI, USA, June 2005.
- [10] C. Wen, Z.-X. Shi, and Y. Xin, "Chinese non-taxonomic relation extraction based on extended association rule," *Computer Engineering*, vol. 24, 2009.
- [11] E. Riloff and R. Jones, "Learning dictionaries for information extraction by multi-level bootstrapping," in *Proceedings of the AAAI/IAAI 1999*, pp. 474–479, Orlando, FL, USA, July 1999.
- [12] M. Sabou, M. d'Aquin, and E. Motta, "SCARLET: semantic relation discovery by harvesting online ontologies," in *Proceedings of the European Semantic Web Conference 2008*, pp. 854–858, Karlsruhe, Germany, October 2008.
- [13] H. Yu, L. Qian, G. Zhou, and Q. Zhu, "Chinese semantic relation extraction based on unified syntactic and entity semantic tree," *Journal of Chinese Information Processing*, vol. 24, no. 5, pp. 17–23, 2010.
- [14] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, "Knowledge-based weak supervision for information extraction of overlapping relations," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies 2011*, pp. 541–550, Portland, ON, USA, June 2011.
- [15] Z. Qin and F. Ye, "Research on reliability of instance and pattern in semi-supervised entity relation extraction," in *Recent Developments in Intelligent Computing, Communication and Devices*, pp. 377–385, Springer, Berlin, Germany, November 2019.
- [16] D. Davidov, A. Rappoport, and M. Koppel, "Fully unsupervised discovery of concept-specific relationships by web mining," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics 2007*, pp. 232–239, Prague, Czech Republic, June 2007.
- [17] Y. Yan, N. Okazaki, Y. Matsuo, Z. Yang, and M. Ishizuka, "Unsupervised relation extraction by mining wikipedia texts using information from the web," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP 2009*, pp. 1021–1029, Singapore, August 2009.
- [18] T.-J. Fu, P.-H. Li, and W.-Y. Ma, "GraphRel: modeling text as relational graphs for joint entity and relation extraction," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019*, pp. 1409–1418, Italy; Venue, January 2019.
- [19] Y. Zhang, Z. Guo, and W. Lu, "Attention guided graph convolutional networks for relation extraction," 2019, <https://arxiv.org/abs/1906.07510>.
- [20] Z. He, Z. Zhou, L. Gan, J. Huang, and Y. Zeng, "Chinese entity attributes extraction based on bidirectional LSTM networks," *International Journal of Computational Science and Engineering*, vol. 18, no. 1, pp. 65–71, 2019.
- [21] F. Bai and A. Ritter, "Structured minimally supervised learning for neural relation extraction," 2019, <https://arxiv.org/abs/1904.00118>.
- [22] G. Ji, K. Liu, S. He, and J. Zhao, "Distant supervision for relation extraction with sentence-level attention and entity descriptions," in *Proceedings of the AAAI 2017*, Francisco, CA, USA, February 2017.
- [23] G. Bekoulis, J. Deleu, T. Demeester, and C. Develder, "Adversarial training for multi-context joint entity and relation extraction," 2018, <https://arxiv.org/abs/1808.06876>.
- [24] H. Xun, Y. Hongliang, and Y. Yang, "A review of relation extraction," *Data Analysis and Knowledge Discovery*, vol. 29, no. 11, pp. 30–39, 2013.
- [25] W. Zhang, Y. Feng, F. Meng, D. You, and Q. Liu, "Bridging the gap between training and inference for neural machine translation," 2019, <https://arxiv.org/abs/1906.02448>.
- [26] L. Luo, Z. Yang, P. Yang et al., "An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition," *Bioinformatics*, vol. 34, no. 8, pp. 1381–1388, 2018.
- [27] W. Huang, X. Cheng, T. Wang, and W. Chu, "BERT-based multi-head selection for joint entity-relation extraction," in *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing 2019*, pp. 713–723, Zhengzhou, China, October 2019.
- [28] D. Zhang and D. Wang, "Relation classification via recurrent neural network," 2015, <https://arxiv.org/abs/1508.01006>.
- [29] X. Yan, L. Mou, G. Li, Y. Chen, H. Peng, and Z. Jin, "Classifying relations via long short term memory networks along

- shortest dependency path,” 2015, <https://arxiv.org/abs/1508.03720>.
- [30] Y. Wang, *Entity Relationship Extraction Based on Deep Convolutional Neural Network*, Taiyuan University of Technology, Taiyuan, China, 2017.
 - [31] X. Zhang, F. Chen, and R. Huang, “A combination of RNN and CNN for attention-based relation classification,” *Procedia Computer Science*, vol. 131, pp. 911–917, 2018.
 - [32] Z. Cui, L. Pan, and S. Liu, “Hybrid bilstm-siamese network for relation extraction,” in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems 2019*, pp. 1907–1909, Montreal, Canada, May 2019.
 - [33] B. Srinivasa-Desikan, *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, spaCy, and Keras*, Packt Publishing Ltd, Birmingham, UK, 2018.
 - [34] B. Wang, R. Hu, and L. Yang, “Constructing the image graph of tang poetry,” in *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing 2019*, pp. 426–434, Dunhuang, China, October 2019.
 - [35] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, “The stanford CoreNLP natural language processing toolkit,” in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations 2014*, pp. 55–60, Baltimore, MD, USA, June 2014.