ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0002

An Improved Entity-Relationship Joint Extraction Model based on Cascade Pointer Network

Danyang Qiao[1, 2], Bin Song[1, 2], Xu Yang[3], Kefeng Fan[3], Zhiyong Zhang[1, 2, *]

[1] School of Information Engineering College, Henan University of Science and Technology, Luoyang, Henan 471023, China

[2] Henan International Joint Laboratory of Cyberspace Security Applications, Henan University of Science and Technology, Luoyang, Henan 471023, China

[3] China Electronics Standardization Institute, Beijing 100007, China

Abstract

Relationship Extraction, as a key part of building knowledge graphs, aims to extract entity-relationship triad information from unstructured text. However, there still exist some issues that occur in the relationship extraction task, for instance, multiple entity miscommunication for the same sentence and the difficulty of overlapping triples extraction. This paper proposes a joint entity relationship extraction model based on an improved cascaded pointer network. First, the constructed Duie_Bert pre-training model is employed to encode the input text. Then, a multi-headed attention mechanism guided by specific relation-entity vectors is used to enhance the feature representation of the output vectors in the coding layer. Finally, the corresponding tail entities are extracted for the head entities and each relation by using an improved cascaded pointer annotation framework, so as to complete the extraction of the relation triples. The performance of the proposed model is compared with the latest three entity-relationship joint extraction models on the Duie Chinese relational dataset. The experimental results show that the proposed model achieves a performance improvement of 9.0% and 5.2% in Recall and F1 score, respectively. The result validate the effectiveness of the proposed method.

Keywords

Relation Extraction; Deep Learning; Knowledge Graph; Natural Language Processing; Bert.

1. Introduction

Relationship Extraction, as a crucial step in building knowledge graphs, has become a focus for researchers since the last decade [1]. Nowadays there is a new challenge to extract entity-relationship triads from natural language texts quickly and efficiently in the context of the big data era, where data objects and interactions are growing geometrically [2]. However, due to the diversity of unstructured textual information representation, it is difficult and challenging to extract relationships from natural language texts. The classical approach focuses on feature engineering, but all features are built on symbolic representations, which suffer from problems such as multiple meanings and ambiguities [3]. With the development of deep learning, deep neural networks have shown significant advantages in many research areas. To address the problems of feature engineering, Zeng et al. used with CNN for the first time in word-level and sentence-level features, which significantly improved the performance of the relationship extraction model [4]. Gao et al. improved on CNN and kernel functions to obtain a multi-entity Chinese relationship extraction model, which achieved good results in the document-

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0002

level relationship extraction task [5]. In contrast to feature engineering, deep neural networks use distributed representations rather than symbols, greatly improving the problem of multiple meanings and ambiguities. At the same time, deep learning-based models can automatically learn feature representations, with considerable success in entity relationship extraction tasks [6]. However, most existing approaches cannot effectively handle the case of sentences containing multiple overlapping relational triples, leading to problems of miscommunication and data redundancy.

In this paper, the Duie_Bert pre-training mode is introduced into the research of entity association relationship extraction, and an entity association extraction model based on the improved cascade pointer network is proposed. In order to solve the problems of error accumulation and data redundancy in the process of entity relation extraction, we introduce a multi-headed attention mechanism. The process is guided by specific relation-entity vectors, which can enhance the feature representation of the output vector of the encoding layer.

The main contributions of this paper are listed below:

(1) This paper proposes the Duie_Bert pre-training model for text encoding.

(2) This paper introduces a multi-headed attention mechanism guided by specific relation-entity vectors in the model. The process makes it possible to obtain semantic vectors that precisely depict the meaning of entities. To some extent, this method improves the precision of relation extraction.

(3) This paper employs cascading pointer networks for joint decoding of entity relations on the basis of pre-trained models, which can effectively extract entity relation triads in sentences and solve the error loss problem caused by overlapping triads.

2. Relates Work

The essence of the relation extraction task is to identify potential entity-relationship triples in text. Li et al. proposed that an entity triple usually consists of a pair of entities and semantic relations between them [7]. As a core task and an important link in the fields of information extraction, knowledge graph construction, natural language understanding, and information retrieval, entity relationship extraction can extract semantic relationships between entity pairs from text.

Most of the early entity relationship extraction models are based on feature engineering and traditional statistical learning methods [3]. However, traditional feature-engineered entity relationship extraction models cannot be separated from the use of manual and natural language processing tools, which greatly reduce the efficiency of relationship extraction. To solve this problem, a number of deep neural network-based models have gradually become the mainstream research direction. In recent years, with the rise of deep learning, researchers have gradually applied deep learning to the task of entity relationship extraction [8]. Among them, the supervised entity relationship extraction method based on deep learning is a hot research topic in recent years, which can reduce the error accumulation problem in the feature extraction process. At the same time, the deep learning neural network model can automatically learn sentence features without the need for complex feature engineering [9-13].

Depending on the order of completion of two relation extraction subtasks, entity relationship extraction methods includes traditional pipeline methods and joint extraction methods. These two methods are based on the three frameworks of RNN, CNN, and LSTM for extended optimization [4,14,15]. The pipeline method divides the task of relation extraction into two sub-tasks, named entity recognition and relation classification, which are carried out sequentially in a pipelined manner, and finally the triples with entity relations are output as prediction results. Among them, named entity recognition refers to the recognition of entities with specific meaning in text, mainly including names of people, places, institutions, proper nouns, etc. Relation extraction is to explore the relationship between related entities in the sentence. However, this type of method is simple to cause errors to accumulate and spread between the two subtasks and affect the accuracy of extraction. Most researchers have focused attention on the joint extraction model's research in recent years, because

ISSN: 2414-1895
DOI: 10.6919/ICJE.202303_9(3).0002

the method can lessen the influence of error propagation in the pipeline technique. For example, Zheng et al. treated relationship extraction as named entity recognition, entity class labels are changed to relationship class labels, and relationship extraction is performed in a sequential annotation. Based on the end2end model of sequence-to-sequence learning with replication mechanism for joint entity-relationship extraction, Zeng et al. introduced three patterns of overlapping triples. These can solve multiple entity-relationship overlap problems by the sequence-to-sequence model with replication mechanism. However, the model has the problem of unidirectional dependence of entity-relationship triples generated by backward and forward sequences in the process of decoding. Fu et al. proposed to regard the original sequence of sentences as a graph, each word in the sentence as a node, and perform feature fusion between each word through a two-stage graph convolutional network. This model can avoid the problem of entity-relationship triples dependent on each other due to the sequence in the decoding process. However, it cannot solve the overlapping relationship of EPO type. Wei et al. proposed a new cascaded binary annotation framework that converts the task of triad extraction into a problem with three levels of head entities, relations and tail entities, effectively solving the problem of overlapping relations of EOP type. However, for the extraction of complex entity-relationship triads, the fine-grained semantic links between the subject and individual words in the sentence are ignored, which reduces the significant effect of long textual relation extraction to some extent [16-20].

In summary, the past decades have witnessed the aboved mentioned remarkable works on entity-relationship joint extraction methods, but there still exist some shortcomings. For instance, the final-grained semantic links between entities and individual words in a sentence are not fully utilized in the encoding process, leading to the miscommunication of semantic information. In order to solve the problems of error accumulation and data redundancy in the process of relation extraction, this paper introduces a multi-headed attention mechanism guided by a specific relation-entity vector in the joint entity-relationship extraction model.

3. Entity Relationship Joint Extraction Model

The entity relationship joint extraction model proposed in this paper is divided into three main parts, that is Duie_Bert encoding layer, head entity identification layer, and relationship-tail entity joint extraction layer, where a multi-headed attention mechanism is introduced in the relationship-tail entity joint extraction layer. The Diagram of the joint extraction model of entity relations is as follows.

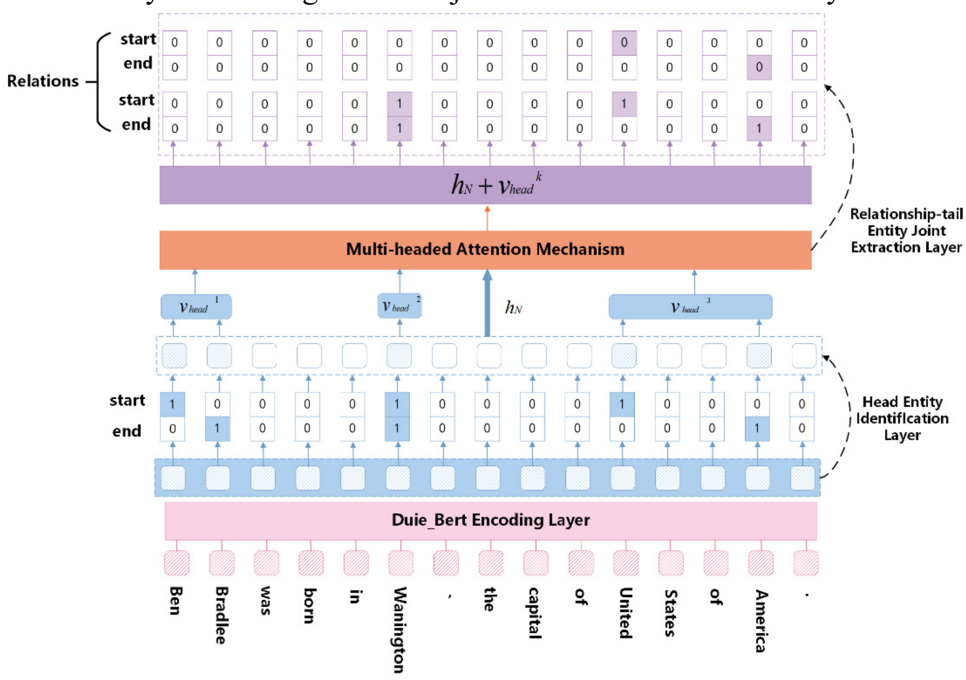

Figure 1. Diagram of the joint extraction model of entity relations

3.1 Duie_Bert Coding Layer

Bert is an efficient pre-trained language model proposed by Devlin et al., which adopts the Transformer encoder structure as the feature extractor and uses the accompanying MLM training method to achieve bidirectional encoding of input sequence text with strong semantic information extraction capability [26]. In order to construct the semantics of input sentences more accurately, the Bert model is trained again using the text in the Duie dataset to obtain the Duie_Bert model. The Bert-based pre-training model is shown as follows.

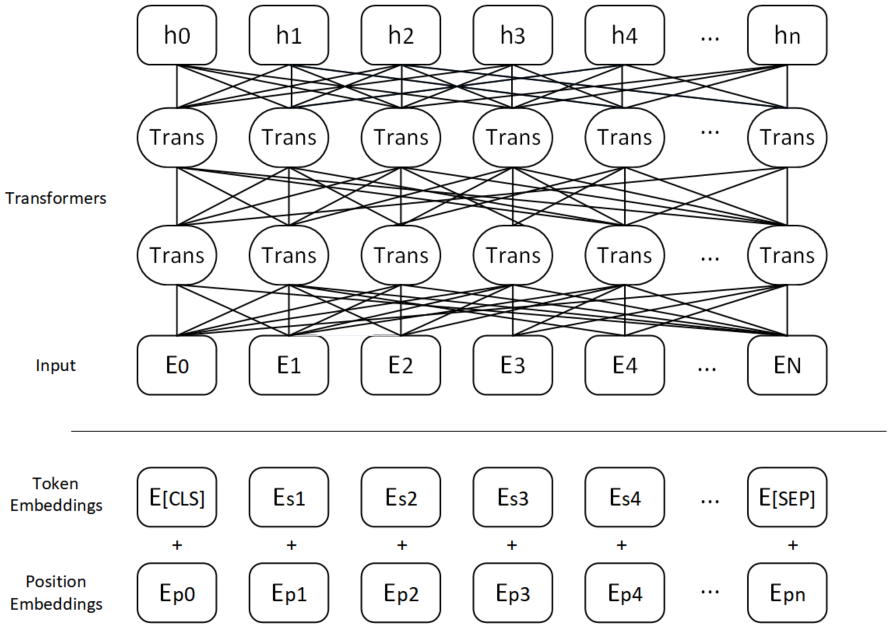

Figure 2. Bert-based pre-training model

The Bert model input contains three parts: word embedding, fragment embedding and location-encoded embedding. As fragment embedding is not applicable in the relational extraction task, the Duie_Bert model discards this part of embedding information and replaces word embedding with word embedding. The word embedding information W_S and location embedding information W_p are summed to obtain the input vector, which is passed through the Transformer network in the first and subsequent layers to obtain the vector representation of the text.

$$h_0 = SW_S + W_p \tag{1}$$

$$h_n = Transformer(h_{n-1}), \quad n \in [1, N] \tag{2}$$

Where S is a single thermal vector matrix of three indices of subwords in the input sentence, P denotes the position index in the input sequence. h_n is a hidden state vector, representing the output of the sentence after it has been encoded by the N-layer Transformer network, which is used as the input to the decoding layer.

3.2 Header Entity Marker Layer

The header entity tagging layer decodes the result of the encoding layer directly. First, for each word embedded token on the input, a linear layer and a $sigmoid$ activation function are used to determine whether it is the start or end part of the header entity. Then, the annotation framework is combined to identify all possible head entities, and for sentences containing multiple head entities, the principle of proximity matching is used to ensure the integrity of the entity span, which is calculated as follows.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0002

$$f_i^{start\_s} = \sigma(W_{start}\,\mathrm{t}_i + b_{start}) \tag{3}$$

$$f_i^{end\_s} = \sigma(W_{end}\,\mathrm{t}_i + b_{end}) \tag{4}$$

Where $f_i^{start\_s}$ and $f_i^{end\_s}$ denote the probability of identifying the ith token in the input sequence as the start position and end position of the topic, respectively. If the probability exceeds a set threshold, token 1 is assigned to the corresponding token, otherwise token 0. t_i is the encoded representation of the ith token in the input sequence, $\mathrm{t}_i = h_n[i]$, where $W_{(.)}$ is the trainable weight, $\mathrm{b}_{(.)}$ is the bias, and σ is the $sigmoid$ activation function.

3.3 Relationship-tail Entity Marker Layer

Unlike traditional recognition methods, the relation-tail entity marking layer first identifies all possible head entities. Then, given a class of relations, goes on to identify the tail entities associated with the head entities. The structure of each tail-entity recognition layer is actually the same as the head-entity recognition layer, differing mainly in the input, which is calculated as follows.

$$f_i^{start\_o} = \sigma(W_{start}^{r}(\mathrm{t}_i + sum(v_{head}^{k})) + b_{start}^{r}) \tag{5}$$

$$f_i^{end\_o} = \sigma(W_{end}^{r}(\mathrm{t}_i + sum(v_{head}^{k})) + b_{end}^{r}) \tag{6}$$

where $f_i^{start\_o}$ and $f_i^{end\_o}$ denote the probability of identifying the ith token in the input sequence as the start and end position of the tail entity under the relationship, respectively. v_{head}^{k} denote the vector of encoded representations of the kth head entity detected in the low-level module, which usually consists of multiple tokens. To maintain consistency t_i with v_{head}^{k} the dimensionality of the two vectors, the average vector representation of all vectors between the start and end tokens of the kth head entity is taken in this paper as $\mathrm{Sum}(v_{head}^{k})$.

3.4 Relationship-specific-entity-directed Attention Mechanisms

The attention mechanism, first proposed by Vaswani et al., is able to selectively filter a small amount of important information from a large amount of textual information and focus on this important information [21].

The focusing process is reflected in the calculation of the weight coefficients, where the correlation between characters in a sentence is obtained by calculating the weight of each Value in the input sentence, then adjusting the weight coefficient matrix to obtain a vector representation of individual Values.

In this paper, a multi-headed attention mechanism is introduced to calculate the attention weights between each specific relationship-entity vector. The process is utilized to obtain a semantic vector that can accurately express the meaning of the entity, thus improving the accuracy of relationship extraction. The specific calculation process is as follows.

$$e_{object^{r}} = \frac{\left(W_{object^{r}} v_{object^{r}} + b_{object^{r}}\right)\left(W_{object^{r}} v_{object^{r}} + b_{object^{r}}\right)^{T}}{\sqrt{d_k}} \tag{7}$$

$$\alpha_i = \frac{\exp\left(e_{object^{r}}\right)}{\sum_{j=1}^{n} \exp\left(e_j\right)} \tag{8}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202303_9(3).0002

First, a scaled dot product is used to calculate the correlation fraction $e_{object}{}^{r}$ between a particular relationship-entity vector. Next, normalisation was carried out to obtain the attention weight α_i . Finally, the attention weight α_i is weighted and summed with the output vector of the encoding layer, and the resulting global vector is passed through the activation function $sigmoid$ to obtain the attention weight output.

Where $v_{object}{}^{r}$ denotes the feature vector of a particular relation-entity, $W_{object}{}^{r}$ denotes the trainable weight matrix, d_k denotes the dimensionality of the input vector, $e_{object}{}^{r}$ denotes the correlation score between the vectors of a particular relation-entity, and α_i denotes the attention weight.

3.5 Model Algorithm Training Process

Table 1. Algorithm 1: Entity relationship joint extraction model algorithm training process

input: Training sentence set $W=\{w_1,w_2,...,w_n\}$, relation set $R=\{r_1,r_2,...,r_n\}$,the pretrained Duie_Bert parameters.

output: Header entity embeddings, relation embeddings, tail entity embeddings.

Initialize position embeddings and learnable parameters

$h_i{}^{r} \leftarrow r_i$

for epoch $n=1$ to N **do**

sample a training batch W_{batch}

initialize the training loss function $L_{start\_s}, L_{end\_s}, L_{start\_o}, L_{end\_o}$.

foreach $w_i \in W_{batch}$ **do**

Obtain word embeddings t_i

$\{f_1{}^{start\_s}, f_2{}^{start\_s},..., f_{n_s}{}^{start\_s}\}$ //Get start positions of the head entity

$\{f_1{}^{end\_s}, f_2{}^{end\_s},..., f_{n_s}{}^{end\_s}\}$ //Get end positions of the head entity

$Q_s=\{h_1{}^{sub}, h_2{}^{sub},..., h_{n_s}{}^{sub}\}$ //Get subject set in t_i

Update object extraction loss L_{start_s}, L_{end_s} .

foreach $h_j{}^{sub} \in Q_s$ **do**

foreach $h_i{}^{r} \in R$ **do**

Obtain new sentence representation t_i with Eq.(1-8) via the identified subject and existing relation embeddings.

$\{f_1{}^{start\_o}, f_2{}^{start\_o},..., f_{n_o}{}^{start\_o}\}$

$\{f_1{}^{end\_o}, f_2{}^{end\_o},..., f_{n_o}{}^{end\_o}\}$

$Q_o=\{h_1{}^{obj}, h_2{}^{obj},..., h_{n_o}{}^{obj}\}$

Update object extraction loss L_{start_o}, L_{end_o} .

end

end

Update extraction loss L .

end

end

 Volume 9 Issue 3, 2023
DOI: 10.6919/ICJE.202303_9(3).0002

Essentially, relationship extraction is to identify potential triples of entity relationships in the text. The steps of entity relationship joint extraction model algorithm are shown in Table 1.

4. Experimental Design and Results Analysis

4.1 Experimental Design

In this paper, DuIE dataset, an open-source Chinese relational extraction dataset from the 2019 Baidu Information Extraction Contest, is chosen to test the performance of the model [22]. The corpus is constructed from the text of Baidu Encyclopedia, Baidu Info Stream and Baidu Posting Bar, and its schema adds multiple complex relationship types to the traditional simple relationship types, comprehensively covering both written and spoken expression corpus, which can fully investigate the ability of relationship extraction in real business scenarios.

Table 2. Information about DuIE dataset

Category	Train	Dev	Test
Triples	314996	34270	43749
Sentences	155931	17178	21639

The experimental environment and configuration are as follows. The server CPU is Intel(R) Core (TM) i5-12500H, the graphics card is RTX 3050, the RAM is 8GB, the hard disk 512GB, the operating system is Windows 11, the development tool is Pycharm, the development language is Python, and the deep learning framework is Pytorch.

4.2 Test Indicators

For fair comparison, we adopt the evaluation criteria to evaluate our model, which is based on Precision(P), Recall(R)and F1 score (macro average F1 score). In addition, for an extracted triplet <head entity, relation, tail entity>, each element in it is considered correct if and only if it is the same as an element in the dataset. The calculation is as follows.

$$Prcision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F_1 = \frac{2 \times Prcision \times Recall}{Prcision + Recall} \tag{11}$$

Where TP indicates the number of correctly predicted triples, FP indicates the number of incorrectly predicted triples and FN is the number of correct triples that were not predicted.

4.3 Analysis of Experimental Results

Since this paper uses a binary cross-entropy loss function to constrain the head and tail entity types, the parameters of different loss function weights may affect the final effect of the model. In order to study the degree of influence of different core parameters α of loss function weight on the entity relationship joint extraction model, the parameter α is taken from 0.1 to 0.5 and divided into nine groups equally for comparison experiments. The values and trends of Precision, Recall and F1 under different loss function weight parameters are shown in Table 3 and Figure 3.

DOI: 10.6919/ICJE.202303_9(3).0002

Table 3. Comparison data of different parameters

Index	α	Precision/%	Recall/%	F1/%
1	0.10	75.5	78.1	76.8
2	0.15	70.6	83.7	76.6
3	0.20	75.3	81.5	78.3
4	0.25	**77.2**	82.0	**79.5**
5	0.30	73.6	82.0	77.5
6	0.35	75.1	83.1	78.9
7	0.40	68.1	82.7	74.7
8	0.45	75.2	81.9	78.4
9	0.50	71.6	**83.9**	77.3

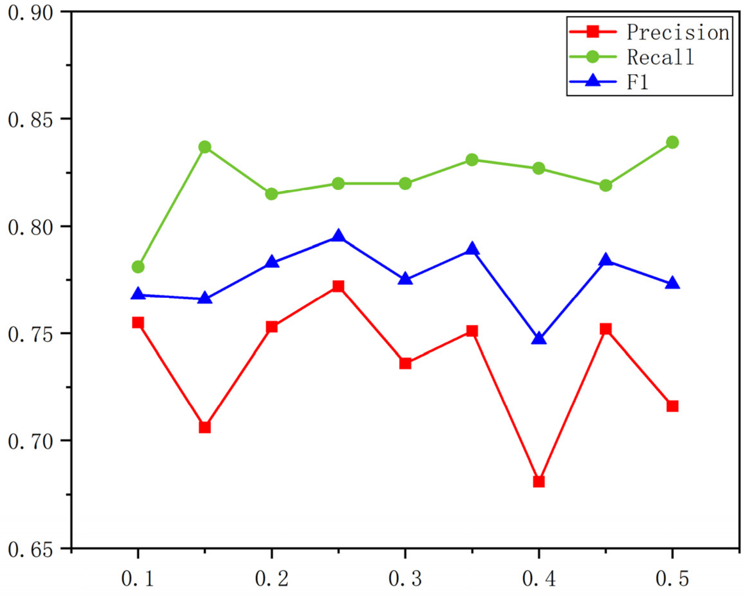

Figure 3. Comparison line chart of different parameters

From Table 3, it can be concluded that the Precision (77.2%) and F1(79.5%) are the highest values when $\alpha = 0.25$. From Figure 3, it can be concluded that the Recall rate increases and then decreases with the increase of α , while the Precision and F1 values show a complex trend with the increase of α . Although the value of recall reaches the highest value when $\alpha = 0.15$, the Precision and F1 values fail to reach the highest value in the same case. Overall, the joint entity relationship extraction model achieves the best results when $\alpha = 0.25$, and the corresponding Precision, Recall, and F1 values are 77.2%, 82.0%, and 79.5%. Therefore, in this paper, the parameter α of the weight of the loss function is set to 0.25 when the head and tail entity types are constrained separately using the binary cross-entropy loss function.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202303_9(3).0002

4.4 Comparison of Related Modelling Techniques

In this paper, we use a representative entity-relationship joint extraction model from recent years as a benchmark to compare and validate the advantages of the proposed model, includes CopyMTL, WDec, Seq2UMTree [23-25].

The results of the experiments are shown in the table, where the best experimental results are marked in bold.

Table 4. Benchmark model comparison data

Models	Precision/%	Recall/%	F1/%
CopyMTL	49.9	39.4	43.9
WDec	64.1	54.2	58.7
Seq2UMTree	75.6	73.0	74.3
Textual model	**77.2**	**82.0**	**79.5**

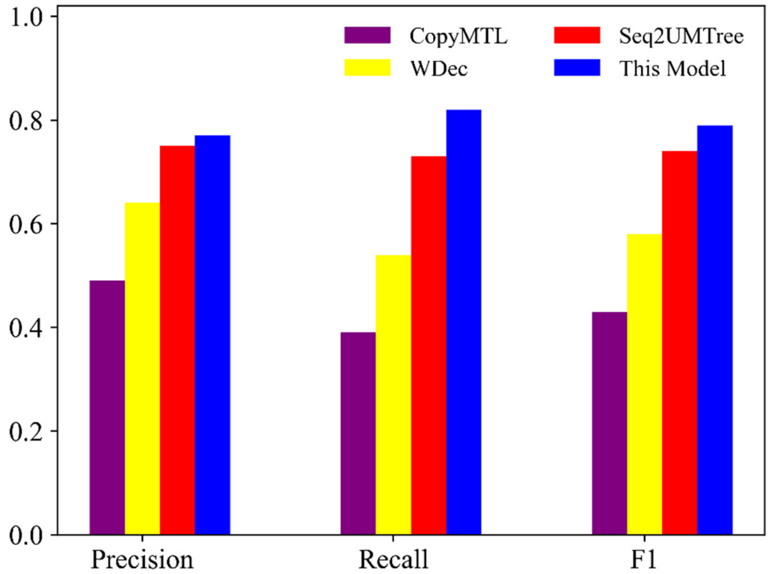

Figure 4. Experimental comparison of different models

The experimental comparison of the data reveals that the model proposed in this paper achieves the highest Precision (77.2%), Recall (82.0%) and F1(79.5%) in this dataset. The comparison revealed that the method using the Duie_Bert pre-trained model for encoding was significantly better than the method using the BiLSTM encoder (CopyMTL), mainly because the encoder modelled using BiLSTM was not able to accurately encode text containing multi-entity overlaps and suffered from the problem of mis-passing. The experimental results also showed that compared to the best performing model in the baseline model, Seq2UMTree, the model proposed in this paper improves 9% in Recall, indicating that the model in this paper has better stability.

In this paper, we implement a cascading pointer annotation approach to relation extraction based on the pre-trained model Duie_Bert, while introducing a specific relation-entity guided multi-headed attention mechanism. The method not only takes full account of the fine-grained semantic information in the sentence, but also effectively improves the accuracy of entity extraction.

4.5 Ablation Experiment

In order to analyses the performance of the modules in the model, ablation experiments are conducted on the Duie dataset to verify the effectiveness of each module.

Table 5. Comparative data from ablation experiments

Models	Precision/%	Recall/%	F1/%
This paper-Duie_Bert	63.1	62.3	62.6
This paper-Attention	65.6	70.6	68.0
This paper	**77.2**	**82.0**	**79.5**

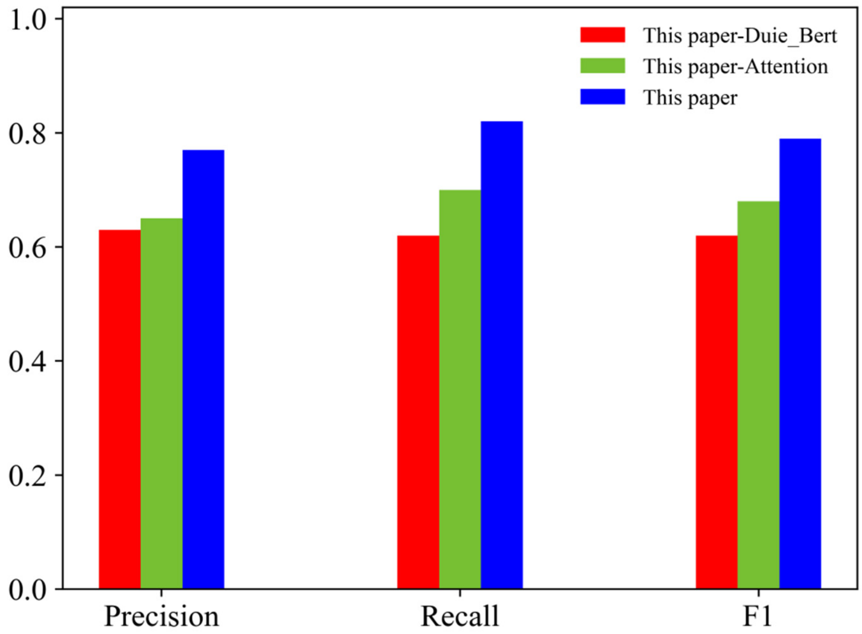

Figure 5. Comparative analysis chart of ablation experiments

The complete relationship extraction model proposed in this paper is compared with the model with the Duie_Bert module removed, the model of the multi-headed attention mechanism for the specific relationship-entity removed, respectively. The experimental results are shown in the table.

From the experimental results, it can be observed that when the Duie_Bert model is removed and the multi-headed attention mechanism is removed, the model decreases to some extent in all three metrics of P, R and F1. With the addition of the Duie_Bert module, the Precision improved 14.1% and the F1 by 16.9%, indicating that the pre-trained BERT model is effective in improving the accuracy and stability of triad extraction.

With the addition of the relationship-entity specific multi-headed attention mechanism, the precision, recall and F1 values improved by 11.6%, 11.4% and 11.5% respectively. The results show that the model proposed in this paper can effectively improve all the performance of relationship extraction, and further demonstrate the impact of each module on the overall performance of the model.

5. Conclusion

In response to the problems of overlapping relational triples and multi-entity miscommunication in recent relation extraction research, this paper proposes a joint entity relationship extraction model based on an improved cascading pointer network. On the Duie Chinese relational dataset, the performance is compared with the current three latest entity-relational joint extraction models. The experimental results show that the proposed model achieves a performance improvement of more than 9.0% and more than 5.2% in the recall and F1 score, respectively. The next step is to study the domain-oriented knowledge graph construction method and complete the construction of the knowledge graph based on the entity-relationship federated extraction model proposed in this paper.

Acknowledgments

The paper is partly supported by Project of Leading Talents in Science and Technology Innovation in Henan Province under Grant 204200510021, Program for Henan Province Key Science and Technology under Grant 222102210177, Grant 222102210072 and Grant 212102210383, Henan Province University Key Scientific Research Project under Grant 23A520008. The authors show gratitude to the constructive comments and guidance's from the reviewers during the whole reviewing process.

References

[1] Chinchor N, Marsh E. Muc-7 information extraction task definition[C]//Proceeding of the seventh message understanding conference (MUC-7), Appendices. 1998: 359-367.

[2] E HH, Zhang WJ, Xiao SQ, et al. Survey of Entity Relationship Extraction Based on Deep Learning[J]. Journal of Software, 2019, 30(6):1793-1818.

[3] Liu KB, Li F, Liu L, et al. Implement of a kernel-based Chinese relation extraction system[J]. Journal of Computer Research and Development, 2007, 44(8): 1406-1411.

[4] Zeng DJ, Liu K, Lai SW, et al. Relation classification via convolutional deep neural network[C]// Proceedings of Coling, pages 2335–2344. 2014.

[5] Gao D, Peng DL, Liu C. Entity Relation Extraction Based on CNN in Large-scale Text Data[J]. Small Microcomputer System, 2018, 39(05):1021-1026.

[6] Liu K. A survey on neural relation extraction[J]. Science China (Technological Sciences), 2020, 63 (10): 1971-1989.

[7] Li DM, Zhang Y, Li DY, et al. Review of entity relation extraction methods[J]. Journal of Computer Research and Development, 2020, 57(7): 1424-1448(in Chinese).

[8] Kumar S. A survey of deep learning methods for relation extraction. arXiv: arXiv preprint arXiv: 1705.03645, 2017.

[9] Huang X, You HL, Yu Y. A Survey of Research on Relationship Extraction Technology. New Technology of Library and Information Service, 2013, 29(11): 30-39(in Chinese with English abstract). [doi:10. 11925/ infotech.1003-3513.2013.11.05]

[10] Jiao LC, Yang SY, Liu F, Wang SG, et al. Neural Network Seventy Years: Retrospect and Prospect[J]. Chinese Journal of Computers, 2016, 39(8): 1697-1716(in Chinese with English abstract).

[11] Golshan PN, Dashti HAR, Azizi S, et al. A study of recent contributions on information extraction. arXiv preprint arXiv: 1803.05667, 2018.

[12] Wang LY. Entity relationship extraction based on deep convolutional neural network [MS. Thesis]. Taiyuan: Taiyuan University of Technology, 2017(in Chinese).

[13] Yang JF, Yu QB, Guan Y, et al. A survey of research on electronic medical record named entity recognition and entity relationship extraction[J]. Acta Automatica Sinica, 2014, 40(8): 1537-1562(in Chinese with English abstract).

[14] Socher R, Huval B, Manning CD, et al. Semantic compositionality through recursive matrix-vector spaces [C]// Proceeding of the Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012.1201-1211.

[15] Xu Y, Mou L, Li G, et al. Classifying relations via long short-term memory networks along shortest dependency paths[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1785-1794.

[16] Zhou D, Zhong D, He Y. Biomedical relation extraction: from binary to complex[J]. Computational and mathematical methods in medicine, 2014, 2014.

[17] Zheng SC, Wang F, Bao HY, et al. Joint extraction of entities and relations based on a novel tagging scheme [C]// Proc of the 55th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2017: 1227–1236.

[18] Zeng XG, Zeng DJ, He SZ, et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]//Proc of the 56th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2018: 506–514.

[19] Fu TJ, Li P, Ma WY. GraphRel: Modeling text as relational graphs for joint entity and relation extraction[C]//Proc of the 57th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2019: 1409–1418.

[20] Wei ZP, Su JL, Wang Y, et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]//Proc of the 58th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, PA: Association for Computational Linguistics, 2020: 1476-1488.

[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceeding of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA: December 4-9, 2017. MIT Press: Cambridge, MA, 2017: 5998-6008.

[22] Li SJ, He W,Shi YB, et al. DuIE: a large-scale Chinese dataset for information extraction [C]// Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing. Beijing, China: CCF,2019:791-800.

[23] Zeng D, Zhang H, Liu Q. Copymtl: copy mechanism for joint extraction of entities and relations with multi-task learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5):9507-9514

[24] Nayak T, Ng HT. Effective modeling of encoder-decoder architecture for joint entity and relation extraction [C]//Proceedings of the AAAI Conference on Artificial Intelli gence,2020,34(5):8528-8535.

[25] Zhang HR, Lliu QY, Fan XM, et al. Minimize exposure Bias of Seq2Seq models in joint entity and relation extraction [EB/OL]. (2020-09-16) [2021-01-02]. https: //arxiv. org/pdf/2009. 07503. pdf.

[26] Devlin J, Chang MW, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of NAACL-HLT, Minneapolis, Minnesota: June 2-7, 2019. Stroudsburg: ACL, 2019: 4171-4186.