



# CLGR-Net: a collaborative local-global reasoning network for document-level relation extraction

Xiaoyao Ding<sup>1</sup> · Gang Zhou<sup>1</sup> · Jicang Lu<sup>1</sup> · Taojie Zhu<sup>1</sup>

Accepted: 7 October 2022 / Published online: 24 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Document-level relation extraction aims to model the reasoning information over multiple sentences of a document and capture complex dependency interactions between inter-sentence entities. However, modeling reasoning information effectively in the document remains a challenging task. In this paper, we propose a Collaborative Local-Global Reasoning Network (CLGR-Net) for the Document-Level Relation Extraction model to effectively predict such relations by integrating rich local and global information from the multi-granularity graph. Specifically, CLGR-Net first constructs a mention-level graph and a concept-level graph. The former aggregates complex local interactions underlying the same entities, the latter captures long-distance global interaction among different entities. Finally, it creates an entity-level graph, the nodes and edges of the entity graph are aggregated by Relational Graph Convolutional Networks (R-GCN) and enriched by probability Knowledge Graphs (KGs), based on which we design a novel hybrid reasoning mechanism to collaborate relevant global and local information for entities. In this way, our model can effectively model reasoning information from these three graphs. The mention-level graph and concept-level graph are used as auxiliary information for the entity-level graph in the form of independent heterogeneous graphs. Our CLGR-Net model achieves more competitive performance than state-of-the-art on three widely used benchmarks.

**Keywords** Relation extraction · Document-level · Local and Global reasoning network · Multi-granularity graph

---

✉ Gang Zhou  
gzhouzhou@126.com

Xiaoyao Ding  
201420959449@stu.ncwu.edu.cn

Jicang Lu  
lujicang@sina.com

Taojie Zhu  
zhutaojie158@163.com

<sup>1</sup> PLA Strategic Support Force Information Engineering University, Zhengzhou 450000, China

# 1 Introduction

Relation extraction plays a significant role in various natural language processing applications, which aims to extract semantic relations between entities from the given text. Previous methods [1–4] have achieved remarkable success in a single sentence. In real-world applications, many relation instances appear in multiple sentences. Many recent studies [5–9] tackle the document-level relation extraction that recognizes relations between all entities in the entire document. Therefore, there is a more complex task for document-level relation extraction.

Document-level relation extraction requires more complex reasoning compared to sentence-level relation extraction. According to an analysis of the Wikipedia corpus [10], at least 61.1% of relations require complex reasoning skills to be extracted. Only 38.9% of relations can recognize simple patterns, which indicates that reasoning techniques play a crucial role in document-level relation extraction. For example, As Fig. 1 shows, in order to reason the relations between *Elizabeth II* in S1 and *United Kingdom* in S3, as well as *Commonwealth* in S3 and *Elizabeth II* in S1. First, the relation of entity pairs between (*United Kingdom*, *Commonwealth*) can easily be identified based solely on the same sentence. Next, we find that the inter-sentential relation between *Elizabeth II* and *Queen* is a mention of *Elizabeth II*. Obviously, to extract (*Elizabeth II*, *United Kingdom*) and (*Commonwealth*, *Elizabeth II*) relational facts, it is necessary to reason local and global context feature information.

Some previous relation extraction methods focus on an utterly certain assumption. However, deterministic knowledge graphs are inconsistent with real-world situations. We introduce probability KGs, such as ConceptNet and ProBase, to model related concepts for entity pairs that provide global reasoning information for related entity pairs. ConceptNet is a knowledge graph that expresses a core set of 36 relations between terms, such as PartOf, IsA and CapableOf. Probase is a

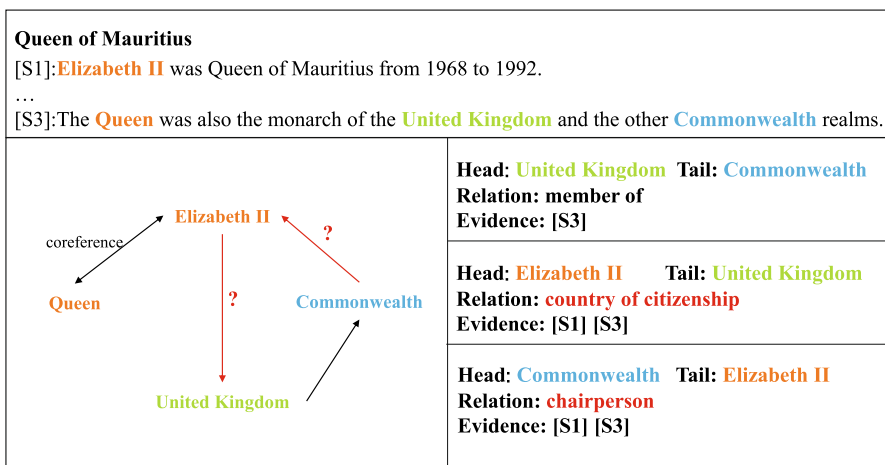


Fig. 1 A case of document-level relation extraction from the DocRED [10], we use black arrows to denote intra-sentential relations and red arrows to denote inter-sentential relations

universal probabilistic taxonomy automatically constructed from a corpus of 1.6 billion HTML texts that contains almost 2.7 million concepts. They use probabilities to model inconsistent, vague, and uncertain information it contains. Many of these ambiguous concepts are helpful in the coarse-grained understanding of document context.

This paper proposes a Collaborative Local-Global Reasoning Network (CLGR-Net), including a novel reasoning mechanism. In detail, we first construct two heterogeneous graphs: a mention-level graph and a concept-level graph. For the mention-level graph, we apply R-GCN [12] on the mention-level graph to get a local-aware representation for each word and mention. For the concept-level graph, we utilize probability KGs, such as ProBase [11] and ConceptNet [13], to aggregate entities effectively to obtain concept feature representations. Considering that different types of edges between two nodes in heterogeneous graphs should have different interaction strengths, we follow the basic idea of R-GCN but add a trainable parameter to the propagation model. Finally, the entity-level graph is constructed by collaborating a mention-level graph and a concept-level graph, integrating the local mentions interaction of entities and the global concepts knowledge of entities. We summarize our major contributions as follows:

- We propose an effective framework called CLGR-Net that captures local and global context interactions, to better address the long-distance dependency problems between entities.
- We propose a collaborative reasoning mechanism to fuse in concert with the entity reasoning block, to further improve the relational reasoning ability among entities.
- Our model achieves better performances than previous models on three public widely document-level relation extraction datasets, and further detailed analysis suggests that CLGR-Net can bring more applicable and reliable predictions.

## 2 Related work

### 2.1 Document-level relation extraction

Early approaches mainly focus on intra-sentence relation extraction [1–4]. However, these approaches do not consider large amounts of relations across mentions and ignore relations expressed across multiple sentences. Following Yao et al. [10], at least 40.7% of relations can only recognize the relation between two entities across sentence boundaries in a document. Therefore, many researchers have begun to explore document-level relation extraction in recent years. Most of them apply the graph-based models to encode heterogeneous graph structures. For example, Christophoulou et al. [5] first utilize heuristics to construct an edge-oriented graph for document-level relation extraction to generate different dependencies over the graph edges. Zeng et al. [14] proposed two heterogeneous graphs to model the document-level different interactions between entities and mentions. However, these

graph-based models are built upon the context of the document itself, which is different from our model that global interactions with the probability KGs.

## 2.2 Reasoning in relation extraction

Some studies also take reasoning into account by introducing an inference structure. Nan et al. [6] constructed a latent structure to perform relational reasoning dynamically. Li et al. [8] distinguished mentions to generate entity representations of mention-based reasoning. Tang et al. [15] proposed a hierarchical inference network to obtain document-level inference information. Wang et al. [16] presented a document-level RST-GRAPH and tackled the evidence reasoning module by introducing the Rhetorical Structure Theory (RST). The SSAN model proposed by Xu et al. [17] can predict entity relation through the interaction of context reasoning and structure reasoning. Zhou et al. [23] introduced two techniques of adaptive thresholding and localized context pooling to deal with multi-label and multi-entity problems. Li et al. [18] proposed a Multi-view Inference framework for relation extraction with Uncertain Knowledge (MIUK), which designed a multi-view reason mechanism that integrated local and global across mention-view, entity-view, and concept-view. In contrast, Li et al. [18] connected the three views in an intertwined way. Moreover, they do not strictly distinguish between local and global cross-view links explicitly, resulting in the impact of local mentions on entity views mixed with global information. Second, the processes of node construction are different. They do not conduct views node representation learning like R-GCN to produce latent feature reasoning on the constructed graph, only calculated by averaging the embeddings to represent nodes. These approaches make the nodes in the graph not contain adequate information, and the edges containing deep reasoning are not connected, failing to predict the relation between entities.

Compared to the MIUK model and other graph-based inference structures for document-level relation extraction, our model has many different external designs and internal principles. We structure a mention-level graph and a concept-level graph to interact with entity graphs using independent connections, respectively.

## 3 Collaborative local-global reasoning module

As is illustrated in Fig. 2, the overall architecture of CLGR-Net mainly consists of three modules: encoding module (Sect. 3.1), collaborative local-global reasoning module (Sect. 3.2), and relation classification model (Sect. 3.3). Specifically, in encoding module, we convert each word in the document into a vector through an encoder. In collaborative local-global reasoning module, we first use logsumexp pooling that can contain rich semantics to generate original mention nodes and construct a mention-level graph, and then abstract the semantics of entities to generate concept nodes and construct a concept-level graph. Finally, the nodes and edges in the entity-level graph are reasoned through R-GCN. In relation classification model, by connecting the global reasoning

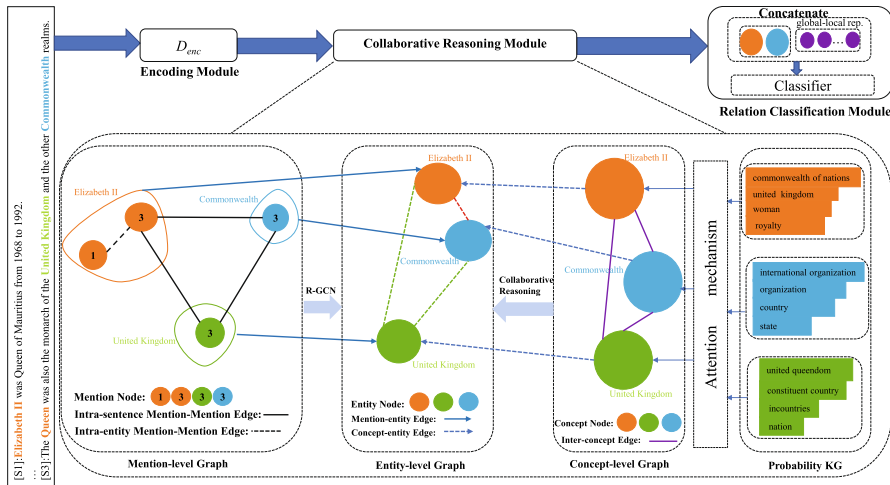


Fig. 2 The overall architecture of CLGR-Net

representation of the target entity and putting the results into a sigmoid function, the entities of relation extraction task are completed.

### 3.1 Encoding module

To represent each word in the document context, we first encode an input document containing  $n$  words  $\{w_i\}_{i=1}^n$  into a vector. In addition to the vectorization of the words themselves, we add two additional features to augment the input. One is type embedding, which is obtained by mapping the entity type of each mentioned word into a vector and has been proved to be useful for relation extraction [5, 10]. The other is coreference embedding, which is assigned words according to the entity to which they belong and help the model catch global interactive coreference information. Therefore, we concatenate each word  $w_i$  with its corresponding entity type embedding  $t_i$  and coreference embedding  $c_i$  to generate mixed input vectors  $x_i = [w_i; t_i; c_i]$ , where  $[\ ; \ ]$  is the concatenation operator. Finally, we feed mixed input vectors into an encoder  $D_{enc}$  to obtain contextualized embedding representation  $h_i$  for each word as follows:

$$h_i = D_{enc}(x_i), i \in [1, n] \tag{1}$$

where the  $D_{enc}$  can use bidirectional LSTM/BERT or other units.

### 3.2 Local-global representation module

#### 3.2.1 Local representation module

In this module, we focus on local representation modeling. Based on the contextual representation of each word, we extract mention nodes to construct a heterogeneous

mention-level graph. Unlike existing approaches, which constructed mention nodes by averaging or maximum pooling words that the mention contains [5, 8], we apply logsumexp pooling [18] to obtain the mention node. This pooling approach is similar to maximum pooling, but it can better accumulate signals from local contextual information, and it also shows better performance compared to averaging pooling in the experiment, the mention node is computed as the logsumexp pooling of the word representations associated with the mention:

$$m_i = \log \sum_{j=s_i}^{t_i} \exp(h_j) \quad (2)$$

where  $s_i$  and  $t_i$  are the start and end of the  $i$ -th mention, respectively.

To model the local context interactions, we treat mention nodes and construct the following two types of edges:

- *Intra-sentence mention-mention edges* Mentions are fully connected with intra-sentence mention-mention edges if the mentions co-occur in the same sentence.
- *Intra-entity mention-mention edges* Mentions are fully connected with intra-entity mention-mention edges if the mentions refer to the same entity.

### 3.2.2 Global representation module

Such uncertain representations with KGs critically capture the uncertain information of document relational fact and provide more rich global representations. Relations between entities can constitute the probabilistic semantic relations of vague concepts, specifically, by observing many individual entity pairs, the possibility of its corresponding concept can be to what extent determined according to a confidence score.

Inspired by Hao et al. [19], for each given entity, we first respectively mappings from the entity to all concepts based on the weight values in ConceptNet and ProBase. If the concept of an entity is not retrieved, the corresponding concept of the entity is marked as [unk], and its weight value is set to 0. Then we use the softmax function to normalize their weight values to obtain the attention weights of the concept corresponding to the entity. The corresponding attention weight values are denoted as  $\{\alpha_s\}_{s=1}^n$  and  $\{\gamma_t\}_{t=1}^n$  in ConceptNet and ProBase. In order to enrich the global reasoning information, we generate the concept vectors  $\mathbf{p}_i$  and  $\mathbf{b}_i$  of the corresponding entity  $e_i$  by introducing the attention mechanism [20] as follows:

$$\mathbf{p}_i = \sum_{s=1}^n \alpha_s \cdot e_i; \alpha_s = \frac{\exp(\alpha_s)}{\sum_{s=1}^n \exp(\alpha_s)} \quad (3)$$

$$\mathbf{b}_i = \sum_{t=1}^n \gamma_t \cdot e_i; \gamma_t = \frac{\exp(\gamma_t)}{\sum_{t=1}^n \exp(\gamma_t)} \quad (4)$$

In order to improve the understanding of attention weights in Eqs. 3 and 4, we visualize the concept of attention weights on relevant entities. As Fig. 3 shows, we

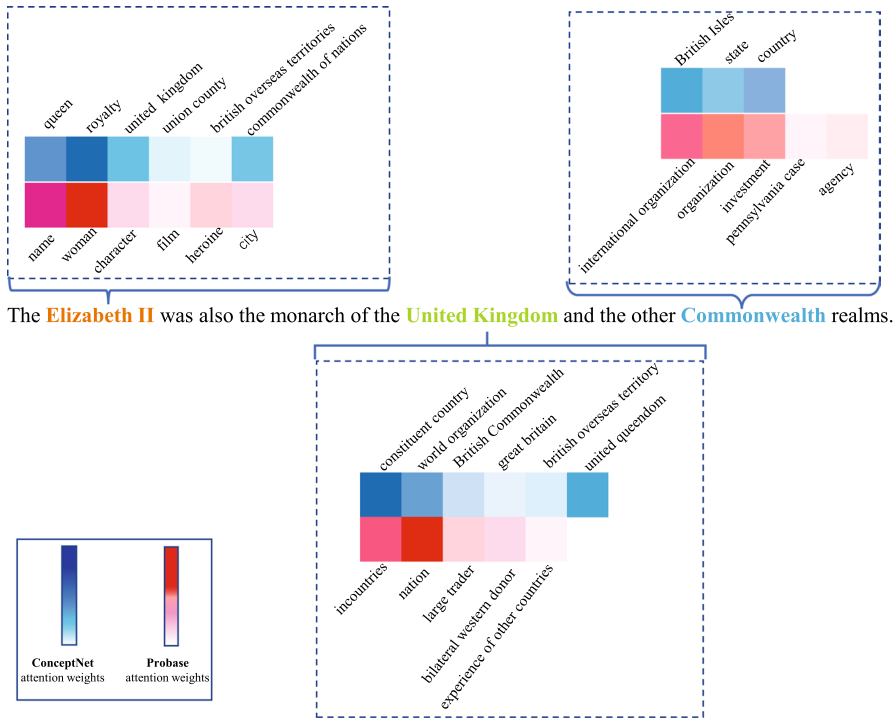
choose “*The Elizabeth II was also the monarch of the United Kingdom and the other Commonwealth realms.*” in the DocRED as an example, there are three entities in this sentence. Through the traversal of ConceptNet and ProBase, the concept corresponding to each entity is obtained, and Eq. 3 and Eq. 4 are used to calculate the concept attention weights on each entity. For example, the entity *Elizabeth II*, the most relevant concept to this entity, is *royalty* in ConceptNet, so the concept has the largest attention weight and the darkest color. Similarly, the most relevant concept to this entity is *woman* in Probase.

Finally, we concatenate  $\mathbf{p}_i$  and  $\mathbf{b}_i$  to obtain concept node representation of  $e_i$  as follows:

$$\mathbf{c}_i = [\mathbf{p}_i; \mathbf{b}_i] \tag{5}$$

To model global document interactions, we construct the Concept-level Graph by connecting concept nodes with the following type of edges:

*Inter-concept edges:* To capture non-local relation reasoning interactions among concepts, we connect all concept nodes.



**Fig. 3** Visualization of different attention weights, the deeper color denotes the higher weight, some concepts are not shown for brevity

### 3.2.3 Collaborative reasoning module

We construct Mention-level Graph (MG) in Sect. 3.2.1 and Concept-level Graph (CG) in Sect. 3.2.2. To further enhance interactions among entity pairs in a document, in this module, we construct Entity-level Graph (EG) and use R-GCN to perform collaborative local-global reasoning.

In this study, we propose rich local-global reasoning information for different edge types. More specifically, R-GCN can aggregate the features across different neighboring nodes. However, various types of edges usually have different dependency influences, they should be treated differential, so we assign unequal weight values  $\lambda_t^l$  by self-learning to different edge types  $\mathcal{T}$ . Formally, node  $i$  and its neighbors  $\mathcal{N}_i^t$  link with edge  $t$  at the  $l$ -th layer, and the modified graph convolutional operation generates transformed representation in the  $(l+1)$ th layer for node  $i$  via as follows:

$$m_i^{(l+1)} = \sigma \left( \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{N}_i^t} \frac{\lambda_t^l}{|\mathcal{N}_i^t|} W_t^l n_k^l + W_o^l m_i^l \right) \quad (6)$$

where  $\sigma$  denotes a bilinear function,  $\lambda_t^l$ ,  $W_t^l$  and  $W_o^l$  are learned parameter.

Next, in order to cover local mention node features of all levels, we concatenate the outputs of all R-GCN layers to generate the final representation of mention node as follows:

$$m_i = [m_i^0; m_i^1; m_i^2 \dots; m_i^N] \quad (7)$$

where  $m_i^0$  is the initial representation of mention node.

Finally similar to previous steps in mention representation, for each entity  $e_i$  with mentions  $\left\{ m_j^i \right\}_{j=1}^{E(e_i)}$ , where  $E(e_i)$  is entity  $e_i$  mentioned times, and we apply logsumexp pooling to define the entity node  $e_i$  as follows:

$$e_i = \log \sum_{j=1}^{E(e_i)} \exp(m_j) \quad (8)$$

We combine Mention-level Graph (MG) and Concept-level Graph (CG) to generate Entity-level Graph (EG). There are two types of edges in EG:

- *Mention-entity edges* To pass the mention-level reasoning message to the entity-level, mentions referring to the entity are fully connected.
- *Concept-entity edges* To pass the concept-level reasoning message to the entity-level, concepts referring to the entity are fully connected.

To enhance local-global reasoning, we first generate the interactive local-global representation  $\eta_{ij}$  for the target entities  $e_i$  and  $e_j$  as follows:

$$\eta_{ij} = \sigma([\mathbf{c}_i; e_i], [\mathbf{c}_j; e_j]) \quad (9)$$



The final local-global reasoning representation can be generated by concatenating the local-global interactive representation  $\eta_{ij}$  and a relative distance representations  $\delta_{ij}$  from the first mention of  $e_i$  to  $e_j$  as follows:

$$o_{ij} = [\eta_{ij}; \delta_{ij}] \quad (10)$$

### 3.3 Relation classification module

Based on the local-global reasoning module introduced above, we first concatenate entity pair representation and local-global reasoning representation to classify the relations, then we use a feedforward network with the sigmoid function to calculate the probability of each relation type  $r$ :

$$P(r | e_i, e_j) = \text{sigmoid}(W_r[o_{ij}; e_i; e_j] + b_r) \quad (11)$$

where  $[\ ; ]$  denotes concatenation and  $W_r$  and  $b_r$  are trainable parameters.

Finally, in our experiments, we use the binary cross-entropy(BCE) to train our model and the loss function as follows:

$$\mathcal{L} = - \sum_{r=1}^{\mathcal{R}} s_r \log(P_r) + (1 - s_r) \log(1 - P_r) \quad (12)$$

where  $s_r \in \{0, 1\}$  denotes the true value on relation label  $r$  and  $\mathcal{R}$  is the number of whole relations.

## 4 Experiments

### 4.1 Dataset and evaluation

We evaluate the effectiveness of our CLGR-Net on three document-level relation extraction datasets, including DocRED [10], CDR [21], and GDA [22]. DocRED is a large-scale human-annotated dataset based on Wikipedia texts. It consists of 132375 entities, 96 frequent relation types, and an “NA” (no relation) relation on the 5053 Wikipedia document. CDR and GDA focus on biomedical area datasets where CDR contains the binary relations between chemical and disease entities with 1500 documents, and GDA contains the binary interactions between gene and disease entities with 30192 documents.

For DocRED, following Yao et al. [10], we use F1 and Ign F1 as the evaluation indicators. Ign F1 is defined by excluding the common relation instances mentioned that exist in both training and dev/test sets. For CDR and GDA, considering that DocRED does not strictly annotate between intra-sentential and inter-sentential relations types but CDR and GDA do, following the previous work [5, 17], we utilize the intra-F1 and inter-F1 metrics to evaluate relation extraction performance respectively on dev set.

## 4.2 Baseline models

We compare our CLGR-Net with the following three types of baseline models:

- *Sequence-based Models* These models introduced different neural architectures to encode the input document, including CNN [10], LSTM [10], Context-Aware [10], bidirectional LSTM (BiLSTM) [10] and HIN [15].
- *Graph-based Models* These models constructed a document graph (homogeneous or heterogeneous graph) by connecting the given entities, including EoG [5], LSR [6], and GAIN [14].
- *BERT-based Models* Some models applied an excellent pre-trained model like BERT [25] to improve the performance of the document-level relation extraction model, including DISCO [16], MRN [8], ATLOP [23], HeterGSAN [7], MIUK [18], and SSAN [17]. Besides that, we also include the SciBERT [27] baseline models, SciBERT is also a pre-trained model based on BERT, which is pre-trained on scientific text.

## 4.3 Implementation details

In our CLGR-Net implementation, we use three word embedding methods. CLGR-BERT uses Uncased BERT-base (768d) [25] as encoder, for DocRED, CLGR-GloVe uses GloVe (100d) [28] embedding with BiLSTM (256d) as word embedding and encoder, for CDR and GDA, CLGR-SciBERT uses SciBERT-base (768d) as the encoder. AdamW [29] is used as our model optimizer, and weight decay is set to  $10^{-4}$ , learning rate to  $10^{-3}$  under PyTorch [30]. We tune all the hyperparameters on the development set and incorporate early stopping based on the best training epoch. All the experiments are run on NVIDIA GeForce GTX TITAN X GPU, with Intel (R) Xeon (R) E5-2620 v4 CPU.

# 5 Experimental results and analyses

## 5.1 DocRED results

Table 1 summarizes the results on DocRED. We observe that BERT-based models remarkably outperform sequence-based and graph-based baselines, and the best graph-based baseline model GAIN-Glove [14] outperforms the best sequence-based baseline model HIN-Glove [15]. It directly proves that BERT can take advantage of the document context interactions, and the graph-based structure compensates for BERT's weakness in capturing long-distance, cross-sentential information.

Moreover, we observe that our model generally outperforms the best graph-based and BERT-based models. In particular, our CLGR-GloVe (F1 56.68) and CLGR-BERT<sub>base</sub> (F1 62.87) achieve substantial improvements of 2.51% and 2.91% in F1 than the existing best GAIN-Glove model (F1 55.29) and ATLOP-BERT<sub>base</sub> model

**Table 1** Performances on DocRED

Groups	Methods	Dev		Test	
		Ign F1	F1	Ign F1	F1
Sequence-based Models	CNN [10]	41.58	43.45	40.33	42.26
	LSTM [10]	48.44	50.68	47.71	50.07
	BiLSTM [10]	48.87	50.94	48.78	51.06
	Context-Aware [10]	48.94	51.09	48.40	50.70
	HIN-GloVe [15]	51.06	52.95	51.15	53.30
	<b>CLGR-Glove</b>	<b>55.29</b>	<b>56.68</b>	<b>54.83</b>	<b>56.36</b>
Graph-based Models	EoG [5]	45.94	52.15	49.48	51.82
	LSR-Glove [6]	48.82	55.17	52.15	54.18
	GAIN-GloVe [14]	53.05	55.29	52.66	55.08
BERT-based Models	HIN-BERT <sub>base</sub> [15]	54.29	56.31	53.70	55.60
	DISCO-BERT <sub>base</sub> [16]	55.91	57.78	55.01	55.70
	SSAN-BERT <sub>base</sub> [17]	57.03	59.19	55.84	58.16
	HeterGSAN-BERT <sub>base</sub> [7]	58.13	60.18	57.12	59.45
	MIUK-BERT <sub>base</sub> [18]	58.27	60.11	58.05	59.99
	ATLOP-BERT <sub>base</sub> [23]	59.22	61.09	59.31	61.30
	<b>CLGR-BERT<sub>base</sub></b>	<b>61.53</b>	<b>62.87</b>	<b>61.14</b>	<b>62.41</b>

(F1 61.09) on the dev set, respectively, it is mainly due to emphasizing the local mention-level contextual reasoning information and global concept-level abstract structure in document-level relation extraction.

## 5.2 CDR and GDA results

Table 2 lists the results on the CDR and GDA, which besides the overall F1, Intra-F1, and Inter-F1. We apply the BERT and SciBERT pre-trained models on the test set, respectively. On the CDR test set, CLGR achieves +1.4 F1/+3.7 F1 gain based on BERT-based/SciBERT-based, which is better than existing state-of-the-art methods. On the GDA test set, the performance of CLGR is 0.83%/1.98% better than the best performing model SSAN-BERT<sub>base</sub>/ATLOP-SciBERT<sub>base</sub> based on BERT-based/SciBERT-based. These results again indicate the substantial effectiveness of CLGR.

## 5.3 Effect of probability knowledge graphs

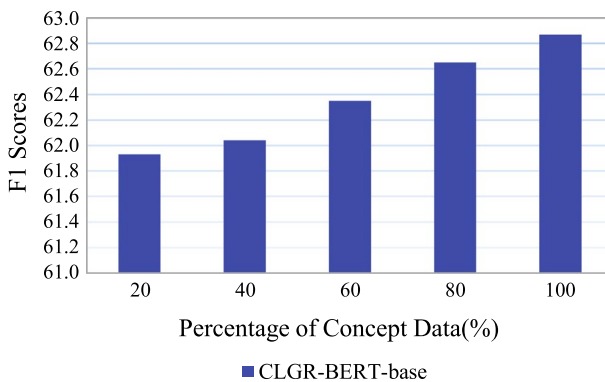
We further explore the effect of probability knowledge graphs. We randomly sample different amounts of concept data from ConceptNet and ProBase, and we consider different amounts of concept data settings ranging from 20–100%. Figure 4 shows the performance of CLGR-Net under the different amounts of concept data. It is observed that the F1 scores increase is slightly larger with more

**Table 2** Results on CDR test set and GDA test set

Dataset	Methods	F1	Intra-F1	Inter-F1
CDR	EoG [5]	63.6	68.2	50.9
	LSR [6]	64.8	68.9	53.1
	SSAN-BERT <sub>base</sub> [17]	62.7	70.4	44.7
	MRN-BERT <sub>base</sub> [8]	65.9	70.4	54.2
	<b>CLGR-BERT<sub>base</sub></b>	<b>67.3</b>	<b>72.8</b>	<b>54.4</b>
	SSAN-SciBERT <sub>base</sub> [17]	68.7	74.5	56.2
	ATLOP-SciBERT <sub>base</sub> [23]	69.4	-	-
	<b>CLGR-SciBERT<sub>base</sub></b>	<b>73.1</b>	<b>79.2</b>	<b>58.4</b>
GDA	EoG [5]	81.5	85.2	50.0
	LSR [6]	82.2	85.4	51.1
	MRN-BERT <sub>base</sub> [8]	82.9	86.1	53.5
	SSAN-BERT <sub>base</sub> [17]	83.4	86.7	62.3
	<b>CLGR-BERT<sub>base</sub></b>	<b>84.1</b>	<b>86.9</b>	<b>65.0</b>
	SSAN-SciBERT <sub>base</sub> [17]	83.7	86.6	65.3
	ATLOP-SciBERT <sub>base</sub> [23]	83.9	-	-
	<b>CLGR-SciBERT<sub>base</sub></b>	<b>85.6</b>	<b>88.7</b>	<b>67.1</b>

concept data. Specifically, when we use 100% of the concept data, the CLGR-Net model can achieve the highest F1 score of 62.87. This result indicates that probability knowledge graphs play a crucial role in our model.

How do ConceptNet and ProBase knowledge affect our model? In document-level relation extraction, when two entities cannot simply obtain the relations from a complex context, we can see the role of probability knowledge. Probability knowledge abstracts entities into higher-level concepts. Many of these concepts are crossed and related so that we can infer the relation between the two entities through the relation between the concepts corresponding to the two entities.

**Fig. 4** Performance comparison with the different amount of concept data on the DocRED dev set

## 5.4 Ablation study

To further analyze our method, we run an ablation experiment to study the effectiveness of each component of CLGR on the DocRED dev set. From Table 3, we find that:

- Without introducing concept node representations in CLGR, drop the result by 0.66/0.95. This shows that concept node representations can capture richer global interaction information in the long-distance non-local dependency relations, thereby enhancing the effect of entity relation extraction.
- When we replace the dynamic weighted R-GCN with the fixed weight R-GCN in node representation computing, F1 drops by 0.71 and 1.06, respectively. Compared with traditional R-GCN, our weighted R-GCN has a more vital ability to reason different weight edge types by self-learning.
- After removing interactive Local-Global reasoning, the performance sharply goes down by F1 for 1.26 and IgnF1 for 1.42, implying that collaborative local-global reasoning is required in order to obtain more practical inference information in heterogeneous graphs.

In order to more clearly show the impact of each part of the ablation experiment on the overall model, we give the following example, as shown in Fig. 5. When the model is not missing any part, CLGR-BERT<sub>base</sub> can predict six types of relations. When the model removes the concept node representations, it can predict five types. Since the concept of *James Poe* in the example cannot be accurately identified, the relation between *They Shoot Horses* and *James Poe* cannot be predicted. When the model removes the dynamic weighted R-GCN, the model cannot perceive the information of neighbor nodes, so it cannot predict the relation between the two entity pairs (*They Shoot Horses*, 1969) and (*They Shoot Horses*, *Sydney Pollack*). When the model removes the interactive Local-Global reasoning. In the example, the relation between two pairs of entities that require local and global reasoning cannot be predicted. In an entity pair (*They Shoot Horses*, *York*), it is necessary to infer that *Susannah York* and *York* are coreferences.

**Table 3** Results of ablation study

Setting	Dev	
	Ign F1	F1
CLGR-BERT <sub>base</sub>	61.53	62.87
Concept node representations	60.87 (↓0.66)	61.92 (↓0.95)
Dynamic weighted R-GCN	60.82 (↓0.71)	61.81 (↓1.06)
Interactive Local-Global reasoning	60.09 (↓1.26)	61.45 (↓1.42)

<p>[S1] <i>They Shoot Horses</i>, Don't They? [S2] is a 1969 American depression-era melodrama film based on Horace McCoy's 1935 novel of the same name and directed by <i>Sydney Pollack</i>. [S3] The screenplay was written by <i>James Poe</i> and Robert E.Thompson. ...[S5] It starts <i>Jane Fonda</i>, Michael Sarrazin, <i>Susannah York</i>, Red Buttons, Bruce Dern, Bonnie Bedelia and Gig Young. ... [S9] The film received nine nominations at the 42nd <i>Academy Awards</i> including: Best Director, Best Actress(for <i>Fonda</i>), <i>Best Supporting Actress</i>(for <i>York</i>), Best Adapted Screenplay and winning Best Supporting Actor(for Young).</p>	<p><b>CLGR-BERT<sub>base</sub>:</b></p>
	<p>(42nd Academy Awards, has part, <i>Best Supporting Actor</i>) (<i>Fonda</i>, award received, <i>Best Actress</i>) (<i>They Shoot Horses</i>, screenwriter, <i>James Poe</i>) (<i>They Shoot Horses</i>, publication date, 1969) (<i>They Shoot Horses</i>, director, <i>Sydney Pollack</i>) (<i>They Shoot Horses</i>, cast member, <i>York</i>)</p>
	<p><i>without</i> <b>Concept node representations:</b></p>
	<p>(42nd Academy Awards, has part, <i>Best Supporting Actor</i>) (<i>Fonda</i>, award received, <i>Best Actress</i>) (<i>They Shoot Horses</i>, publication date, 1969) (<i>They Shoot Horses</i>, director, <i>Sydney Pollack</i>) (<i>They Shoot Horses</i>, cast member, <i>York</i>)</p>
	<p><i>without</i> <b>Dynamic weighted R-GCN:</b></p>
	<p>(42nd Academy Awards, has part, <i>Best Supporting Actor</i>) (<i>Fonda</i>, award received, <i>Best Actress</i>) (<i>They Shoot Horses</i>, screenwriter, <i>James Poe</i>) (<i>They Shoot Horses</i>, cast member, <i>York</i>)</p>
	<p><i>without</i> <b>Interactive Local-Global reasoning:</b></p>
<p>(42nd Academy Awards, has part, <i>Best Supporting Actor</i>) (<i>Fonda</i>, award received, <i>Best Actress</i>) (<i>They Shoot Horses</i>, screenwriter, <i>James Poe</i>) (<i>They Shoot Horses</i>, publication date, 1969)</p>	

Fig. 5 An example from DocRED to illustrate the effect of each part on the model

### 5.5 Case study

Figures 6 and 7 show two cases study to further demonstrate the effectiveness of our proposed CLGR-Net model compared with the several baselines on DocRED and CDR, respectively.

As is shown in Fig. 6, we notice that both BERT<sub>base</sub>-RE and ATLOP-BERT<sub>base</sub> can successfully predict three entity pairs (*Gregorio Pacheco Leyes, Bolivia*), (*Gregorio Pacheco Leyes, Livi Livi*) and (*Livi Livi, Province of Potosi*) of relation on DocRED. However, due to a lack of supplementary knowledge, ATLOP-BERT<sub>base</sub> fails to predict the relation between *Bolivia* and *Province of Potosi*, while it deduces *Livi Livi* and *Bolivia* across sentences successfully. Our CLGR-Net can identify correct logical reasoning chains: *Livi Livi* → *Bolivia* → *Province of Potosi*. This case indicates that our CLGR-Net has better collaborative local-global reasoning ability.

As is shown in Fig. 7, the “chemical-induced disease” relation of four entity pairs (*ethambutol, Bilateral optic neuropathy*), (*isoniazid, Bilateral optic neuropathy*), (*ethambutol, scotoma*) and (*isoniazid, scotoma*) can be accurately predicted by SSAN-SciBERT<sub>base</sub> and ATOLP-SciBERT<sub>base</sub> models. However, when it comes to entity pair (*scotoma, Bilateral optic neuropathy*), it is necessary to identify that *Bilateral optic neuropathy* and *bilateral retrobulbar neuropathy* are coreferences, which requires the model to have the ability of cross-sentence coreference reasoning. Obviously, our model is better at this aspect.

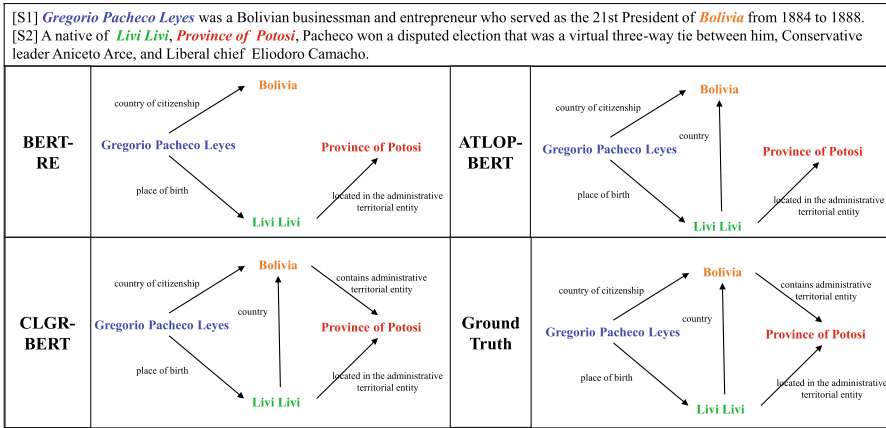


Fig. 6 A case study for our model and several baseline models on DocRED

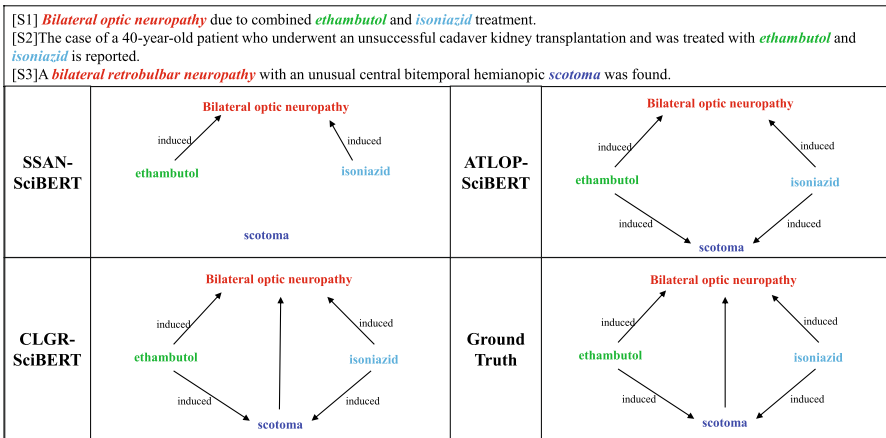


Fig. 7 A case study for our model and several baseline models on CDR

### 6 Conclusion

In this paper, we propose CLGR-Net, a collaborative local-global reasoning network for document-level relation extraction. We construct a mention-level graph and a concept-level graph to help alleviate long-distance dependency problems. Based on two heterogeneous graphs, we introduce a novel form of collaborative reasoning module and employ R-GCN to capture intrinsic clues and perform reasoning ability among entity pairs. Experimental results on three-wide datasets show that our CLGR-Net outperforms existing models.

**Acknowledgements** This work is supported by the Natural Science Foundation of Henan Province, China, under grant No. 222300420590. Also, we thank all the reviewers and editors for their feedback.

**Author Contributions** XD Conceptualization, Design, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review & editing, and Visualization. GZ Conceptualization, Writing—review & editing, and Supervision. JL Conceptualization, Writing—review & editing, and Supervision. TZ Conceptualization, Writing—review & editing, and Supervision.

**Funding** This study was financed in part by the Natural Science Foundation of Henan Province, China under grant No. 222300420590.

**Data Availability** The datasets used in the experiments are publicly available in the online repository.

## Declarations

**Conflict of interest** There are no conflicts or competing interests.

**Consent for publication** There is the consent of all authors.

## References

- Zhang Y, Qi P, Manning CD (2018) Graph convolution over pruned dependency trees improves relation extraction. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 2205–2215. <https://doi.org/10.18653/v1/d18-1244>
- Soares LB, FitzGerald N, Ling J, Kwiatkowski T (2019) Matching the blanks: distributional similarity for relation learning. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp 2895–2905. <https://doi.org/10.18653/v1/p19-1279>
- Alt C, Gabryszak A, Hennig L (2020) Probing linguistic features of sentence-level representations in neural relation extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pp 2895–2905. <https://doi.org/10.18653/v1/2020.acl-main.140>
- Hu X, Zhang C, Ma F, Liu C, Wen L, Yu PS (2021) Semi-supervised relation extraction via incremental meta self-training. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 1753–1762. <https://doi.org/10.18653/v1/2021.findings-emnlp.44>
- Christopoulou F, Miwa M, Ananiadou S (2019) 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 (EMNLP), pp 4925–4936. <https://doi.org/10.18653/v1/d19-1498>
- Nan G, Guo Z, Sekulić I, Lu W (2020) Reasoning with latent structure refinement for document-level relation extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), pp 1546–1557. <https://doi.org/10.18653/v1/2020.acl-main.141>
- Xu W, Chen K, Zhao T (2021) Document-level relation extraction with reconstruction. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI), pp 14167–14175
- Li J, Xu K, Li F, Fei H, Ren Y, Ji D (2021) MRN: A locally and globally mention-based reasoning network for document-level relation extraction. In: Findings of the Association for Computational Linguistics (ACL-IJCNLP), pp 1359–1370. <https://doi.org/10.18653/v1/2021.findings-acl.117>
- Yuan C, Huang H, Feng C, Shi G, Wei X (2021) Document-level relation extraction with entity-selection attention. *Inf Sci* 568:163–174. <https://doi.org/10.1016/j.ins.2021.04.007>
- Yao Y, Ye D, Li P, Han X, Lin Y, Liu Z, Sun M (2019) DocRED: A large-scale document-level relation extraction dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp 764–777. <https://doi.org/10.18653/v1/p19-1074>
- Wu W, Li H, Wang H, Zhu KQ (2012) Probase: a probabilistic taxonomy for text understanding. In: Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp 481–492. <https://doi.org/10.1145/2213836.2213891>
- Schlichtkrull M, Kipf TN, Bloem P, Berg RVD, Titov I, Welling M (2018) Modeling relational data with graph convolutional networks. In: European Semantic Web Conference (ESWC), pp 593–607. [https://doi.org/10.1007/978-3-319-93417-4\\_38](https://doi.org/10.1007/978-3-319-93417-4_38)
- Speer R, Chin J, Havasi C (2017) Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI), pp 4444–4451



14. Zeng S, Xu R, Chang, B, Li L (2020) 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. <https://doi.org/10.18653/v1/2020.emnlp-main.127>
15. Tang H, Cao Y, Zhang Z, Cao J, Fang F, Wang S, Yin P (2020) HIN: hierarchical inference network for document-level relation extraction. In: Proceedings of the 2019 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp 197–209. [https://doi.org/10.1007/978-3-030-47426-3\\_16](https://doi.org/10.1007/978-3-030-47426-3_16)
16. Wang H, Qin K, Lu G, Yin J, Zakari RY, Owusu JW (2021) Document-level relation extraction using evidence reasoning on RST-GRAPH. *Knowl-Based Syst* 228:107274. <https://doi.org/10.1016/j.knosys.2021.107274>
17. Xu B, Wang Q, Lyu Y, Zhu Y, Mao Z (2021) Entity structure within and throughout: modeling mention dependencies for document-level relation extraction. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI), pp 14149–14157
18. Li B, Ye W, Huang C, Zhang S (2021) Multi-view inference for relation extraction with uncertain knowledge. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI), pp 13234–13242
19. Hao J, Chen M, Yu W, Sun Y, Wang W (2019) Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp 1709–1719. <https://doi.org/10.1145/3292500.3330838>
20. Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: 3rd International Conference on Learning Representations (ICLR), <https://doi.org/10.48550/arXiv.1409.0473>
21. Li J, Sun Y, Johnson RJ, Sciaky D, Wei CH, Leaman R, Lu Z (2016) BioCreative V CDR task corpus: a resource for chemical disease relation extraction. *Database: J Biol Databases Curation*, <https://doi.org/10.1093/database/baw068>
22. Wu Y, Luo R, Leung H, Ting HF, Lam TW (2019) Renet: a deep learning approach for extracting gene-disease associations from literature. *Res Comput Mol Biol* 25:272–284. [https://doi.org/10.1007/978-3-030-17083-7\\_17](https://doi.org/10.1007/978-3-030-17083-7_17)
23. Zhou W, Huang K, Ma T, Huang J (2021) Document-level relation extraction with adaptive thresholding and localized context pooling. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), pp 14612–14620
24. Wang D, Hu W, Cao E, Sun W (2020) 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. <https://doi.org/10.18653/v1/2020.emnlp-main.303>
25. Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pp 4171–4186
26. Wang H, Focke C, Sylvester R, Mishra N, Wang W (2019) Fine-tune Bert for DocRED with two-step process. <http://arxiv.org/abs/1909.11898>
27. Beltagy I, Lo, K.; and Cohan, A. 2019. SciBERT: A pretrained language model for scientific text. 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 3615–3620. <https://doi.org/10.18653/v1/d19-1371>
28. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 1532–1543. <https://doi.org/10.3115/v1/d14-1162>
29. Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR), <https://doi.org/10.48550/arXiv:1711.05101>
30. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lerer A (2017) Automatic differentiation in pytorch. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). <https://doi.org/10.18653/v1/d18-1244>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.