



CorefDRE: Coref-Aware Document-Level Relation Extraction

Zhongxuan Xue[✉], Jiang Zhong[✉], Qizhu Dai[✉], and Rongzhen Li[✉]

College of Computer Science, Chongqing University, Chongqing 400044, China
{xuezhongxuan,zhongjiang,daiqizhu,lirongzhen}@cqu.edu.cn

Abstract. Document-level Relation Extraction (Doc-level RE) aims to extract relations among entities from a document, which requires reasoning over multiple sentences. The pronouns are ubiquitous in the document, which can provide reasoning clues for Doc-level RE. However, previous works do not take the pronouns into account. In this paper, we propose **Coref-aware Doc-level RE** based on Graph Inference Network (CorefDRE) to infer relations. CorefDRE first dynamically constructs the heterogeneous **Mention-Pronoun Affinity Graph** (MPAG) by integrating coreference information of pronouns. Then, **Entity Graph** (EG) is aggregated from MPAG through the weight of mention-pronoun pairs, calculated by the noise suppression mechanism, and GCN. Finally, we infer relations between entities based the normalized EG. Moreover, We introduce the noise suppression mechanism via calculating affinity between pronouns and corresponding mentions to filter the noise caused by pronouns. Experimental results significantly outperform baselines by nearly 1.7–2.0 in F1 on three public datasets, DocRED, DialogRE, and MPDD. We further conduct ablation experiments to demonstrate the effectiveness of the proposed MPAG structure and the noise suppression mechanism.

Keywords: Document-level relation extraction · Mention-pronoun affinity graph · Noise suppression

1 Introduction

Relation Extraction (RE), a task that automatically extracts relational facts among entities from raw texts, is widely used in knowledge base construction [22] and question answering [18]. Previous researches mainly focus on sentence-level RE, which aims to identify relations between an entity pair in a single sentence. However, large amounts of relational facts are expressed by multiple sentences, which cannot be achieved by sentence-level RE. Therefore, researchers gradually pay more attention to document-level RE.

Doc-level RE not only handles the sentence-level RE but also captures complex interactions among cross-sentence entities in the document. Recent studies focus on graph-based reasoning skills [5, 14, 16], where coreference information,

[1] **Colette de Jouvenel**, also known as **Bel-Gazou**, July 1913 - 1981, was the daughter of French writer **Colette** and her second husband, Henri de Jouvenel. [2]..... [3] Born at **Castel - Novel** in Corr, **she** spent **her** childhood in the care of her English nanny, **Miss Draper**, only rarely seeing her famous mother. [4~7]... ... [8] **She** is buried next to **her mother** at **Lachaise** in Paris.

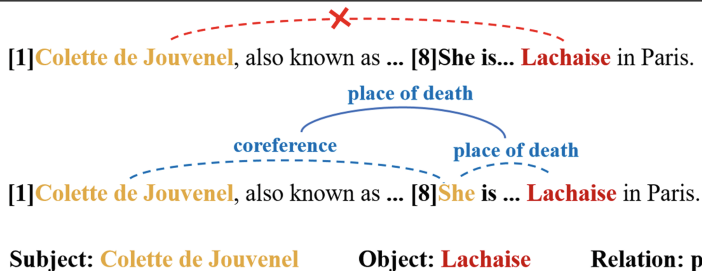


Fig. 1. An example document from DocRED. Entities are distinguished by color, with the reasoning clue and relation label listed offside.

especially produced by mentions, is extensively used for logical inference. However, the coreference information of pronouns, beneficial to obtaining interactive information across sentences [3] and multi-hop graph convolution, is ignored.

Figure 1 shows an example from the DocRED dataset [15]. As will be readily seen, only based on the fact that mention *Colette de Jouvenel* (in the 1st sentence) and pronouns *she* (in the 8th sentence) refer to the same entity, can we infer the relation of entity pair (*Colette de Jouvenel*, *Lachaise*) is the *place of death*. And the relational reasoning pattern of entity pairs (*Colette de Jouvenel*, *Castel–Novel*) and (*Colette*, *Lachaise*) is the same as above. Therefore, the pronouns in documents can produce rich semantic information, which is extremely vital to Doc-level RE. To verify the hypothesis, we randomly sample 100 documents from the DocRED training set and measure the number of pronouns and mention-pronoun pairs. Table 1 describes that each document has approximate 32 pronouns (“*he*”, “*him*”, “*his*”, “*she*”, “*her*”, etc.) and 14 mention-pronoun pairs. Obviously, pronouns can provide significant clues to Doc-level RE if some strategies are designed ingeniously.

Table 1. Statistics of pronouns and mention-pronoun pairs.

Type	Count
Pronouns	32
Mention-Pronoun pairs	14

To capture the feature produced by pronouns, we propose a novel Coref-aware Doc-level RE based on Graph Inference Network (CorefDRE). CorefDRE is a fine-tuned coreference-aware approach that instructs the model directly to learn the coreference information produced by mentions and pronouns. Specifically, we

propose a heterogeneous graph, **Mention-Pronoun Affinity Graph** (MPAG) with two types of nodes, namely mention node and pronoun node, as well as three types of edges (i.e., intra-sentence edge, intra-entity edge, and pronoun-mention edge) to capture semantic information of pronouns in the document for relation extraction. MPAG is a fusion of **Mention Graph**(MG) [20] and **Mention-Pronoun Graph**(MPG), which is constructed according to NeuralCoref, an extension to the Spacy. After that, we apply GCNs [6] to MPAG to get the representation for each mention and pronoun. Then we merge mentions and pronouns that refer to the same entity to get the Entity Graph (EG), and based on EG we infer multi-hop relations between entities. Meanwhile, to reduce the noise brought by NeuralCoref, we propose the noise suppression mechanism that first calculates the affinity of each mention-pronoun pair as edge weigh of MPAG and then suppresses the low weight edge during the fusion of MPAG into EG.

Our contributions are summarized as follows:

- We introduce a novel heterogeneous graph, Mention-Pronoun Affinity Graph (MPAG), which integrates the coreference information produced by mentions and pronouns to better adapt to Doc-level RE task.
- We propose a noise suppression mechanism to calculate the affinity between mention and corresponding pronoun for suppressing noise produced by false mention-pronoun pairs.
- We conduct experiments and the results outperform baseline by nearly 1.7–2.0 in F1 on the public datasets, DocRED, DialogRE, and MPDD, which demonstrate the effectiveness of our CorefDRE model.

This paper is organized as follows: Sect. 1 outlines the research on doc-level RE and the main contribution of this paper, Sect. 2 and section 3 detail the proposed model and the experimental results, respectively. Section 4 describes the related work of graph-based doc-RE and Sect. 5 summarizes the advantages and disadvantages of this paper and provides the direction for future research.

2 Proposed Approach

We formulate the task of document-level RE (doc-level RE) in the following way: **Document** D : the document is the raw text containing multiple sentences, namely $D = \{s_1, s_2, \dots, s_n\}$. **Entity** E : the entity set E consists of the entities that appear in the document. For each entity e_i , it is represented by a set of mentions in the document as well as an entity type: $e_i = (\{m_{i1}, m_{i2}, \dots\}, t_i)$, where $t_i \in R_e$ (the set of predefined entity types in the datasets). **Mention** m : the mentions refer to the representation of entities in a document, and each mention refers to a span of words: $m = \{w_1, w_2, \dots\}$. **Pronoun** p : pronouns are words that can refer to mention in a document (e.g., it, he, she, etc.).

Given the document D and entity set E , Doc-level RE is required to predict the relational facts between entities, namely $r_{s,o} = f(e_s, e_o)$, where e_s, e_o are subject entity and object entity in E , $r_{s,o}$ is a relational fact in pre-defined relation set R . In order to produce the above described output, our model, Coref-aware

Doc-level RE based on Graph Inference Network (CorefDRE), mainly consists of 3 modules: Mention-Pronoun Affinity Graph construction module (Subsect. 2.1), noise suppression mechanism (Subsect. 2.2), graph inference module (Subsect. 2.3), as is shown in Fig. 2.

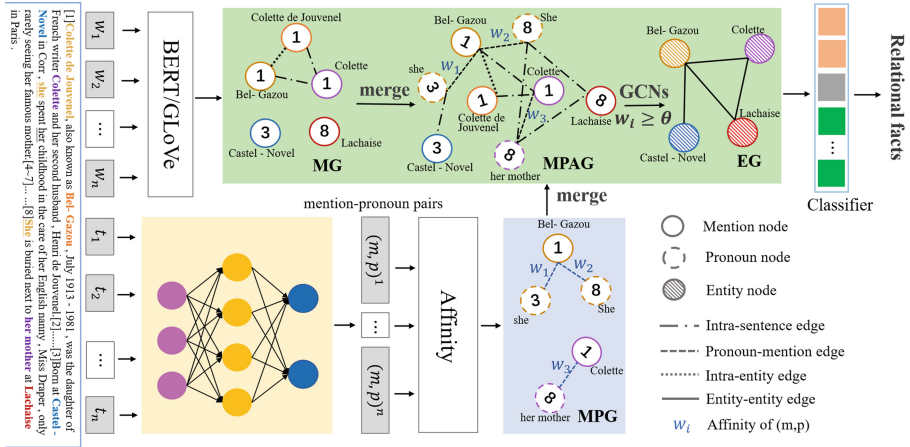


Fig. 2. The architecture of our CorefDRE. First, the document is fed into the encoder respectively, and then MG is constructed. Second, find the mention-pronoun pairs and use the noise suppression mechanism to calculates the affinity of mention-pronoun pairs as the weight of mention-pronoun edge, and then MPG is constructed. Third, merge MG and MPG to MPAG and mention-pronoun pairs with low affinity are inhibited when merging. Finally, after applying GCNs, MPAG is transformed into EG, where the paths between entities are identified for reasoning. Different entities are drawn with colors and the number i in each node denotes that it belongs to the i -th sentence.

2.1 Mention-Pronoun Affinity Graph Construction Module

To model the coreference relationships and enhanced the interactions between entities, Mention-Pronoun Affinity Graph (MPAG), a combination of MG and MPG, is constructed. MG is constructed according to Zeng et al. [20] but no document node here. MPG is constructed according to the mention-pronoun pairs generated by NeuralCoref. Specifically, the NeuralCoref first identify the pronouns that refer to the same mention and cluster the mention-pronoun pairs together as coreference clusters. For instance, we can obtain mention-pronoun pair clusters simply (e.g., $[(Bel\ Gazou, she), (Bel\ Gazou, She), \dots, (Colette, her\ mother)]$) from the sentences as shown in Fig. 2. And m and p in the pair (m, p) correspond a mention node m and a pronoun node p of MPG respectively, and there is a mention-pronoun edge between node m and node p .

There are two types of nodes and three types of edges in MPG:

Mention Node: each mention node in graph corresponds to a particular mention of an entity. The representation of the mention node m_i is defined

by the concatenation of semantic embedding, coreference embedding and type embedding [15], namely $m_i = [avg_{w_k \in m_i}(h_k); t_m; c_m]$, where $t_m \in R_e$, c_m represents which entity it refers to, and $avg_{w_k \in m_i}(h_k)$ is the average representation of mention containing words encoded by encoder.

Pronoun Node: Each pronoun (e.g. *it*, *his*, *she*, etc.) refers to a mention in the document corresponding to a pronoun node. The representation of the pronoun node is similar to that of the mention, where the *type embedding* and *coreference embedding* are the same as that of the corresponding mention nodes.

Intra-entity Edge: Nodes that refer to the same entity are fully connected with the intra-entity edge between them. The edge can model the interaction among different mentions and pronouns of the same entity and establish the interaction of cross-sentence.

Intra-sentence Edge: If two nodes co-occur in a single sentence, there is an intra-sentence edge between them. The edge can model the interaction among the mentions and pronouns referring to different entities.

Mention-Pronoun Edge: The mention-pronoun edge is the same as the mention-pronoun edge of MPG. The edge can strengthen the interaction of semantic information among sentences through the coreference information.

To initialize MPAG, we follow the GAIN proposed by Zeng et al. [20] and then dynamically update MPAG by applying Graph Convolution Network [6] to convolute the heterogeneous graph. Given node n_i at the l -th layer, the heterogeneous graph convolutional operation is formed as follows:

$$n_i^{l+1} = \sigma \left(\sum_{e \in E} \sum_{j \in N} \frac{1}{|N|} W_e^l n_j^l + b_e^l \right) \quad (1)$$

where $\sigma(\cdot)$ is the activation function. E denotes the set of different edges, N denotes the set of different neighbors of node n_i , and $W_e^l, b_e^l \in R^{d \times d}$ are trainable parameters.

To cover features of all levels, the node n_i is defined as the concatenation of hidden states of each layer:

$$n_i = [n_i^0; n_i^1; \dots; n_i^N] \quad (2)$$

where n_i^l is the representation of node n_i at layer l , and n_i^0 is the initial representation of node n_i , which is formed by the document representation from encoder.

2.2 Noise Suppression Mechanism

Mention-pronoun pairs will produce noise because of the weak adaptability between the datasets and the NeuralCoref. Therefore, we propose the noise suppression mechanism to filter noise in the process of graph inferencing (Subsect. 2.3). In our framework, we adopt the BERT to measure the affinity of the mention-pronoun pairs as the weight of the mention-pronoun edge. For each pair (*mention*, *pronoun*), we concatenate the context of *mention* and *pronoun*

as input and produce a single affinity scalar for every pair when constructing MPAG. The input form of tokens is as follows:

$$[CLS] \langle \text{Mention} \rangle [SEP] \langle \text{Pronoun} \rangle [SEP] \quad (3)$$

where $\langle \star \rangle := c_l[START] \star [END]c_r$

where \star is mention tokens or pronoun tokens, c_l and c_r represent the text on left and right of “ \star ” respectively. The $[START]$ and $[END]$ are two special tokens fine-tuned that indicate the start and end of “ \star ” in the context respectively.

Inspired by Angell et al. [1], we make affinity symmetric by averaging the representation of $(\text{mention}, \text{pronoun})$ and $(\text{pronoun}, \text{mention})$ to improve the representation. And then the affinity of the mention-pronoun pair is calculated by passing the enhanced representation of pairs into a linear layer with sigmoid activation. For instance, the affinity between mention-pronoun pair $(\text{mention}, \text{pronoun})$ is set 1, which is a strong signal for the fusion of MPAG. To calculate the affinity between the mention-pronoun pair accurately, we subtly design the positive sampling and negative sampling to train the affinity calculation. We screen out 300 positive samples D_p from the data D obtained by NeuralCoref and replace the mention m of the positive sample with other mentions m' randomly. To train the affinity module, we minimize the following triplet max-margin loss when training.

$$L_\varphi = \sum_{p_+, m \in P^+} \sum_{p_-, m \in P^-} l(m, p_+, p_-) \quad (4)$$

$$l(g, p, n) = [aff(g, n)^2 - (1 - aff(p, n))^2]_+ \quad (5)$$

where m and p are mention and pronoun in mention-pronoun pair (m, p) and $aff(m, p)$ is the affinity between m and p . The g, p, n in Eq. (5) are mention, negative pronoun, and positive pronoun referring to mention.

2.3 Graph Inference Module

Graph Merging. Inspired by Zeng et al. [20], we predict relational facts between entity pairs by reasoning on Entity Graph (EG), which is transformed from MPAG. Furthermore, the dynamic process of merging MPAG to EG is divided into three steps:

Step 1: Pronoun nodes that refer to the same mention are merged with the corresponding mention node to form a new mention node. Note that if the affinity between the mention-pronoun pair is less than the threshold θ , the pronoun does not participate in the merging process so that noise is depressed simply. For the i -th mention node merged from N pronoun nodes, it is represented by concatenating the mention and the average of its N pronoun node representations, and the representation of new mention node is defined as:

$$m_i = \bar{m}_i \oplus \frac{1}{N} \sum_n aff_n p_n (aff_n \geq \theta) \quad (6)$$

where \bar{m}_i denotes the mention representation. p_n is the n -th pronoun referred to the mention m_i , aff_n is the affinity of (\bar{m}_i, p_n) pair and \oplus denotes concatenate operation.

Step 2: Mention nodes that refer to the same entity are merged to an entity node in EG. For the i -th entity node merged from N mention nodes, it is represented by the average of its N mention node representation:

$$e_i = \frac{1}{N} \sum_n m_n \quad (7)$$

Step 3: Intuitively, intra-sentence edges between the mentions, which refer to the two entities, are merged as the bi-directional edge in EG. The directed edge from entity node e_i to e_j in EG is defined as:

$$edge_{ij} = \sigma(W_q[e_i; e_j]) + b_q \quad (8)$$

where W_q and b_q are trainable parameters and σ is an activation function (e.g., ReLU).

We model the potential reasoning clue between the entity nodes in EG through the path between the entity nodes. Based on the representation of the edge, *two-hop* path between entity nodes e_s and e_o is defined as:

$$P_{s,o}^k = [edge_{s,i}; edge_{i,o}; edge_{o,i}; edge_{i,s}] \quad (9)$$

where i stands for the intermediate node. Since there are multiple paths between two entity nodes, an attention mechanism is introduced to fuse the path information and pay more attention to the strong path. Path information of the entity in EG is defined as:

$$s_i = \sigma([e_s; e_o] \cdot W_l \cdot p_{s,o}^i) \quad (10)$$

$$\alpha_i = \frac{e^{s_i}}{\sum_j e^{s_j}} \quad (11)$$

$$p_{s,o} = \sum_i \alpha_i p_{s,o}^i \quad (12)$$

where α_i is the attention weight for i^{th} path and σ is an activation function (e.g., ReLU).

Relation Inference. According to the fusion of MPAG and noise suppression mechanism, The isomorphic graph EG is dynamically constructed, and the relationship between entity nodes can be predicted by the path inference. To identify the relationship of entity pair (e_s, e_o) , we concatenate the following representations as $I_{s,o}$ and compute the probability of relation r from the pre-specified relation schema as Eq. (14):

$$I_{s,o} = [e_s; e_t; |e_s - e_o|; e_s \odot e_o; p_{s,o}] \quad (13)$$

$$P(r | \mathbf{e}_s, \mathbf{e}_o) = \text{sigmoid}(W_b \sigma(W_a I_{s,o} + b_a) + b_b) \quad (14)$$

where e_s and e_o are the representation of subject and object entity in EG, and $p_{s,0}$ is the comprehensive inferential path information. W_a, W_b, b_a, b_b are trainable parameters and σ is an activation function (e.g., ReLU). We use binary cross entropy as the loss function to train our model.:

$$L = - \sum_{D \in S} \sum_{s \neq o} \sum_{r \in R} \text{CrossEntropy}(P(r | \mathbf{e}_s, \mathbf{e}_o), \bar{P}(r | \mathbf{e}_s, \mathbf{e}_o)) \quad (15)$$

where S denotes the whole corpus, and $\bar{P}(r | \mathbf{e}_s, \mathbf{e}_o)$ refers to ground truth.

3 Experiments

3.1 Dataset and Experimental Settings

DocRED [15]: DocRED consists of 3053 documents for training, 1000 for development and 1000 for test. And more than 40.7% of the relation facts require reasoning over multiple sentences. **DialogRE** [17]: DialogRE includes 1073 for training, 358 for development and 357 for test, and 95.6% of relational triples can be inferred through multiple sentences, where pronouns are extensively used. **MPDD** [2]: A publicly available Chinese dialogue dataset with the emotion and interpersonal relation labels and a mass of pronouns. To learn an effective representation for documents and capture the context of each mention, Following Yao et al. [15], for each word, we concatenate its word embedding, entity type embedding, and entity id embedding. And then we feed all the word representations into GloVe/BERT to get the representation of the document. We extract the relation between pronoun and mention based on Huggingface’s NeuralCoref and use BERT to pretrain the affinity for the mention-pronoun pair. We use 2 layers of GCN to encode the MPAG and EG. Our model is optimized with AdamW [9] and sets the dropout rate of GCN to 0.6, learning rate to 0.001.

3.2 Baseline Models

We use the following models as baselines.

CNN & BiLSTM: Yao et al. [15] proposed CNN and BiLSTM to encode the document into a sequence of the hidden state vectors. **Context-Aware**: Yao et al. [15] also proposed LSTM to encode the document and attention mechanism to fuse contextual information for predicting. **CorefBERT**: a pre-trained model was proposed by Ye et al. [16] for word embedding. **DocuNet-BERT**: Zhang et al. [21] proposed a U-shaped segmentation module to capture global information among relational triples. **GAIN-GloVe/GAIN-BERT**: Zeng et al. [20] proposed GAIN, which designed mention graph and entity graph to predict target relations, and make use of GloVe or BERT for word embedding, GCN for representation of the graph.

Table 2. Performance on DocRED. Models above the first double line do not use pre-trained models. Results with * are reported in their original papers. Ign F1 refers to excluding the relational facts shared by the training and dev/test sets.

Model	Dev		Test	
	Ign F1	F1	Ign F1	F1
CNN* [15]	41.58	43.45	40.33	42.46
BiLSTM* [15]	48.87	50.94	48.78	51.06
ConText-Aware* [15]	48.94	51.09	48.40	50.70
GAIN-GloVe [20]	53.12	55.37	52.71	55.18
CorefDRE-GloVe	55.01	57.33	54.37	56.74
BERT _{base} * [13]	-	54.16	-	53.20
CorefBERT _{base} * [16]	55.32	57.51	54.54	56.96
DocuNet-BERT _{base} * [21]	59.86	61.83	59.93	61.86
GAIN-BERT _{base} [20]	59.21	61.25	59.03	59.12
CorefDRE-BERT_{base}	60.85	63.06	60.78	60.82

Table 3. Performance on the datasets DialogRE and MPDD.

Model	F1-DialogRE	Acc-MPDD
CNN [17]	46.1	-
BERT [8]	60.6	31.0
GAIN [8]	69.8	42.2
CoIn [8]	71.1	46.5
CorefDRE	71.4	46.7

3.3 Main Results

We compare our CorefDRE model with other baselines on the DocRED dataset. The results are shown in Table 2. We use F1 and Ign F1 as evaluation indicators to evaluate the effect of models. Compared with the models based on GloVe, CorefDRE outperforms strong baselines by 1.7–2.0 F1 scores on the development set and test set. Compared with the models on BERT-base, CorefDRE outperforms strong baselines by 1.6–1.9. These results suggest that MPAG can capture the interaction relationship of multi-sentences for better Doc-level RE. Although we only conduct the experiments on DocRED, DialogRE, and MDPP shown in Table 3, our model is fit to others since pronouns are the essential grammar and syntax of the natural language.

3.4 Ablation Study

To verify the effectiveness of different modules in CorefDRE, we further analyze our model and the results of the ablation study shown in Table 4.

First, we remove the noise suppression mechanism. We set the weight of the mention-pronoun edge directly to 1 and merge all the pronoun nodes with

Table 4. Performance of CorefDRE with different embeddings and submodules.

Model	Dev		Test	
	Ign F1	F1	Ign F1	F1
CorefDRE-GloVe	55.01	57.33	54.37	56.74
-noise suppression	53.57	55.85	53.02	55.26
-pronoun node	53.12	55.37	52.71	55.18
CorefDRE-BERT _{base}	60.85	63.06	60.78	60.82
-noise suppression	59.57	61.67	59.43	59.64
-pronoun node	59.21	61.25	59.03	59.12

the corresponding mention node when generating EG. Without the weight between pronoun node and mention node, the performance of CorefDRE-GloVe/CorefDRE-BERT_{base} sharply drops by 1.39 F1 on the development set. This drop shows that the affinity between pronoun node and mention node plays a vital role in suppressing the noise caused by unsuitable mention-pronoun pairs.

Next, we remove the pronoun nodes. Specifically, we convert MPAG into MG proposed by Zeng et al. [20]. Without pronoun nodes, the result drops by an average of 1.88 F1 on the development set. This suggests that the pronoun nodes can capture richer information that mention node and document cannot capture effectively.

3.5 Case Study

Document Text	Model	Graph Inference	Prediction	Ground Truth
[0]... [2]Born in Victoria , Texas , Conrad Johnson was nine when his family moved to Houston. [3]Following studies at Yates High School , he attended Houston College for Negroes and graduated from Wiley College	GAIN		(Conrad Johnson, Wiley College, NA) ✗	(Conrad Johnson, Wiley College, Educated at)
	CorefDRE		(Conrad Johnson, Wiley College, Educated at) ✓	
[0] Samuel C. Brightman (1911 – 1992) was a journalist , war correspondent , freelance writer and adult educator. ... [3] He enlisted in the army in World War II and was assigned as a Public Relations Officer , responsible for transport, billeting , and general assistance for a group of newspaper reporters. ...	GAIN		(World War II, Public Relations Officer, NA) ✗	(World War II, Public Relations Officer, Conflict)
	CorefDRE		(World War II, Public Relations Officer, Conflict) ✓	

Fig. 3. Case Study on our CorefDRE model and baseline model. The models take the document as input and predict relations among different entities in different colors. The *Graph Inference* is reasoning process on graphs and the *NA* stands for no relation.

Figure 3 illustrates the case study of our CorefDRE compared with baseline. As is shown, GAIN can not predict the relation of entity pairs (*Conrad Johnson, Wiley College*) and (*Samuel C. Brightman, World War II*), while CorefDRE can predict the relation between *Conrad Johnson* and *Wiley College* is *educated at* and the relation between *Samuel C. Brightman* and *World War II* is *conflict*, because pronoun nodes *he* and *He* can connect the entity pair (*Conrad Johnson, Wiley College*) and (*Samuel C. Brightman, World War II*) respectively. We observe that relation extraction among those entities needs pronouns to connect them across sentences. The observation indicates that the information introduced by pronouns is beneficial to relation extraction.

4 Related Work

Relation Extraction is to extract relation facts from a given text, while early researches mainly focus on predicting relation fact between two entities within a sentence [4, 22]. These approaches include sequence-based methods, graph-based methods, and pre-training methods, which can tackle sentence-level RE effectively, and the dataset contains a fixed number of relation types and entity types. However, large amounts of real-world relational facts only can be extracted through multiple sentences.

Doc-Level Relation Extraction. Researchers extend sentence-level to Doc-level RE [12, 21] and explore two trends. The first is the sequence-based method that uses the pre-trained model to get the contextual representation of each word in a document, which directly uses the pre-trained model to obtain the relationship between entities [7, 10, 16]. These methods adopt transformers to model long-distance dependencies implicitly and get the entities embedding, and feed them into a classifier to get relation labels. However the sequence-based methods cannot capture enough entity interactions when the document length is out of the capability of the encoder at a time. In order to model these interactions, the graph-based method are proposed to constructs graphs according to documents, which can model entity structure more intuitively [11, 20, 23]. These methods take advantage of LSTM or BERT to encode the input documents and obtain the representation of entities, then utilize the GCNs to update representation, and finally feed them into the classifier to get relation labels.

Coreference Dependency Relation Reasoning. Some previous efforts on Doc-level RE introducing coreference dependency for multi-hop inference are useful for solving multi-hop reasoning. Previous works [10, 11, 23] have shown that graph-based coreference resolution is obviously beneficial to construct dependencies among mentions for relation reasoning. [19] proposed intra-and-inter-sentential reasoning based on R-GCN to model multiple paths by covering all cases of logical reasoning chains in the graph. [14] introduced a reconstructor to rebuild the graph reasoning paths to guide the relation inference by multiple reasoning skills including coreference and entity bridge. However, none of the above methods model the influence of pronouns on relation extraction and

reasoning directly. Our CorefDRE model deals with the problem by introducing a novel heterogeneous graph with mention-pronoun coreference resolution and noise suppression mechanism.

5 Conclusion and Future Work

This paper proposed CorefDRE which features two novel skills: the coref-aware heterogeneous graph, MPAG, and the noise suppression mechanism. Based on the proposed method, the model can extract document-level entity pair relation more effectively due to the richer semantic information brought by pronouns. Experiments demonstrated that our CorefDRE outperforms most previous models significantly and is orthogonal to pre-trained language models. However, there are still some problems not been completely solved. The noise generated by pronouns hinders the improvement of the performance of our model. In the future, we will explore other methods to construct less noisy mention-pronoun pairs to optimize CorefDRE.

Acknowledgements. The authors would like to thank the Associate Editor and anonymous reviewers for their valuable comments and suggestions. This work is funded in part by the National Natural Science Foundation of China under Grants No.62176029, and in part by the graduate research and innovation foundation of Chongqing, China under Grants No. CYB21063. This work also is supported in part by the National Key Research, Development Program of China under Grants 2017YFB1402400, Major Project of Chongqing Higher Education Teaching Reform Research (191003), and the New Engineering Research and Practice Project of the Ministry of Education (E-JSJRJ20201335).

References

1. Angell, R., Monath, N., Mohan, S., Yadav, N., McCallum, A.: Clustering-based inference for biomedical entity linking. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2021)
2. Chen, Y.T., Huang, H.H., Chen, H.H.: MPDD: a multi-party dialogue dataset for analysis of emotions and interpersonal relationships. In: Proceedings of the 12th Language Resources and Evaluation Conference, pp. 610–614 (2020)
3. Dasigi, P., Liu, N.F., Marasovi, A., Smith, N.A., Gardner, M.: Quoref: a reading comprehension dataset with questions requiring coreferential reasoning. 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) (2019)
4. Guo, Z., Nan, G., Lu, W., Cohen, S.B.: Learning latent forests for medical relation extraction. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, pp. 3651–3657 (2021)
5. Huang, H., Lei, M., Feng, C.: Graph-based reasoning model for multiple relation extraction. *Neurocomputing* **420**, 162–170 (2021)

6. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint [arXiv:1609.02907](https://arxiv.org/abs/1609.02907) (2017)
7. Li, Y., Song, Y., Jia, L., Gao, S., Li, Q., Qiu, M.: Intelligent fault diagnosis by fusing domain adversarial training and maximum mean discrepancy via ensemble learning. *IEEE Trans. Ind. Inform.* **17**(4), 2833–2841 (2021)
8. Long, X., Niu, S., Li, Y.: Consistent inference for dialogue relation extraction. In: *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (2021)
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: *ICLR* (2019)
10. Qiu, H., Zheng, Q., Msahli, M., Memmi, G., Qiu, M., Lu, J.: Topological graph convolutional network-based urban traffic flow and density prediction. *IEEE Trans. Intell. Trans. Syst.* **22**(7), 4560–4569 (2021)
11. Sahu, S.K., Christopoulou, F., Miwa, M., Ananiadou, S.: Inter-sentence relation extraction with document-level graph convolutional neural network. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4309–4316 (2019)
12. Wang, D., Hu, W., Cao, E., Sun, W.: Global-to-local neural networks for document-level relation extraction. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 3711–3721 (2020)
13. Wang, H., Focke, C., Sylvester, R., Mishra, N., Wang, W.: Fine-tune BERT for DocRED with two-step process. arXiv preprint [arXiv:1909.11898](https://arxiv.org/abs/1909.11898) (2019)
14. Xu, W., Chen, K., Zhao, T.: Discriminative reasoning for document-level relation extraction. arXiv preprint [arXiv:2106.01562](https://arxiv.org/abs/2106.01562) (2021)
15. 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)
16. 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)
17. Yu, D., Sun, K., Cardie, C., Yu, D.: Dialogue-based relation extraction. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4927–4940 (2020)
18. Yu, M., Yin, W., Hasan, K.S., dos Santos, C., Xiang, B., Zhou, B.: Improved neural relation detection for knowledge base question answering. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 571–581 (2017)
19. Zeng, S., Wu, Y., Chang, B.: SIRE: Separate intra-and inter-sentential reasoning for document-level relation extraction. arXiv preprint [arXiv:2106.01709](https://arxiv.org/abs/2106.01709) (2021)
20. 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)
21. Zhang, N., et al.: Document-level relation extraction as semantic segmentation. arXiv preprint [arXiv:2106.03618](https://arxiv.org/abs/2106.03618) (2021)
22. Zhang, Y., Zhong, V., Chen, D., Angeli, G., Manning, C.D.: Position-aware attention and supervised data improve slot filling. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 35–45 (2017)
23. Zhu, H., Lin, Y., Liu, Z., Fu, J., Chua, T.S., Sun, M.: Graph neural networks with generated parameters for relation extraction. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1331–1339 (2019)