Document-Level Relation Extraction with Relational Reasoning and Heterogeneous Graph Neural Networks

Wanting Ji, Yanting Dong, and Tingwei Chen(✉)

School of Information, Liaoning University, Shenyang, China
twchen@lnu.edu.cn

Abstract. Document-level relation extraction aims to identify the relations between the entities in an unstructured text and represents them in a structured way for downstream tasks such as knowledge graphs and question answering. In recent years, graph neural network-based methods have made significant progress in relation extraction. However, these methods usually require extracting all the entities in the document first, then a classifier is used to analyze the relations between the entities regardless of whether they have any relation. This wastes a lot of time analyzing the relations of irrelevant entity pairs and reduces the classifier's attention to relevant entity pairs. To address this issue, this paper proposes a relation extraction module that integrates **R**elational **R**easoning and **H**eterogeneous **G**raph neural **N**etworks (RRHGN). The method finds a meta-path for each entity pair in a document and uses multi-hop reasoning to analyze the entities on the meta-path to determine whether there is a strong reasoning path between the entity pair. The relational reasoning module built into the method makes the classifier focus more on the relevant entity pairs in the document, thus reducing the task burden of the classifier and improving the accuracy of entity relation extraction. Experimental results on the large-scale document-level relation extraction dataset DocRED show that the proposed method achieves significant performance improvement compared with existing methods.

Keywords: Document-level relation extraction · Heterogeneous graph neural network · Multi-hop reasoning

1 Introduction

The purpose of document-level relation extraction is to extract the relations between different entities in a document and to represent them in a structured way. It plays an important role in natural language processing tasks such as information retrieval [1], question answering [2], and dialogue system [3]. Usually, document-level relation extraction involves a large number of entities, and

X. Yang et al. (Eds.): ADMA 2023, LNAI 14179, pp. 182–195, 2023.
https://doi.org/10.1007/978-3-031-46674-8_13

Kungliga Hovkapellet

[1] ***Kungliga Hovkapellet*** (The ***Royal Court Orchestra***) is a *Swedish* orchestra, originally part of the *Royal Court* in ***Sweden***'s capital *Stockholm*. [2] The orchestra originally consisted of both musicians and singers. [3] It had only male members until *1727,* when *Sophia Schroder* and *Judith Fischer* were employed as vocalists; in the *1850s*, the harpist *Marie Pauline Ahman* became the first female instrumentalist.[4] From *1731* public concerts were performed at ***Riddarhuset*** in *Stockholm*, [5] Since 1773, when the ***Royal Swedish Opera*** was founded by *Gustav III* of ***Sweden***, the ***Kungliga Hovkapellet*** has been part of the opera's company.

Subject: ***Riddarhuset***
Object: ***Sweden***
Relation: country
Supporting Evidence: 1, 4

Fig. 1. A example from the dataset DocRED.

these entities are sparsely distributed in multiple sentences that constitute a document. According to the statistics of human tagged corpus extracted from Wikipedia documents, more than 40.7% of entity relationship facts need to be jointly extracted from multiple sentences. Therefore, it is very necessary to study document-level relationship extraction methods [4,5]. Recently, some researches introduced graph data structure into the task of document-level relationship extraction [6–8]. The common way is to construct document-level heterogeneous graphs according to different entity types, and then encoding the graphs using attention mechanism, finally classifying the relationships among entities in the graphs using classifiers. However, such methods need to extract all the entities in a document first, and then classify the relationships among the entities. In this process, relation analysis for a large number of unrelated entity pairs not only distracts the attention of the classifier, but also reduces the efficiency of the classifier.

Figure 1 is an example in DocRED dataset. Entities (Riddarhuset, Sweden) are pairs of entities to be classified. They are located in sentences 1 and 4 respectively, and need to be obtained by relational reasoning between sentences. However, sentences 1 and 4 contain a large number of irrelevant entities (e.g., Kunglia Hovkapellet, Roval Court, 1731), and the relational classifier needs to classify them regardless of whether there is a relationship between these entities. Obviously, these irrelevant entity pairs distract the classifier. Usually, judging whether there is a relationship between two entities across sentences requires reasoning, and there is often a reasoning path for related entity pairs, such as sentences 1 and 4. We can judge that Stockholm is the capital of Sweden through the first sentence, and that Riddarhuset is an area of Stockholm through the fourth

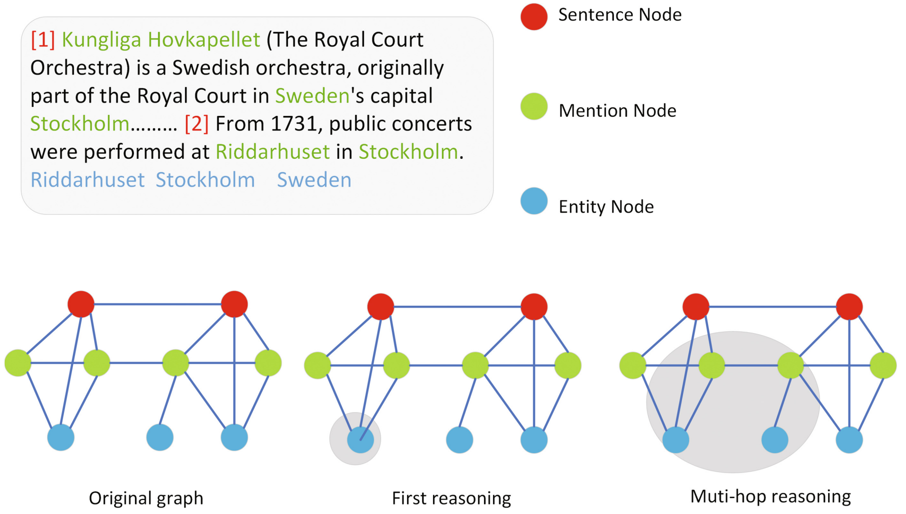

Fig. 2. The reasoning process of RRHGN.

sentence. Therefore, through reasoning between these two sentences, we can get (Riddarhuset, Sweden) that the relationship between the two entities is (country). If there is a relationship between two entities, we can find a reasoning path. However, the existing methods need to extract the relationship between entities regardless of whether there is a relationship between them, which greatly reduces the efficiency of task execution.

To solve the above problems, this paper proposes a document-level relationship extraction method based on relational reasoning and heterogeneous graph neural network (RRHGN). In this method, a relational reasoning module is built to judge whether there is a strong reasoning path between two entities, to predict the probability of relationship between these two entities. By constructing a dynamic graph structure, relational reasoning module is completed through multiple iterations on the selected meta-path. In each reasoning process, only the related nodes are reserved, and the irrelevant nodes are shielded, thus ensuring that all useful information is transmitted. Figure 2 shows the reasoning process of RRHGN, which tries to find a strong reasoning path between two entity pairs, so that the classifier pays more attention to those related entity pairs and completes the relationship extraction better. Gray denotes nodes in the reasoning process.

The main contributions of this paper are as follows:

(1) A relation reasoning module is proposed for relation extraction of graph structure to solve the problem that irrelevant entity pairs will distract the attention of entities in the process of relation reasoning using graph structure.
(2) Through multi-hop reasoning on meta-path nodes, it can be judged whether there is a strong reasoning path between entities.

(3) Experiments on DocRED, a large relational extraction dataset, show that the proposed method can accurately predict the relationships between entities.

2 Related Work

Existing document-level relation extraction methods can be roughly divided into sequence-based methods and graph-based methods.

2.1 Sequence-Based Document-Level Relationship Extraction Methods

Sequence-based document-level relationship extraction methods directly use neural networks to learn entity representations in documents, and classify all sent entity pairs. Zhou et al. [9] proposed a global context-enhanced graph convolution network model, which combined Transformer encoder with graph neural network, and considered both global and local dependencies among entities. Ye et al. [10] proposed a pre-training model based on BERT, which enhanced the reference reasoning ability of language representation by introducing reference resolution task, carried out document-level relation extraction experiments on DocRED dataset, and achieved very good results. Zeng et al. [11] proposed a model for separating intra-sentence and cross-sentence reasoning, which uses Transformer encoder to process each sentence and the whole document respectively, and uses graph convolution network to classify relations. Giorgi et al. [12] developed a sequence-to-sequence approach, seq2rel, that can learn the subtasks of DocRE (entity extraction, coreference resolution and relation extraction) end-to-end, replacing a pipeline of task-specific components. Liu et al. [13] proposed an effective structure enhanced transformer encoder model (SETE), integrating entity structural information into the transformer encoder. However, for long documents, sequence based methods are prone to losing semantic relationships and cannot effectively obtain global information.

2.2 Graph-Based Document-Level Relation Extraction Methods

Graph-based document-level relationship extraction methods often need to model documents according to the relationship between entities and sentences in documents, and use graph neural network [14] to build document graphs and learn the related information between entities. Some researches use graph convolution network (GCN) [15] to extract document-level relations, but these methods do not make full use of the global information of documents. To solve this problem, Sahu et al. [7] proposed a labeled graph convolution neural network model (GCNN), which uses cross-sentence and intra-sentence dependencies to capture local and non-local dependency information. Park et al. [16] used a graph structure and an entity attention awareness mechanism to capture the global information of documents. Hu et al. [17] proposed a multi-granularity interactive network (HAIN) to capture global information at three levels: word, sentence

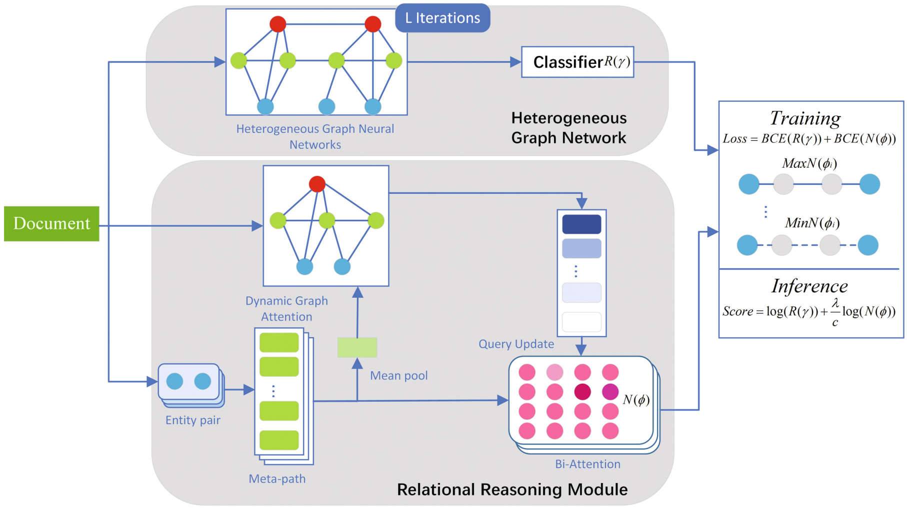

Fig. 3. Overall overview of the method.

and document-level. In addition, Zeng et al. [18] proposed a graph aggregation reasoning network (GAIN) with double graph features, considering that a single global graph can not get complete global information, which used multiple hierarchical networks to extract structured features. Nan et al. [19] used graph structure for multi-hop reasoning, which effectively solves the problems of long distance and implicit relationship. Sun et al. [20] proposed Dual-Channel and Hierarchical Graph Convolutional Networks (DHGCN), which constructed three graphs in token-level, mention-level, and entity-level to model complex interactions among different semantic representations across the document. Based on the multi-level graphs, they applied the Graph Convolutional Network (GCN) for each level to aggregate the relevant information scattered throughout the document for better inferring the implicit relations. Although these methods capture the global information well, they do not take into account that not all entity pairs need relation extraction, and some irrelevant entity pairs will distract the classifier's attention from related entity pairs. Therefore, this paper proposes a relational reasoning module based on graph structure, which tries to find a strong reasoning path through reasoning analysis on meta-path, and helps classifiers to extract relations better.

3 The Proposed Method

In this paper, a document-level relationship extraction method based on relation reasoning and heterogeneous graph neural network (RRHGN) is proposed for document-level relationship extraction. As shown in Fig. 3, the method mainly includes two parts: a heterogeneous graph network and a relational reasoning module. Heterogeneous graph network uses self-attention mechanism to encode entities in heterogeneous graph, and a multi-layer perception is used as a classifier to extract the relationship of entity pairs. Considering that most irrelevant entities will distract the attention of related entity pairs, this paper proposes a relational reasoning module, which starts from one of the entity pairs, makes reasoning analysis on the meta-path, calculates the probability of the existence of meta-path, and judges whether there is a relationship between entity pairs, so that the classifier pays more attention to related entity pairs, which increases the efficiency and accuracy of relational classification.

3.1 Heterogeneous Graph Network

Construction of Heterogeneous Graph. Referring to the construction of heterogeneous graph by Xu et al [27], this paper defines three types of nodes: Sentence node, Mention node and Entity node, and defines six types of edges: Sentence-Sentence (SS), Mention-Sentence (MS), Mention-Mention (MM), Entity-Mention (EM), Entity-Sentence (ES) and Mention-Coref- erence (CO). Therefore, a document can generate an adjacency matrix to represent the connection between nodes. The final document can be represented by a heterogeneous diagram $G = \{V, E\}$.

Encoder. Following the work of Guo et al. [21], we use graph attention network to encode each node in the heterogeneous graph to obtain an effective graph representation. Let h_n^l be the initial node, we first concatenate the outputs of all the previous l layers of $\{s_n^1, s_n^2, \cdot, \cdot\cdot, s_n^{l-1}\}$ and transform them into a fixed-dimensional vector:

$$h_n^l = W_e^l \cdot [v_n : s_n^1 : s_n^2 : \cdots : s_n^{l-1}] \tag{1}$$

where $s_n^{l-1} \in \mathbb{R}^{d_0}$,$W_e^l \in \mathbb{R}^{d_0 \times (l \times d_0)}$. We use self-attention mechanism[20] to extract the feature relations of C neighbor nodes $\{h_{a1}^l, h_{a2}^l, \cdot, \cdot\cdot, h_{ac}^l\}$ and h_n^l connected to v_n. Here, K and V are key-value matrices determined by the types of edges of the neighbor nodes:

$$s_n^l = \mathrm{softmax}(\frac{h_n^l K^T}{\sqrt{d_0}})V \tag{2}$$

Finally, combine the node v_n and the relation information of the document through a non-linear layer to obtain the global information of the document:

$$q_n = \mathrm{Relu}(W_0 \cdot [v_n : s_n^1 : \cdots : s_n^l] \tag{3}$$

The heterogeneous graph is finally represented as: $G = (q_1, q_2, \cdot \cdot \cdot, q_n)$.

Classifier. Classifier is a sigmod function with a multi-layer perceptron (MLP) to compute the probability of the relation:

$$R(r) = P(r|\{e_i, e_j\} = \mathrm{sigmod}(\mathrm{MLP}([q_i, q_j])) \tag{4}$$

3.2 Relation Reasoning Module

The classification strategy of classifying all entity pairs by the classifier is obviously unreasonable. Therefore, the relation reasoning module is used to judge whether there is a relation between entity pairs, so that the classifier can pay more attention to the related entity pairs and improve the efficiency and accuracy of classification.

Meta-Path. When there is a relation between two entity pairs, a strong reasoning path can usually be found to prove that there is indeed a relation between the entity pairs. Conversely, when there is no relation between two entity pairs, such a strong reasoning path cannot be found.

Therefore, we need to find such a strong reasoning path to prove that there is indeed a relation between the entity pairs. Hence, this paper defines three meta-paths to infer whether there is such a strong reasoning path between the entity pairs [23].

(1) Pattern recognition: In this form of reasoning, two entities are connected by a sentence, and the relation pattern is EM-MM-ME.
(2) Logical reasoning: In this form, two entities are connected by a common entity, and the relation pattern is EM-MM-CO-MM-ME.
(3) Coreference reasoning: In this form of reasoning, two entities appear in a sentence, and the relation pattern is ES-SS-SE.

Different entity pairs have one or more meta-paths between them. Therefore, we prioritize the meta-paths according to the priority: pattern recognition > logical reasoning > coreference reasoning. Many entity pairs have multiple paths for the same meta-path type, but according to the document writing habit, the entities that appear later are often replaced by pronouns, and the entities usually appear for the first time at the beginning of the article. Therefore, we choose the first meta-path that appears in the document.

Relation Reasoning. For each entity pair $\{e_{1n}, e_{2n}\}$, a meta-path $\phi_n = \{Q^1, Q^2, \cdots, Q^t\}$ can be found. A graph neural network is used to propagate the node information of an entity to its neighboring nodes. A dynamic graph attention mechanism is employed to simulate the reasoning process. In the reasoning stage, if each node needs to propagate information to its neighboring nodes, then the more relevant two nodes are, the more information will be propagated. This paper only allows the node information that is related to the query to be propagated. An attention network between the query and the entities is used to predict a mask m_t, which obtains the starting entity in the t-th reasoning step.

$$\tilde{q}^t = \text{Meanpool}(Q^t) \tag{5}$$

$$\gamma^{(t)} = \frac{\tilde{q}^t V^{(t+1)} e_i^t}{\sqrt{d_2}} \tag{6}$$

$$m^{(t+1)} = \text{sigmod}([\gamma_1^{(t+1)}, \cdots, \gamma_N^{(t+1)}]) \tag{7}$$

$$\tilde{E^{(t)}} = [m_1^{(t+1)} e_1^t, \cdots, m_N^{(t+1)} e_N^t] \tag{8}$$

where V_t is a linear projection function that multiplies the entity node with the mask, encouraging the required initial entities, and other unnecessary entities will be penalized, so this can limit the information dissemination of irrelevant nodes.

Then the graph attention method (GAT) [24] is used to compute the attention score α between a meta-path node and its neighbors:

$$h_i^{(t+1)} = U_t e_i^{(t+1)} + b_t \tag{9}$$

$$\beta_{i,j}^{(t+1)} = \text{LeakyReLU}(W_t^T [h_i^{(t+1)}, h_j^{(t+1)}]) \tag{10}$$

$$\alpha_{i,j}^{(t+1)} = \frac{\exp(\beta_{i,j}^{(t+1)})}{\sum_k exp(\beta_{i,k}^{(t+1)})} \tag{11}$$

where $U_t \in \mathbb{R}^{d_2 \times 2d_2}$ and $W_t \in \mathbb{R}^{2d_2}$ are linear projection parameters, and α represents the proportion of neighbor information assigned to neighbor entity j in row i.

Summing each node column-wise yields a new entity containing information gathered from neighbor nodes:

$$e_i^{(t+1)} = \text{ReLU}(\sum_{j \in B_i} \alpha_{j,i}^{(t+1)} h_j^{(t+1)}) \tag{12}$$

where B_i is the set of neighbors of entity i, and finally we get the updated entity embedding$E^{(t+1)} = [e_1^{(t+1)}, \cdots, e_N^{(t+1)}]$.

Relational reasoning consists of multiple steps, and the newly visited entity in the previous step will be the starting entity for the next step. Here we use the modified Bi-Attention network [25] to update the probability of query reasoning to the next step:

$$p(Q^{(t+1)}|Q^t) = \text{Bi-Attention}(Q^{(t)}, E^{(t+1)}, Q^{(t+1)}) \tag{13}$$

The probability of the final whole meta-path is expressed as:

$$N(\phi_n) = \prod_1^C p(Q_{(c+1)}^{(t+1)}|Q_c^t) \tag{14}$$

where C is the number of probabilities reasoned on the meta-path.

3.3 Path Reasoning

Using the relational reasoning module as a relational indicator, when classifying relations, the auxiliary classifier performs relational classification:

$$S(r) = \log(R(r)) + \lambda \cdot \frac{1}{C} \log\left(N\left(\phi_n\right)\right) \tag{15}$$

where λ is a hyper-parameter that controls the importance of relational reasoning.

3.4 Loss Function

When training the proposed method, this paper uses the binary cross loss to train the triplet (subject, object, relation) in the dataset, namely $\{\{e1_n^t, e2_n^t, r_n^t\}_{n=1}^{N_t}\}_{t=1}^{T}$, to optimize the parameters of the neural network.

Loss function for the heterogeneous graph network:

$$Loss_h = -\frac{1}{\sum_{t=0}^{T} N_t} \sum_{t=1}^{T} \sum_{n=1}^{N_t} \left\{ r_n^t \log \left(R \left(r_n^t \right) \right) \right\} + \left(1 - r_n^t \right) \log \left(1 - R \left(r_n^t \right) \right) \quad (16)$$

Loss function for the relational reasoning module:

$$Loss_r = -\frac{1}{\sum_{t=0}^{T} N_t} \sum_{t=1}^{T} \sum_{n=1}^{N_t} \left\{ r_n^t \log N \left(\phi_n \right) \right\} + \left(1 - r_n^t \right) \log \left(1 - N \left(\phi_n \right) \right) \quad (17)$$

where $r_n^t \in (0, 1)$. Finally, the whole loss of RRHGN is the sum of the heterogeneous graph network loss and the relational reasoning module loss:

$$Loss = Loss_r + Loss_h \quad (18)$$

4 Experiments

4.1 Dataset

This paper uses a widely used document-level relation extraction dataset DocRED for experiments. DocRED is a large-scale human-annotated document-level relation extraction dataset built from Wikipedia and Wikidata. DocRED contains 132,375 entities and 56,354 relational facts, which are annotated on 5,053 Wikipedia documents. It is currently the largest human-annotated document-level relationship extraction Dataset [4].

4.2 Experimental Setting

All the experiments in this paper are completed on the Ubuntu20.4 platform, the CPU uses Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz, and the graphics card uses NVIDIA A40 GPU. The language used is Python3.9, the encoder uses GloVe embedding (100d), the model optimizer uses Adam, and the optimal parameter settings of the model are shown in Table 1.

For evaluation, on DocRED, following Yao et al. [4], we use the widely employed F1 and Ign F1 as the evaluation metrics. Ign F1 refers to excluding the relational facts shared by the training and dev/test sets. F1 is defined as:

$$F1 = \frac{2 \times P \times R}{P + R}$$

where precision(P) and recall(R) are defined as:

$$P = \frac{\text{True Positives}}{\text{True Positives+False Positives}}$$

$$R = \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$$

Table 1. Model parameter values.

parameter	values
learning rate	1e-4
l	2
batch size	32
epoch	300

4.3 Baseline

This paper compares the proposed method with existing sequence-based document-level relation extraction methods (convolution neural networks (CNN) [4], bidirectional LSTM (BiLSTM) [4], Context-Aware LSTM [4], HIN-Glove [26]) and graph-based document-level relation extraction methods (GAT [24], GCNN [8], EOG [6], AGGCN [21], LSR-Glove [19], GAIN-Glove [18]) in DocRED. The performance on the DocRED dataset was compared.

4.4 Experimental Results

Table 2 presents the experimental results of different document-level relation extraction methods on the DocRED dataset.

Table 2. Comparison of Model Experiment Results.

Model Name	Dev		Test	
	Ign F1	F1	Ign F1	F1
CNN	41.58	43.45	40.33	42.26
LSTM	48.44	50.68	47.71	50.70
BiLSTM	48.87	50.94	48.78	51.06
Context-Aware	48.94	51.09	48.40	50.70
HIN-GloVe	51.06	52.95	51.15	53.30
GAT	45.17	51.44	47.36	49.51
GCNN	46.22	51.52	49.59	51.62
EOG	45.94	52.15	49.48	51.82
AGGCN	46.29	52.47	48.89	51.45
LSR-GloVe	48.82	55.17	52.15	54.18
GAIN-GloVe	53.05	55.29	52.66	55.08
RRHGN-GloVe(ours)	**54.23**	**55.80**	**53.47**	**55.54**

In the model based on Glove for word vector representation, the F1 of the method proposed in this paper is higher than that of the sequence-based and

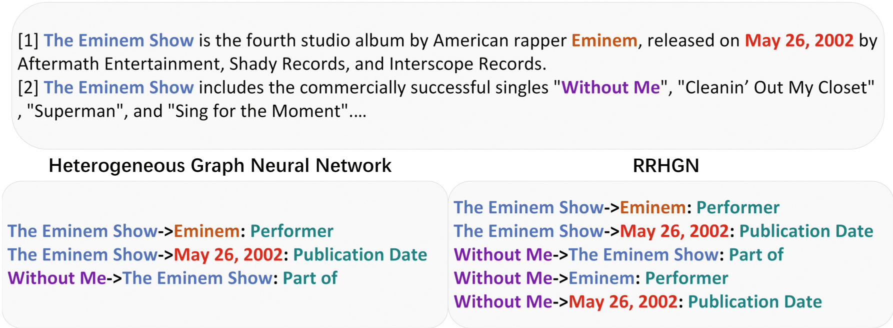

Fig. 4. The case study of our proposed RRHGN and baseline.

graph-based baseline models by 0.46-13.28 in the test set, and achieved good experimental results, reflecting the superiority of the RRHGN. The relational reasoning module in this paper judges whether there is a relationship between the entity pairs by analyzing the meta-path of the entity pair, so that the model can pay more attention to the entity pairs that have relationships, and is more conducive to relationship classification.

Table 3. Results of ablation experiments.

Model Name	F1
Heterogeneous Graph Network	53.52
Heterogeneous Graph Network + Relational Reasoning Module(ours)	55.54

4.5 Ablation Experiments

In order to verify the gain effect of the relational reasoning module on heterogeneous graph network, this paper conducted an ablation experiment on the DocRED dataset, and the experimental results are shown in Table 3.

Among them, the heterogeneous graph network means that only heterogeneous graphs are used to directly extract relationships.

Experimental results show that RRHGN proposed in this paper makes a positive contribution on the task of relation extraction. RRHGN improves the original basic model by 2.02% points. It can be seen that the RRHGN model can improve the accuracy of relation extraction by constructing meta-paths and performing reasoning analysis.

4.6 Case Study

Figure 4 shows a case study of our proposed method RRHGN and baseline. Heterogeneous graph network can identify the relationships between entity pairs within the same sentence, but their performance across sentences is not ideal. RRHGN has achieved good results in cross sentence relationship extraction through relational reasoning.

5 Conclusion

This paper proposes RRHGN to solve the problem that a large number of irrelevant entity pairs distracts the classifier from relational entity pairs. The proposed method judges whether there is a relationship between entity pairs by reasoning and analyzing the meta-paths between entity pairs, and provides a basis for the classifier. RRHGN acts as an indicator to assist classifiers when classifying relations. Experiments show that the proposed method improves the accuracy and efficiency of relation extraction. Although the model proposed in this paper has a certain degree of improvement in the task of relation extraction, how to further reduce the expenditure of computing resources, and how to update and query through the simplest and most effective method while be considered in our future work.

References

1. Kadry, A., Dietz, L.: Open relation extraction for support passage retrieval: merit and open issues. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1149–1152 (2017)
2. Yu, M., Yin, W., Hasan, K.S., dos Santos, C., Xiang, B., Zhou, B.: Improved neural relation detection for knowledge base question answering. In: Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics (ACL) (2017)
3. Young, T., Cambria, E., Chaturvedi, I., Huang, M., Zhou, H., Biswas, S.: Augmenting end-to-end dialog systems with commonsense knowledge (2017). arXiv preprint arXiv:1709.05453
4. Yao, Y., et al.: Docred: a large-scale document-level relation extraction dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 764–777 (2019)
5. Cheng, Q., et al.: Hacred: a large-scale relation extraction dataset toward hard cases in practical applications. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 2819–2831 (2021)
6. Christopoulou, F., Miwa, M., Ananiadou, S.: Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 4925–4936 (2019)
7. Sahu, S.K., Christopoulou, F., Miwa, M., Ananiadou, S.: Inter-sentence relation extraction with document-level graph convolutional neural network. arXiv preprint arXiv:1906.04684 (2019)
8. Li, B., Ye, W., Sheng, Z., Xie, R., Xi, X., Zhang, S.: Graph enhanced dual attention network for document-level relation extraction. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 1551–1560 (2020)

9. Zhou, H., Xu, Y., Yao, W., Liu, Z., Lang, C., Jiang, H.: Global context-enhanced graph convolutional networks for document-level relation extraction. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5259–5270 (2020)
10. Ye, D., et al.: Coreferential reasoning learning for language representation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7170–7186 (2020)
11. Zeng, S., Wu, Y., Chang, B.: Sire: separate intra-and inter-sentential reasoning for document-level relation extraction. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 524–534 (2021)
12. Giorgi, J., Bader, G., Wang, B.: A sequence-to-sequence approach for document-level relation extraction. In: Proceedings of the 21st Workshop on Biomedical Language Processing, pp. 10–25 (2022)
13. Liu, W., Zhou, L., Zeng, D., Qu, H.: Document-level relation extraction with structure enhanced transformer encoder. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2022)
14. Zhou, J., et al.: Graph neural networks: a review of methods and applications. AI Open **1**, 57–81 (2020)
15. Zhang, S., Tong, H., Xu, J., Maciejewski, R.: Graph convolutional networks: a comprehensive review. Comput. Soc. Networks **6**(1), 1–23 (2019)
16. Park, S., Yoon, D., Kim, H.: Improving graph-based document-level relation extraction model with novel graph structure. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 4379–4383 (2022)
17. Hu, N., Zhang, T., Yang, S., Nong, W., He, X.: HAIN: hierarchical aggregation and inference network for document-level relation extraction. In: Wang, L., Feng, Y., Hong, Yu., He, R. (eds.) NLPCC 2021. LNCS (LNAI), vol. 13028, pp. 325–337. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88480-2_26
18. Zeng, S., Xu, R., Chang, B., Li, L.: Double graph based reasoning for document-level relation extraction. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1630–1640 (2020)
19. Nan, G., Guo, Z., Sekulić, I., Lu, W.: Reasoning with latent structure refinement for document-level relation extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 1546–1557 (2020)
20. Sun, Q., et al.: Dual-channel and hierarchical graph convolutional networks for document-level relation extraction. Expert Syst. Appl. **205**, 117678 (2022)
21. Guo, Z., Zhang, Y., Lu, W.: Attention guided graph convolutional networks for relation extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 241–251 (2019)
22. Vaswani, A., et al.: Attention is all you need. Advances in neural information processing systems 30 (2017)
23. Sun, Y., Han, J.: Mining heterogeneous information networks: a structural analysis approach. ACM SIGKDD Explorations Newsl. **14**(2), 20–28 (2013)
24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
25. Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi, H.: Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603 (2016)

26. Tang, H., et al.: HIN: hierarchical inference network for document-level relation extraction. In: Lauw, H.W., Wong, R.C.-W., Ntoulas, A., Lim, E.-P., Ng, S.-K., Pan, S.J. (eds.) PAKDD 2020. LNCS (LNAI), vol. 12084, pp. 197–209. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-47426-3_16
27. Xu, W., Chen, K., Zhao, T.: Document-level relation extraction with reconstruction. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 14167–14175 (2021)