Joint Extraction of Entities and Relations in the News Domain

Zeyu Li[1(✉)], Haoyang Ma[2], Yingying Lv[1], and Hao Shen[1]

[1] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China
lizeyu@cuc.edu.cn

[2] North China Institute of Computing Technology, Beijing, China

Abstract. Extracting entities and relationships between entities from news text information is the core task of building news knowledge graphs. In recent years, with the rise of knowledge graphs, the joint extraction of entity relationships has become a research hotspot in the field of natural language processing. Aiming at the problem that there are many entities in news text data and overlapping relationships between entities are common, this paper first proposes a labeling strategy around the central entity, which transforms the extraction of entities and relationships into sequence labeling problems. After that, this paper also proposes a joint extraction model, which is based on pre-trained language and combined with the improved Bi-directional Long Short-Term Memory (BiLSTM) and Conditional Random Field (CRF) model to achieve entity and relationship extraction. The experimental results on two public news datasets show that our proposed joint extraction model has different degrees of improvement in accuracy and recall compared with other popular joint extraction models. The F1 value on NYT and DuIE both achieved the highest values, reaching 71.6% and 81.4%, which proves that the method proposed in this paper is effective.

Keywords: Joint extraction · Deep learning · Relation overlap

1 Introduction

In modern society, with the rapid development of network media and of information networks. With the exponential growth of data resources, how to extract valuable information from massive unstructured or semi-structured news data has become an important issue in the field of natural language processing. Therefore, the knowledge graph, which can be rich in objective world knowledge and store information in a structured form [1], has gradually become a research hotspot in the field of news. By building a knowledge map in the news field, the rapid development of intelligent information service technology in the news industry can be promoted, thereby bringing greater economic benefits and better news dissemination effects.

The triplet (h, r, t) composed of entities and the relationships between entities is the key semantic information in the knowledge graph, where h represents the head entity and t represents the tail entity, r represents the relationship that exists between entities.

W. Chen et al. (Eds.): ADMA 2022, LNAI 13725, pp. 81–92, 2022.
https://doi.org/10.1007/978-3-031-22064-7_7

How to extract entities and relationships from a large amount of news text data are the two core tasks of constructing and updating news knowledge graphs. There are usually a large number of entities in the text information in the news domain, and the problem of overlapping relationships is common. For example, in the short text "Alibaba company founded by Jack Ma is located in Hangzhou", there are two overlapping relation triples (Jack Ma, created, Alibaba) and (Alibaba, located in, Hangzhou). Therefore, if one wants to improve the accuracy of relation extraction in the field of news, in addition to improving the accuracy of named entity recognition, one must also be able to accurately identify overlapping relationships between entities. There are currently three joint extraction methods, which are method based on shared parameters, the method based on sequence annotation, and the method based on graph structure. These joint models fall into two paradigms. The first normal form can be expressed as $(h, t) \to r$, which first identifies all entities in a sentence, and then performs relation classification based on each extracted entity pair. However, these methods need to enumerate all possible entity pairs, and relation classification may suffer from redundant entities. Bekoulis et al. [4] proposed another paradigm, denoted as $h \to (r, t)$, which first detects head entities and then predicts the corresponding relations and tail entities. Compared with the first paradigm, the second paradigm can jointly identify entities and all possible relationships between them at one time, which can better solve the problem of overlapping entity relationships [5, 6]. Therefore, in order to effectively extract entities and overlapping relationships while reducing redundant information extraction, this paper proposes a central entity-oriented labeling strategy, which transforms the entity-relationship joint extraction task into a sequence labeling task to distinguish entities at different positions. Then, based on the theoretical basis of the second joint extraction paradigm, bidirectional long short-term memory (BiLSTM) and conditional random field (CRF), this paper proposes a RoBERTa-BiLSTM*-CRF entity-relationship joint extraction model, which is based on the RoBERTa pre-training model [7] and improved BiLSTM-CRF. The F1 values on the open-source datasets NYT and DuIE published in the news field reached 0.716 and 0.814. The experimental results confirmed the effectiveness and feasibility of the method in this paper.

2 Research Status

Since the concept of entity relation extraction task was proposed, after more than 20 years of continuous research, there have been relatively rich research results [2]. Traditional entity relationship extraction generally adopts the pipeline method, which divides named entity recognition and relationship extraction into two independent subtasks, and directly extracts the relationship between entities on the basis of entity recognition has been completed. Miwa used text features to construct feature vectors, and Kambhatla et al. [8] used the lexical and semantic features of text to extract relationships through the maximum entropy model. Because the construction of feature vectors requires researchers to have a large amount of linguistic knowledge, and features need to be manually designed, a lot of work is required to preprocess the data, which may also lead to the transmission of errors, so the accuracy of the early extraction model is not too high. Deep learning can automatically learn the features in the text and can have high accuracy, so it has gradually become the mainstream method for researching row entity and relationship extraction in recent years. At present, the relationship extraction methods based on deep learning can be roughly divided into pipeline relationship extraction methods and joint relation-ship extraction methods.

The pipeline relationship extraction method based on deep learning still de-composes the relationship extraction task into two independent subtasks, such as Nayak et al. [9] Perform classification and transform the relation extraction task into a classification prediction task [10]. Although the pipeline extraction method based on deep learning is easy to implement and has strong flexibility, the error in the process of entity extraction may be transmitted to the process of relation classification, resulting in the transmission of errors. Therefore, Miwa et al. [3] used the neural network model to solve the joint entity-relation extraction task for the first time, and integrated the two tasks into the same model through the method of sharing parameters, but the two tasks are still separate processes, resulting in a lot of redundant remaining information. In order to solve this problem, Zheng et al. [11] designed a novel labeling method, which extracts entities and relationships at the same time, transforms the extraction problem into a labeling task, and uses neural networks to model, avoiding the complexity of feature engineering. Wang et al. [12] proposed a new joint learning model based on the graph structure, which can enhance the correlation between related entities using a loss function of paranoid weight. However, these three joint extraction methods all face the problem of overlapping relationships in the extraction process. In order to effectively solve this problem, in recent years, several studies have tried to apply the Attention mechanism and pre-training models to joint extraction tasks. Liu et al. [13] proposed an attention-based joint relation extraction model, which designed a supervised multi-head self-attention mechanism as a relation detection module to separately learn the associations between each relation type to identify overlapping relations and relationship types. The joint extraction models proposed in the past two years basically introduce the Attention mechanism [14] and different pre-training models [15] to solve the problem of overlapping relationships in texts.

3 Methodology

This paper proposes a RoBERTa-BiLSTM*-CRF entity-relationship joint extraction model based on pre-trained embedding vectors, which map the sentences to the vector space and input the model, and which use an improved BiLSTM net-work to capture the semantic information in the sentence. Then combined with the CRF, for each input sentence, the predicted entity and related annotation sequence is obtained as the output of the model. The overall framework of the model is shown in Fig. 1.

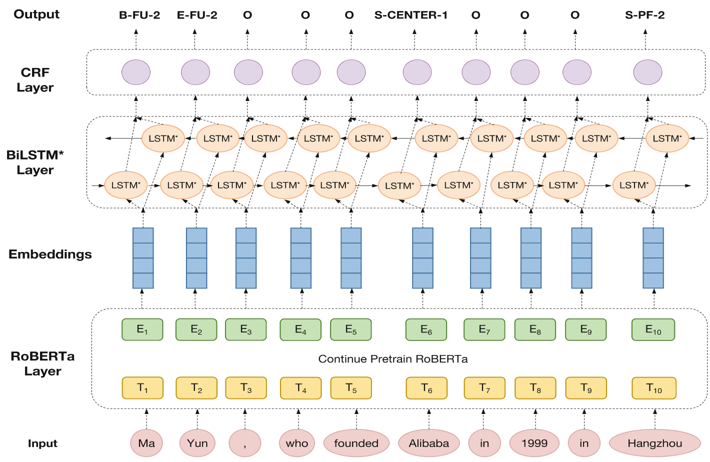

Fig. 1. Model framework

3.1 Labeling Strategy for Central Entities

The labeling strategy proposed in this paper is to divide labels into three parts: boundary information, relation classification, and entity order, as shown in Fig. 2. The boundary information part uses the 'BIESO' labeling method to represent the boundary information of a single word and character in the entity, B represents the first word in the entity head, and I represent the word in the middle of the entity, and E represents the last word of the entity, S represents an entity with only one word or character, and O represents other words that are not entities. The relation classification part is composed of pre-defined relation type labels. Furthermore, we specifically propose a new relation class "CENTER", which is the highest-level relation class to label the narrative body, the central entity, in news texts. The entity order is the serial number marked by a number to distinguish the internal head and tail entities of the triplet. The serial number of the central entity as the head entity of the triplet is fixed as '1', and the other entities are the tail entities in the triplet. The serial number is '2'.

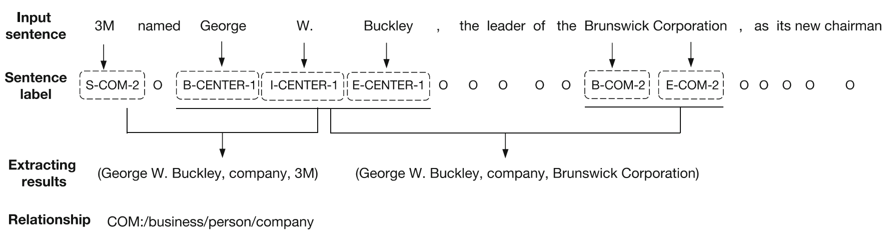

Fig. 2. An example of text labeling

First necessary to obtain entities according to entity boundaries and to determine entity relationships and all possible combinations according to the nearest neighbor principle. When the entity-relationship category is not "CENTER", find the entities with the closest distance, the same relationship category, and different entity sequence numbers in the text before and after the entity to form entity-relationship triples. When the entity-relationship category is "CENTER", the entity-relationship triplet is formed by finding entities with different entity positions in the two text directions before and after and matching them.

3.2 RoBERTa Presentation Layer

In view of the fact that the distance between entity pairs with existing relationships is relatively far in news text data [19], in order to facilitate the construction of the input of the RoBERTa representation layer, we define each input news text sentence as a word sequence d of length m. So $d = (w_1, w_2, w_3, \ldots, w_m)$, the RoBERTa model maps each word $w_i(1 \le i \le m)$ in d to a word vector of dimension n, and the transformed vector represents the sequence V such as Formula 1.

$$V = (v_1, v_2, v_3, \ldots, v_m), v_i \in \mathbb{R}^n (1 \le i \le m) \tag{1}$$

3.3 Improved BiLSTM* Layer

Due to its design characteristics, the LSTM is very suitable for processing text data. However, due to the tanh function used by the LSTM neural network, the gradient disappears during the training process of the network. Therefore, this paper proposes a new activation function RTLU, which is used to replace the tanh function inside the LSTM neuron. The function expression is shown in Formula 2.

$$f(x) = \begin{cases} x, x \ge 0 \\ tanhx, x < 0 \end{cases} \tag{2}$$

Specifically for the neuron x, since the ReLU function does not have the problem of neuron saturation when x is in the positive interval, it can easily solve the gradient disappearance problem caused by the soft saturation of the tanh activation function when x is in the positive interval. When x is in the negative range during the backpropagation process, the tanh function will not cause the problem that the weight is not updated due to the gradient being 0, which can well solve the neuron death problem caused by the ReLU function not updating the weight.

In order to improve the performance of the model, this paper replaces the tanh function of the LSTM neuron with the RTLU activation function. The internal structure of the improved LSTM neuron LSTM* is shown in Fig. 3.

Output of the hidden layer is h_{t-1}. Then the control formula of the forgetting gate at time t is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, V_t] + b_f) \tag{3}$$

where W_f is the weight matrix of the forget gate, b_f is the bias term of the forget gate, $[h_{t-1}, V_t]$ represents the splicing of two vectors, σ represents the activation function, f_t represents how many times t-1 should be retained unit status.

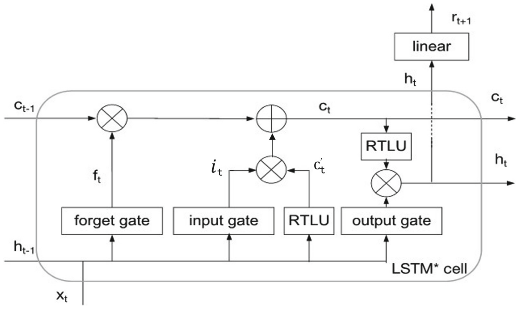

Fig. 3. Internal Structure of the improved LSTM*

Similarly, the input gate control formula at time t is:

$$i_t = \sigma(W_i \cdot [h_{t-1}, V_t] + b_i) \tag{4}$$

where W_i is the weight matrix of the input gate, and b_i is the input gate bias term. The input of the unit state at time t is jointly determined by the output at time t-1 and the input at time t. The formula is as follows:

$$c_t^{'} = \mathrm{RTLU}(W_c \cdot [h_{t-1}, V_t] + b_c) \tag{5}$$

where W_c is the weight matrix, b_c is the bias term. The unit state c_t at time t is determined by the unit state c_{t-1} at time t-1 and the calculation results of formulas (3), (4), (5), and the formula is as follows:

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t^{'} \tag{6}$$

where $\odot$ represents the multiplication of the corresponding position elements in the matrix. Then the control formula of the "output gate" at time t is:

$$o_t = \sigma(W_o \cdot [h_{t-1}, V_t] + b_o) \tag{7}$$

where W_o is the weight matrix, b_o is the bias term. The final output of the forward LSTM* at time t is obtained by multiplying the unit state and the judgment condition obtained by the output gate. The formula is as follows:

$$\overrightarrow{h_t} = o_t \cdot \mathrm{RTLU}(c_t) \tag{8}$$

The hidden layer of BiLSTM* also saves the forward output vector $\overrightarrow{h}$ and the reverse output vector $\overleftarrow{h}$. .The output of the improved BiLSTM* encoding layer at time t is:

$$\mathrm{H} = \overrightarrow{h_t} + \overleftarrow{h_t} \tag{9}$$

Table 1. Dataset size.

Dataset	NYT	DuIE
Total number of sentences	61385	70128
Total number of relationships	68312	76254
Number of relationship types	24	27
Train set sentences	42970	49089
Test set sentences	12277	14026
Validation set sentences	6138	7013

Table 2. Experiment environment.

Hardware equipment	CPU: Intel I7–9700 3.00 GHz
	RAM: 32.0 GB
Software	Windows 10 64bit Python 3.6.0
	tensorflow 1.14.0 keras 2.4.3 gensim 3.4.3 jieba 0.42.1

4 Experiment

4.1 Experimental Data and Experimental Environment

To verify the effectiveness of the model proposed in this paper for entity relation extraction in the news domain, it is tested on public datasets NYT [20] and DuIE [21]. These two open-source datasets contain multiple relational triples, making them ideal for evaluating models for extracting overlapping relational triples. This paper filters out overly complex long sentences with more than 100 words in the two datasets and some sentences that are not closely related to the news domain, and randomly divide them into the training set, test set, and validation set according to the ratio of 7:2:1. The specific information of the two datasets is shown in Table 1. The experimental environment of this experiment is shown in Table 2.

4.2 Evaluation Standard

In the entity-relationship joint extraction experiment, the evaluation criteria used are international standards, including the precision (P), recall (R) and F1 value. The extracted relation is considered correct when both the category and head and tail entities of the relation are correct. The formula parameters are defined as follows.

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

where TP is the number of correct identifications, FP is the number of irrelevant identified objects, and FN is the number of unidentified objects that exist in the dataset.

It is usually necessary to comprehensively consider the harmonic mean of precision and recall, that is F1 value, which is defined as follows.

$$F_1 = \frac{2 * P * R}{P + R} \tag{12}$$

4.3 Experimental Parameters

In the model training process, 20% of the data is used as the validation set to adjust the parameters during training. After several experiments and fine-tuning, the experimental parameters set for the NYT and DuIE datasets are slightly different.

During the training process, the batch_size is 64, the number of LSTM units is 256, and the maximum text length is 128. The RoBERTa model and the BERT model used in pre-training are both Google's open-source pre-training models, both of which are 12-layer and hidden. The layer is 768-dimensional and adopts the 12-head mode, while the Glove model is trained on the open source news corpus of Sogou Lab. Its feature vector is 400-dimensional, and the remaining parameters are the default values in gensim. In addition, the Adam optimizer [22] with initial learning rates of 1e-3 and 1e-5 is used to learn 100 epochs on the training sets of NYT and DuIE, respectively, with dropout sizes of 0.5 and 0.2 to speed up training and prevent overfitting. Obtain the best F1 value model on the validation set.

4.4 Experimental Design

In order to prove the effectiveness of the model in this paper, we have compared it with the joint relation extraction models in recent years. The baseline models used in this paper for comparison are as follows:

GraphRel: The model proposed by Fu et al. [23] is a joint extraction model based on graph structure. The model divides the overall joint entity relationship extraction into two stages. Both stages use a bidirectional graph convolutional neural network for feature extraction and prediction.

Glove-BiLSTM-CRF: Hu et al. [6] proposed a joint learning model that can identify overlapping relationships between entities, using a parameter sharing method to achieve joint extraction tasks, by allowing the two tasks to share except the last layer of relationship classifiers The neural network parameters of all layers use the connection between tasks to optimize the effect of the two tasks.

TETI: Based on the encoder-decoder structure, Chen et al. [24] fused entity category information to construct the entity-relationship joint extraction model FETI. The prediction of the head and tail entity categories is added in the decoding stage and constrained by an auxiliary loss function so that the model can use the entity category information more effectively.

RoBERTa-BiLSTM-CRF: Li et al. [18] proposed this new model for the evaluation object extraction task, and this paper uses this model for the joint extraction of entity relations, aiming to verify the effectiveness of our model improvement.

CASREL: Wei et al. [16] proposed a joint extraction method of sequence labeling based on the BERT [17] pre-training model, and implemented a Cascade Binary Tagging Framework that is not troubled by the overlapping triple problem. The relation in triple is modeled as a function that maps a head entity to a tail entity, rather than treating it as a label on an entity pair.

4.5 Result Analysis

In order to verify whether the RoBERTa-BiLSTM*-CRF model proposed in this paper can effectively extract overlapping relationships in news texts, experiments were conducted on the open-source English NYT news dataset and Chinese DuIE dataset. The results are shown in Table 3.

It can be seen that on the two news datasets, the RoBERTa-BiLSTM*-CRF model proposed in this paper has achieved the highest F1 value for relation ex-traction, indicating that our model is superior at the joint entity relation extraction task of news texts. From the comparison of indicators, the precision rate of each model is generally higher than the recall rate. The extraction effect of sever-al joint extraction models based on character vectors is significantly better than the Glove-BiLSTM-CRF model based on word vectors and the GraphRel model based on bi-RNN and GCN. Specifically, the RoBERTa-BiLSTM-CRF model based on character vectors and the model proposed in this paper has improved F1 values on the NYT dataset by 15.7% and 18.2%, respectively, compared with the Glove-BiLSTM-CRF model based on Glove word vectors. The F1 value on the DuIE dataset has increased by 15.9% and 18.1% respectively, which also shows that adding the RoBERTa pre-trained language model can obtain dynamic word vectors according to the context information of words, and use a self-attention mechanism to obtain bidirectional semantic features, which greatly improves the effectiveness of entity relation extraction.

The proposed model in this paper improves the F1 value of the RoBERTa-BiLSTM-CRF model by 2.14% and 1.88% on the NYT and DuIE datasets, respectively, indicating that the comprehensive performance of the BiLSTM*-CRF model is due to the BiLSTM-CRF model, indicating that the improved LSTM neural Meta can effectively improve the effect of entity-relationship joint extraction task. Both CASREL and our model based on the BERT training model are joint extraction methods based on sequence annotation, but CASREL is not as good as our model, but both models are better than the RoBERTa-BiLSTM-CRF model. It shows that the RTLU activation function proposed in this paper has a positive effect on entity relation extraction.

In addition, this paper also explores the different effects of different parameters on the joint extraction performance of our model. First, set the batch_size to 32 and the number of LSTM units to 256, and test the impact of dropout changes on the F1 value of the comprehensive evaluation index of the model, as shown in Table 4. As the value of Dropout increases, the F1 value of the model will first increase to the highest value, and then as the Dropout value continues to increase, the value of F1 will decrease instead.

Afterward, the dropout values on the NYT dataset and the DuIE dataset are set to 0.5 and 0.2 respectively, and the batch_size is uniformly set to 32 to test the impact of the change in the number of LSTM units on the comprehensive evaluation index F1

Table 3. Comparison with baseline model.

Models	NYT			DuIE		
	P	R	F1	P	R	F1
GraphRel	0.632	0.597	0.614	0.705	0.683	0.694
Glove-BiLSTM-CRF	0.611	0.602	0.606	0.692	0.687	0.689
TETI	0.686	0.653	0.669	0.758	0.754	0.756
CASREL	0.715	0.695	0.705	0.806	0.798	0.802
RoBERTa-BiLSTM-CRF	0.713	0.689	0.701	0.792	0.786	0.789
RoBERTa-BiLSTM*-CRF	**0.724**	**0.709**	**0.716**	**0.823**	**0.805**	**0.814**

Table 4. Effect of dropout on RoBERTa-BiLSTM*-CRF model.

Dropout	NYT			DuIE		
	P	R	F1	P	R	F1
0.1	0.715	0.699	0.707	0.819	0.803	0.811
0.2	0.716	0.701	0.708	**0.823**	**0.805**	**0.814**
0.3	0.719	0.704	0.711	0.821	0.802	0.811
0.4	0.722	0.706	0.714	0.817	0.799	0.808
0.5	**0.724**	**0.709**	**0.716**	0.812	0.791	0.801
0.6	0.720	0.707	0.713	0.811	0.785	0.798

value of the model, as shown in Table 5. It can be seen that in the RoBERTa-BiLSTM*-CRF model, increasing the number of LSTM units in the model will improve the model performance to a certain extent, but as the number of LSTM units continues to increase, the model will have problems with overfitting and increase the cost of model training. The overfitting effect of deep neural networks can be reduced by adding dropout, but the larger the dropout, the more information is discarded, and the model performance will gradually decrease. Therefore, selecting the appropriate number of LSTM units and dropout values can effectively improve model performance and alleviate the over-fitting problem caused by too few training samples. It can be seen from Table 4 andTable **5** that on the NYT dataset and DuIE dataset, when the dropout of the joint.

entity-relation extraction task is 0.5 and 0.2, the F1 value reaches the highest. When the number of LSTM units is 256, the F1 value reaches the highest, and the entire model is optimal at this time. To sum up, the RoBERTa-BiLSTM*-CRF model proposed in this paper has the highest F1 value for the joint entity-relation extraction task on the English and Chinese news domain datasets compared with other methods. Compared with other joint extraction methods based on sequence annotation, the F1 value of the model in this paper can be improved by 1.5%3.2%, which shows the superiority of the LSTM improvement strategy in this paper.

Table 5. Effect of the number of LSTM units on model.

LSTM units	NYT			DuIE		
	P	R	F1	P	R	F1
64	0.718	0.702	0.710	0.816	0.795	0.805
128	0.721	0.708	0.714	0.822	0.803	0.812
256	**0.724**	**0.709**	**0.716**	**0.823**	**0.805**	**0.814**
512	0.720	0.704	0.712	0.819	0.800	0.809

5 Conclusion

In this paper, we propose an entity-relationship joint extraction method for news domain text information and a central entity-oriented labeling strategy, which transforms the entity-relationship joint extraction task into a sequence of labeling tasks. The experimental results show that the model proposed in this paper can effectively extract the overlapping relationship between entities and entities in the text data in the news domain with the help of the central entity-oriented annotation strategy. However, in the process of extraction, it is also a problem to be solved that the way of entity semantic distinction is too simple. Therefore, future work considers improving the accuracy of multiple relation extraction between the same entity pair by further improving the central entity-oriented annotation strategy.

References

1. Pujara, J., Miao, H., Getoor, L., Cohen, W.: Knowledge graph identification. In: Alani, H., et al. (eds.) ISWC 2013. LNCS, vol. 8218, pp. 542–557. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41335-3_34
2. Li, D., Zhang, Y., Li, D., et al.: Review of entity relation extraction methods. J. Computer Res. Dev. **57**(7), 1424–1448 (2020)
3. Miwa, M., Bansal, M.: End-to-end relation extraction using LSTMs on sequences and tree structures. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 1105–1116. Association for Computational Linguistics, Strasbourg (2016)
4. Bekoulis, G., Deleu, J., Demeester, T., et al.: Joint entity recognition and relation extraction as a multi-head selection problem. Expert Syst. Appl. **114**, 34–45 (2018)
5. Zhao, T., Yan, Z., Cao, Y., Li, Z.: Entity relative position representation based multi-head selection for joint entity and relation extraction. In: Sun, M., Li, S., Zhang, Y., Liu, Y., He, S., Rao, G. (eds.) CCL 2020. LNCS (LNAI), vol. 12522, pp. 184–198. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-63031-7_14
6. Hu, Y., Yan, H., Chen, C.: Joint entity and relation extraction for constructing financial knowledge graph. J. Chongqing University Technol. (Natural Science) **34**(5), 139–149 (2020)
7. Liu, Y., Ott, M., Goyal, N., et al.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
8. Kambhatla, N.: Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions, pp. 178–181. Association for Computational Linguistics, Strasbourg (2004)

9. Nayak, T., Ng, H.T.: Effective attention modeling for neural relation extraction. In: Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), pp. 603–612. Association for Computational Linguistics, Strasbourg (2019)
10. Zhang, D., Peng, D.: ENT-BERT: entity relation classification model combining bert and entity information. J. Chinese Computer Syst. **41**(12), 2557–2562 (2020)
11. Zheng, S., Wang, F., Bao, H., et al.: 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, pp. 1227–1236. Association for Computational Linguistics, Strasbourg (2017)
12. Wang, S., Yue, Z., Che, W., et al.: Joint extraction of entities and relations based on a novel graph scheme. In: Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 4461–4467. AAAI Press, Menlo Park (2018)
13. Liu, J., Chen, S., Wang, B., et al.: Attention as relation: learning supervised multi-head self-attention for relation extraction. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, pp. 3787–3793. Springer (2021). https://doi.org/10.24963/ijcai.2020/524
14. Lai, T., Cheng, L., Wang, D., et al.: RMAN: Relational multi-head attention neural network for joint extraction of entities and relations. Applied Intelligence **52**(3), 3132–3142 (2021)
15. Qiao, B., Zou, Z., Huang, Y., et al.: A joint model for entity and relation extraction based on BERT. Neural Computing Appl. **34**(5), 3471–3481 (2022)
16. Wei, Z., Su, J., Wang, Y., et al.: A novel cascade binary tagging framework for relational triple extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 1476–1488. Association for Computational Linguistics, Strasbourg (2020)
17. Devlin, J., Chang, M. W., Lee, K., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
18. Li, Z., Yu, T., Shen, H.: Research on Opinion Targets Extraction of Travel Reviews Based on RoBERTa Em-bedded BILSTM-CRF Model. In: 2021 International Conference on Culture-oriented Science & Technology (ICCST), pp. 114–118. IEEE, Piscataway (2021)
19. Fu, R., Li, J., Wang, J., et al.: Joint extraction of entities and relations for domain knowledge graph. J. East China Norm. Univ. Nat. Sci. **2021**(5), 24–36 (2021)
20. Riedel, S., Yao, L., McCallum, A.: Modeling relations and their mentions without labeled text. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010, LNCS, vol. 6323, pp. 148–163. Springer, Heidelberg (2010)
21. Li, S., et al.: DuIE: a large-scale chinese dataset for information extraction. In: Tang, J., Kan, M.-Y., Zhao, D., Li, S., Zan, H. (eds.) NLPCC 2019. LNCS (LNAI), vol. 11839, pp. 791–800. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32236-6_72
22. Kingma, D., Ba, J.: Adam: A Method for Stochastic Optimization. Computer Science (2014)
23. Fu, T., Li, P., Ma, W.: 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, pp. 1409–1418. Association for Computational Linguistics, Strasbourg (2019)
24. Chen, R., Zheng, X., Zhu, Y.: Joint entity and relation extraction via fusing entity type information. Comput. Eng. **48**(3), 46–53 (2022)