

Document-level Biomedical Relation Extraction Based on Multi-Dimensional Fusion Information and Multi-Granularity Logical Reasoning

Lishuang Li* and Ruiyuan Lian† and Hongbin Lu and Jingyao Tang

School of Computer Science and Technology

Dalian University of Technology

Dalian, China

lils@dlut.edu.cn

Abstract

Document-level biomedical relation extraction (Bio-DocuRE) is an important branch of biomedical text mining that aims to automatically extract all relation facts from the biomedical text. Since there are a considerable number of relations in biomedical documents that need to be judged by other existing relations, logical reasoning has become a research hotspot in the past two years. However, current models with reasoning are single-granularity only based on one element information, ignoring the complementary fact of different granularity reasoning information. In addition, obtaining rich document information is a prerequisite for logical reasoning, but most of the previous models cannot sufficiently utilize document information, which limits the reasoning ability of the model. In this paper, we propose a novel Bio-DocuRE model called **FILR**, based on Multi-Dimensional Fusion Information and Multi-Granularity Logical Reasoning. Specifically, FILR presents a multi-dimensional information fusion module MDIF to extract sufficient global document information. Then FILR proposes a multi-granularity reasoning module MGLR to obtain rich inference information through the reasoning of both entity-pairs and mention-pairs. We evaluate our FILR model on two widely used biomedical corpora CDR and GDA¹. Experimental results show that FILR achieves state-of-the-art performance.

1 Introduction

Biomedical relation extraction (Bio-RE) is an important branch of biomedical text mining, which always draws researchers' attention. According to the different length of processing text, Bio-RE is generally divided into sentence-level RE and document-level RE. In sentence-level RE, all involved entities and relations are within a sentence.

*Corresponding author

†The author contributes equally to the first author in this work.

¹<https://github.com/Luguo-ry/FILR>

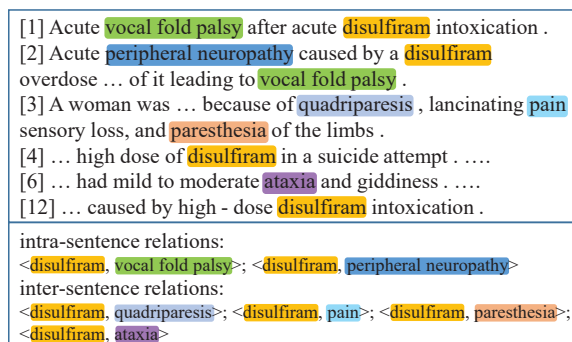


Figure 1: An example of Bio-DocuRE from the CDR dataset. We use the same color to denote the mentions of the same entity.

Document-level RE is more complex than its sentence counterpart. As shown in Fig. 1, in document-level biomedical relation extraction (Bio-DocuRE), the processed text is a whole document containing multiple sentences. “*disulfiram*” is a chemical and other entities are diseases. Each entity always has more than one mention, for example, “*disulfiram*” has four mentions in the document. In this example, there are two intra-sentence and four inter-sentence Chemical Induce Disease (CID) relations, and the inter-sentence relation means that the two entities involved in a relation may span multiple sentences or even the entire document.

Early research (Guo et al., 2015; Jiang et al., 2016; Zheng et al., 2017; Xiong et al., 2019) efforts focus on sentence-level RE. However, in real-world scenarios, many relations are expressed in multiple sentences, and the sentence-level models fail to recognize them. Therefore, more and more attention has been paid to DocuRE (Zhou et al., 2016; Gu et al., 2017; Verga et al., 2018). Most existed Bio-DocuRE studies are sequence-based (Nguyen and Verspoor, 2018; Li et al., 2020; Jiang et al., 2020), graph-based (Tran et al., 2020; Zhang et al., 2020a; Wang et al., 2020; Li et al., 2021b) or transformer-based (Liu et al., 2020; Zhou et al., 2021; Xu et al., 2021) methods, which never con-

sider the problem of relation reasoning and have limited effect. In the past two years, researchers noticed that there are a lot of relations in biomedical text that need to be judged relying on other relevant relations. So, Bio-DocuRE methods integrating logical reasoning have become a research hotspot (Li et al., 2021a; Christopoulou et al., 2019; Nan et al., 2020; Zhang et al., 2020b; Zeng et al., 2021; Zhao et al., 2022).

Mention, entity, mention-pair, and entity-pair are four document elements in different granularity, and we notice that the purpose of logical reasoning is essential to capture the dependencies between certain document elements. Consequently, we divide the logical reasoning mechanisms in Bio-DocuRE into the following four granularity: mention-based, entity-based, mention-pair-based, and entity-pair-based. All current methods with reasoning only consider single-granularity, and they perform mention-based (Li et al., 2021a), or entity-based (Christopoulou et al., 2019; Nan et al., 2020; Zhang et al., 2020b; Zeng et al., 2021), or entity-pair-based (Zhao et al., 2022) reasoning. Mention-based and entity-based reasoning mainly focuses on local token-level interactions and results in limited effect. While entity-pair-based reasoning focuses on global entity-relation interaction and could provide direct guidance for document-level relation classification. Compared with entity-pair-based reasoning, mention-pair-based reasoning can capture the global dependencies among all mention-pairs and provide fine-grained mention-relation reasoning information, which complements each other with the coarse-grained entity-relation reasoning information. Regrettably, no study notices the importance of mention-pair-based reasoning in Bio-DocuRE so far. Meanwhile, the current single-granularity methods leads to insufficient reasoning. Therefore, in order to obtain the effective and sufficient reasoning information, we propose a multi-granularity logical reasoning module with both mention-pair-based and entity-pair-based reasoning.

In addition, adequately mining global document information is an important step for Bio-DocuRE and is also the premise of performing logical reasoning. However, most of the current models suffer from the insufficient utilization of the document information because they either obtain the document information from a single perspective (Li et al., 2021a; Christopoulou et al., 2019; Nan

et al., 2020; Zhang et al., 2020b; Zeng et al., 2021; Zhao et al., 2022), or ignore the interaction and fusion of different document information (Zhang et al., 2020a; Wang et al., 2020; Xu et al., 2021). Both the global context and structural information of a document are important, which express the document from different views. Therefore, we extract both the global context and structural information of documents in this paper and propose a multi-dimensional information fusion module to fuse them.

Furthermore, in the process of obtaining the document global structural information, most current methods use GCN (Kipf and Welling, 2016) to iteratively update the document graph. However, GCN has been proved (Li et al., 2018; Luan et al., 2019; Zhao and Akoglu, 2019) to suffer from the over-smoothing problem when stacking the convolution layers, that is, the features of the nodes in graph would converge to similar values after deep iterations. Therefore, to relieve the GCN over-smoothing problem, we improve the commonly used document graph structure (Wang et al., 2020; Christopoulou et al., 2019; Li et al., 2021b) by designing a bridge node to capture the document global structural information more effectively within a finite number of iteration layers.

In this paper, to address the above issues, we present a novel Bio-DocuRE model called **FILR** based on multi-dimensional **F**usion **I**nformation and multi-granularity **L**ogical **R**easoning. We conduct experiments on two Bio-DocuRE datasets CDR (Li et al., 2016) and GDA (Wu et al., 2019). Our contributions can be summarized as follows:

- We propose a novel Bio-DocuRE model FILR. Experimental results on two benchmark datasets show that our model FILR can achieve state-of-the-art performance.
- We propose a multi-dimensional information fusion module called MDIF to adequately extract the global information of documents. Meanwhile, a bridge node is designed on the document graph to relieve the over-smoothing problem.
- We propose a multi-granularity logical reasoning module called MGLR to simultaneously conduct mention-pair-based and entity-pair-based reasoning.

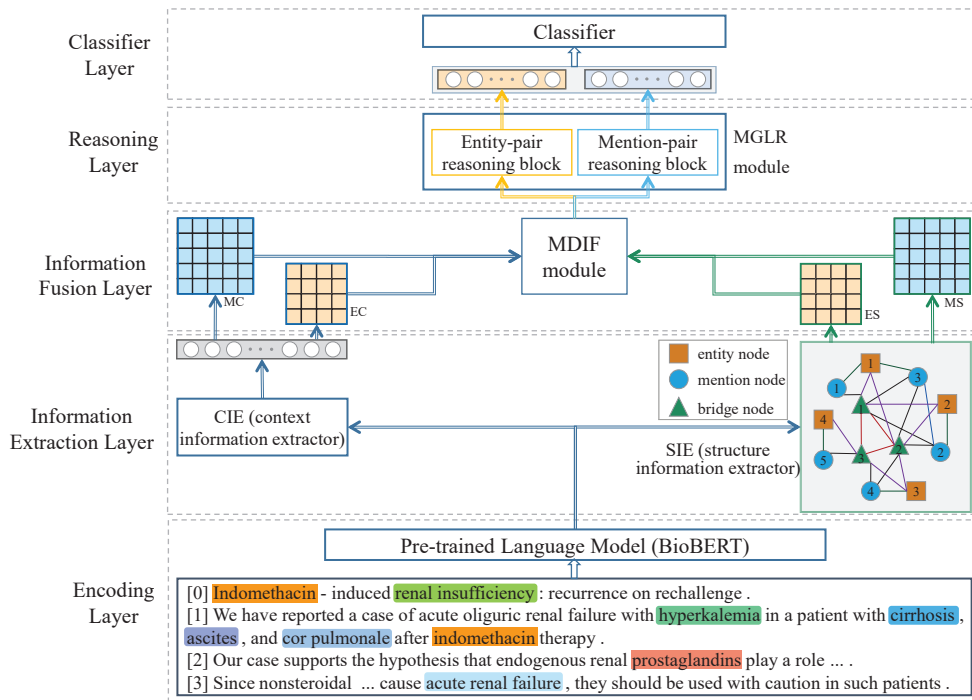


Figure 2: The overall framework of FILR, where MC means “mention-context map”, EC means “entity-context map”, ES means “entity-structure map” and MS means “mention-structure map”.

2 Model

In this section, we describe the overall framework of the proposed model FILR in Fig. 2. FILR makes an effective use of biomedical documents information and conducts multi-granularity reasoning to improve the performance of Bio-DocuRE.

2.1 Task Definition

Document-level biomedical relation extraction task can be defined as a classification problem. Given an annotated document T and a set of entities $e_i (1 \leq i \leq N_e)$, our goal is to extract the relations between all target entity pairs (e_h, e_t) . Note that in one document, an entity pair, also called a relation instance, usually owns more than one corresponding mention pairs because each entity e_i can occur multiple times by entity mentions $m_{ij} (1 \leq j \leq N_{e_i})$.

2.2 Encoding Layer

To better model the semantics of input document, FILR adopts BioBERT (Lee et al., 2020) as the document encoder, which is a domain-specific language representation model pre-trained on the large biomedical dataset for understanding complex biomedical texts. And its effectiveness has been proven surprisingly in various biomedical text mining tasks.

Let $D_T = [w_1, w_2, \dots, w_N]$ be the input document T with N tokens. To better represent the mentions which usually span a few consecutive tokens in T , we first insert a special marker “&” at the start and end of mentions to mark the mention’s span (Zhou et al., 2021). Then, the embedding representations of each token and marker “&” with rich semantic information can be got by the encoder. The embedding representation of T can be described by:

$$X_T = [x_1, x_2, \dots, x_i, \dots, x_{N'}], x_i \in R^{d_{emb}}, \quad (1)$$

where N' is the total number of tokens in T after inserting the marker, and d_{emb} denotes the embedding dim of words. In this paper, for each mention, we take the embedding of the start marker “&” as the mention embedding m_{ij} , and apply *logsumexp* (Jia et al., 2019) on the mentions referring to the same entity to obtain the entity embedding e_i :

$$e_i = \log \sum_{j=1}^{N_{e_i}} \exp(m_{ij}). \quad (2)$$

2.3 Information Extraction Layer

In this section, to obtain mention and entity representations that contain the global context and structural information of documents, we design two information extractors, respectively.

Context information extractor(CIE) Zhou et al. (2021) and Zhao et al. (2022) have proven the effectiveness of the pre-trained multi-head attention matrix $A \in R^{H \times N' \times N'}$ in aggregating the entity context information. So in this paper, we obtain the entity and mention representations with rich context information using similar methods. A_{kab} represents the attention from token a to token b in the k^{th} attention head. For mention m_{ij} , we first take the attention from start marker “&” to all tokens as the mention attention and get the mention-context attention matrix $A^{m_{ij}} \in R^{1 \times N'}$ by averaging all attention heads. For entity e_i , we get the entity-context attention matrix $A^{e_i} \in R^{1 \times N'}$ by averaging all related mention-context attention. Then we obtain the entity representations E_c and mention representations M_c with context information by:

$$A_1^m = \frac{A^m}{\mathbf{1}^T A^m}, A^m \in R^{N_m \times N'}, \quad (3)$$

$$M_c = A_1^m * X_T, M_c \in R^{N_m \times d_{emb}}, \quad (4)$$

$$A_1^e = \frac{A^e}{\mathbf{1}^T A^e}, A^e \in R^{N_e \times N'}, \quad (5)$$

$$E_c = A_1^e * X_T, E_c \in R^{N_e \times d_{emb}}, \quad (6)$$

where N_m is the total number of mentions in the document, A^m and A^e are the attention matrix of all mentions and entities, respectively.

Structure information extractor(SIE) In this part, we improve the commonly used document graph structure to avoid the over-smoothing problem of GCN. The main idea is to make the nodes in the document graph cover a wider information field in the initial state so that the model can capture the global structure information of the document before the over-smoothing problem occurs. So we construct a document-level graph with three types of nodes: entity nodes, mention nodes, and bridge nodes. The bridge node contains information across two sentences, so it broadens the information field and retains the structure of sentences in the meantime. And it is more conducive to the interaction and aggregation of long-distance nodes' information. With this document graph, the shallow layers of GCN can also well model the structure of documents, which alleviates the over-smoothing problem of GCN to some extent.

Specifically, we let $n_{b_i} = [b_i; t_b], n_{b_i} \in R^{d_{emb} + d_t}$ as the representation of bridge node b_i , where b_i is defined by averaging the representation of all tokens in the linked two sentences:

$b_i = avg_{x_k \in \{s_i, s_{i+1}\}}(x_k)$. Meanwhile, we let $n_{m_{ij}} = [m_{ij}; t_m]$ and $n_{e_i} = [e_i; t_e]$ as the representations of mention and entity nodes, respectively. $t_b, t_m, t_e \in R^{d_t}$ are type representations of nodes.

Then, we construct five types of edges: (1) Mention-Mention edges(MM): two mentions are connected with MM edge if they co-occur in a single sentence. (2) Mention-Bridge edges(MB): a mention node is connected to a bridge node with MB edge if this mention appears in the sentences linked with the bridge node. (3) Mention-Entity edge(ME): ME edge is added between a mention node and an entity node if the mention refers to the entity. (4) Entity-Bridge edge(EB): EB edge is added between an entity node and a bridge node if at least one mention of the entity is connected to the bridge node. (5) Bridge-Bridge edge(BB): BB edge is added between any two bridge nodes. In this paper, all edges are undirected.

After the document graph is constructed, R-GCN (Schlichtkrull et al., 2018) is applied on the graph to model the structure information of documents, and the iterative update process is defined as follows:

$$n_i^{l+1} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{|N_i^r|} W_r^l n_j^l + W_0^l n_i^l \right), \quad (7)$$

where $\sigma(\cdot)$ is an activation function, W_r^l, W_0^l are the parameters for edge type r and self-connection in l^{th} layer. N_i^r is the set of neighbouring nodes connected to n_i by edge type r . R is the set of edge types, and n_i^{l+1} is the i^{th} node representation resulted from the l^{th} R-GCN layer.

After L -layer R-GCN is stacked, mention representations $M_s \in R^{N_m \times d_{hid}}$ and entity representations $E_s \in R^{N_e \times d_{hid}}$ with structural information can be obtained, where d_{hid} is the hidden size of R-GCN, and d_{hid} is equal to the d_{emb} in this paper.

2.4 Information Fusion Layer

In this section, we propose a multi-dimensional information fusion module called MDIF to fuse the global context and structure information of documents.

First, four relation maps M_C (mention-context map), E_C (entity-context map), M_S (mention-structure map), E_S (entity-structure map) are built

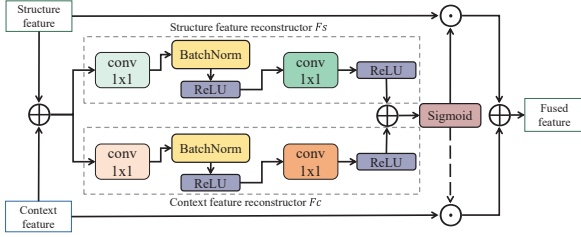


Figure 3: The structure of MDIF. \oplus denotes addition and \odot denotes the element-wise multiplication.

based on M_c , E_c , M_s and E_s as follows:

$$M_C = M_c \otimes M_c^T, M_C \in R^{d_{emb} \times N_m \times N_m}, \quad (8)$$

$$E_C = E_c \otimes E_c^T, E_C \in R^{d_{emb} \times N_e \times N_e}, \quad (9)$$

$$M_S = M_s \otimes M_s^T, M_S \in R^{d_{hid} \times N_m \times N_m}, \quad (10)$$

$$E_S = E_s \otimes E_s^T, E_S \in R^{d_{hid} \times N_e \times N_e}, \quad (11)$$

where \otimes is the multiplication by Einstein summation convention, $*^T$ is the matrix transpose.

Then MDIF is proposed for Bio-DocuRE inspired by Dai et al. (Dai et al., 2021) in computer vision. Our goal is to obtain entity-pair and mention-pair representations with global context and structural information.

The structure of MDIF is shown in Fig. 3. It receives the context and structure features as input. First, these two features are fused to obtain the shallow fusion feature. Then, the refined context and structure features are re-extracted with two similar information reconstructors. Both of them are mainly composed of two layers of point-wise convolution (PWConv), which only exploits point-wise channel interactions for the shallow fusion feature. Next, the refined structure and context features are fused again to obtain the deep fusion feature, and the channel-aware weight is calculated by a sigmoid function. Finally, the original structure and context features are fused by the channel-aware weight. The entity-pair representations E_{sc} and mention-pair representations M_{sc} with rich fusion information can be expressed as:

$$E_{sc} = E_S \odot W_{sc}^e + E_C \odot (1 - W_{sc}^e), \quad (12)$$

$$M_{sc} = M_S \odot W_{sc}^m + M_C \odot (1 - W_{sc}^m), \quad (13)$$

$$W_{sc}^e = \delta(F_s^e(E_S \odot E_C) + F_c^e(E_S \odot E_C)), \quad (14)$$

$$W_{sc}^m = \delta(F_s^m(M_S \odot M_C) + F_c^m(M_S \odot M_C)), \quad (15)$$

where $E_{sc} \in R^{d_{emb} \times N_e \times N_e}$, and $M_{sc} \in R^{d_{emb} \times N_m \times N_m}$. \odot denotes the element-wise multiplication. F_s^* and F_c^* are the structure and context feature reconstructors, respectively. $\delta(\cdot)$ is

the *sigmoid* function. W_{sc}^e and W_{sc}^m denote the channel-aware weights for entity-pair and mention-pair fusion, respectively. Entity-pairs and mention-pairs belong to different granularity of relation representations, so the parameters for entity-pairs and mention-pairs are independent and trained parallelly.

2.5 Reasoning Layer

In this section, we propose a multi-granularity logical reasoning module named MGLR based on CNN (Krizhevsky et al., 2012). There are two reasoning blocks in MGLR, one is the mention-pair reasoning block and the other is the entity-pair reasoning block. The structure of the two reasoning blocks is similar and the specific structure is shown in Fig. 4.

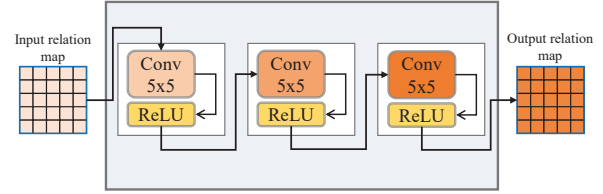


Figure 4: The structure of reasoning block.

The input of the mention-pair reasoning block is the mention-pair representations with rich global context and structure information M_{sc} . We apply the 2D convolution with kernel size 5×5 for reasoning, and use the ReLU activation function after each convolutional layer. In the mention-pair map, each mention-pair can interact with other pairs without the limitation of their position in the document, so the global interaction information of mention-pairs can be achieved only using one layer of convolution. While in the implementation process, we stack three convolution layers for deeper and more adequate fine-grained reasoning. We let M_{rs} as the representation of mention-relation reasoning information, it can be expressed as:

$$M_{rs} = Conv_n^M(M_{sc}), \quad (16)$$

where $Conv_n^M(\cdot)$ denotes the n -layers convolutional network for mention-pair-based reasoning and $M_{rs} \in R^{d_{emb} \times N_m \times N_m}$.

The structure of the entity-pair reasoning block is the same as counterpart mention-pair, and its input is the entity-pair representations with rich global context and structure information E_{sc} . The representation of entity-relation reasoning informa-

tion E_{rs} can be expressed as:

$$E_{rs} = \text{Conv}_n^E(E_{sc}), \quad (17)$$

where $\text{Conv}_n^E(\cdot)$ denotes the n -layers convolutional network for entity-pair-based reasoning and $E_{rs} \in R^{d_{emb} \times N_e \times N_e}$.

2.6 Classifier Layer

In this section, to classify the relations for a relation instance $r = (e_h, e_t)$, we need to get the representation of r .

First, let r_{rs}^e and r_{rs}^m as the entity-relation and mention-relation reasoning representation of r , respectively. r_{rs}^e is the representation of r in E_{rs} , and r_{rs}^m can be calculated by aggregating all mention-pairs referring to r :

$$r_{rs}^m = \text{avg}_{i \in e_h} (\text{avg}_{j \in e_t} (M_{rs[r;i;j]})). \quad (18)$$

Then, r_{rs}^e and r_{rs}^m are concatenated as the reasoning representation of r : $r_{rs} = [r_{rs}^e; r_{rs}^m]$, $r_{rs} \in R^{2d_{emb}}$.

Next, to enrich the relation representation, the entity-pair representation before reasoning and the entity representation obtained by R-GCN are also concatenated to represent the relation. The final relation representation can be expressed as:

$$r = [r_{rs}; r_{ht}; r_f], \quad (19)$$

where $r_{ht} = W_s[e_h^s; e_t^s] + b_s$, and $r_f \in R^{d_{emb}}$ denotes the entity-pair representation of r from E_{sc} . e_h^s and e_t^s are the entity representations of e_h and e_t from E_s . $W_s \in R^{2d_{emb} \times 2d_{emb}}$, and b_s are learnable parameters.

Finally, we use a feed-forward neural network to calculate the probability for each relation instance:

$$P(r|e_h, e_t) = \delta(W_b \sigma(W_a r + b_a) + b_b), \quad (20)$$

where W_* , b_* are learnable parameters, $\sigma(\cdot)$ is activation function, and $\delta(\cdot)$ is the *sigmoid* function. Global thresholding does not consider the variations of model confidence in different instances that results in suboptimal performance, so we adopt an adaptive-thresholding loss as the classification loss to train our model following Zhou et al. (2021). Specifically, they introduce a threshold class TH, and push the logits of all positive classes to be higher than the TH class, and the logits of negative

classes to be lower than it:

$$L = - \sum_{r \in P_T} \log \left(\frac{\exp(\text{logit}_r)}{\sum_{r' \in P_T \cup \{TH\}} \exp(\text{logit}_{r'})} \right) - \log \left(\frac{\exp(\text{logit}_{TH})}{\sum_{r' \in N_T \cup \{TH\}} \exp(\text{logit}_{r'})} \right), \quad (21)$$

where *logit* is the output in the last layer before *sigmoid* function.

3 Experiments

3.1 Datasets

We evaluate our FILT model on two widely used Bio-DocuRE datasets: CDR (Li et al., 2016) and GDA (Wu et al., 2019). The dataset statistics are shown in Table 1.

Table 1: The statistic results of CDR and GDA datasets.

Dataset		Docs	relations	N/A	Intra-	Inter-
CDR	Train	500	1038	4280	755	283
	Dev	500	1012	4136	766	246
	Test	500	1066	4270	763	303
GDA	Train	23353	36079	96399	30199	5880
	Dev	5839	8762	24362	7408	1354
	Test	1000	1502	3720	1273	229

3.2 Experimental Settings

FILR is developed using PyTorch and based on Huggingface’s Transformers (Wolf et al., 2020). The experimental settings are the same on CDR and GDA. We use BioBERT-Base v1.1 (Lee et al., 2020) as the encoder with learning rate 3e-5, and train FILR with learning rate 1e-4 using 3-layers of R-GCN. In FILR, in order to satisfy the requirement for matrix dimensions when performing reasoning, we set batch-max-entity-number and batch-max-mention-number for each batch. And all entity-pairs and mention-pairs representations of documents in the same batch are aligned with padding value 0.

3.3 Results

The baseline models can be divided into non-reasoning models and reasoning models. And each of them can also be divided into sequence-, graph-, or transformer-based methods according to the way of encoding documents.

Results on CDR Dataset. The performances of FILR and baseline models on CDR dataset are shown in Table 2. We can find that FILR achieves 85.7%, 89.1%, and 77.2% in terms of overall, intra-, and inter-F1, which outperforms all existing models and achieves a new state-of-the-art on

Table 2: Experimental results of the FILR on CDR and GDA datasets. All results given in this table are from their related papers. “-” means the data not given in original papers.

Model		CDR F1(%)			GDA F1(%)		
		Overall	Intra-	Inter-	Overall	Intra-	Inter-
non-reasoning model							
CNN+ME+PP(Gu et al., 2017) (Gu et al., 2017)	Sequence-based	61.3	57.2	11.7	-	-	-
BRAN(Verga et al., 2018) (Verga et al., 2018)	Sequence-based	62.1	-	-	-	-	-
EncAttAgg(Jiang et al., 2020) (Jiang et al., 2020)	Sequence-based	64.9	-	-	-	-	-
EoGANE(Tran et al., 2020) (Tran et al., 2020)	Graph-based	66.1	70.7	53.5	82.8	86.3	58.6
GLRE(Wang et al., 2020) (Wang et al., 2020)	Graph-based	68.5	-	-	-	-	-
DAM-GAN(Li et al., 2021) (Li et al., 2021b)	Graph-based	68.6	78.8	56.2	83.6	86.9	63.5
ATLOP(Zhou et al., 2021) (Zhou et al., 2021)	Transformer-based	69.4	-	-	83.9	-	-
SSAN(Xu et al., 2021) (Xu et al., 2021)	Transformer-based	68.7	74.5	56.2	83.7	86.6	65.3
reasoning model							
MRN(Li et al., 2021) (Li et al., 2021a)	Sequence-based	65.9	70.4	54.2	82.9	86.1	53.5
EoG(Christopoulou et al., 2019) (Christopoulou et al., 2019)	Graph-based	63.6	68.2	50.9	81.5	85.2	51.1
LSR(Nan et al., 2020) (Nan et al., 2020)	Graph-based	64.8	68.9	53.1	82.2	85.4	53.1
DHG(Zhang et al., 2020) (Zhang et al., 2020b)	Graph-based	65.9	70.1	54.6	83.1	85.6	58.8
SIRE(Zeng et al., 2021) (Zeng et al., 2021)	Graph-based	70.8	-	-	84.7	-	-
CGM2IR(Zhao et al., 2022) (Zhao et al., 2022)	Transformer-based	73.8	79.2	55.1	84.7	88.3	59.0
FILR(ours)		85.7	89.1	77.2	84.7	87.2	68.9

CDR dataset. Compared with the previous state-of-the-art model CGM2IR, FILR achieves improvements of 11.9%, 9.9%, and 22.1% in overall, intra- and inter-F1, respectively, which can be attributed to the reasoning of global mention-pairs and the utilization of global structure information of documents. In general, FILR provides rich premise information for logical reasoning through the extraction and fusion of global document information. Meanwhile, the multi-granularity logical reasoning based on entity-pair and mention-pair can effectively model the global interaction among different granularity relation pairs, improving the effect distinctly in all evaluation metrics, especially the inter-sentence.

Results on GDA Dataset. As shown in Table 2, we can see that FILR achieves 84.7% in overall F1, which outperforms most of the existing models and achieves the same result as the state-of-the-art models on GDA dataset. It is important to note that the scale of GDA is nearly 50 times larger than CDR, so GDA is less sensitive to models and the improvements on it in all evaluation metrics are not significant. Compared with CGM2IR, on intra-F1 and inter-F1, we can observe that FILR is 1.1% lower than CGM2IR in intra-F1, but 9.9% higher in inter-F1, which shows that FILR mainly improves the performance of inter-sentence relations. However, since the number of inter-sentence relations only accounts for 16% of the total relations shown in Table 1, the increase in inter-sentence F1 scores fails to bring the same gain in overall F1 scores.

3.4 Ablation Study

Table 3: Ablation study of FILR on CDR and GDA. “o-*” refers to the model removing “*” module. “o-context” and “o-structure” are the models without using context and structure information of documents, respectively. In “o-MDIF-cat” model, MDIF module is replaced with concat operation. “o-e-r” refers to the model removing entity-pair-based reasoning block.

Model	CDR F1(%)			GDA F1(%)		
	Overall	Intra-	Inter-	Overall	Intra-	Inter-
FILR	85.7	89.1	77.2	84.7	87.2	68.9
o-context	83.4	87.1	75.0	84.0	87.3	63.8
o-structure	84.6	88.8	74.8	83.7	86.8	63.6
o-MDIF-cat	83.6	88.4	73.5	84.2	87.2	66.7
o-e-r	79.1	83.2	70.3	83.4	86.7	63.1
o-MGLR	70.1	74.8	60.1	83.1	86.1	64.2

We conduct a thorough ablation study as shown in Table 3 and Fig. 5 to study the effectiveness of our contributions: MDIF, MGLR modules and the improved document graph. From Table 3 and Fig. 5, we can observe that:

(1) Both global context and structure document information are important for Bio-DocuRE. And the MDIF module is able to integrate the two global documents information effectively by the channel-based weighted fusion, which brings at least 2.1% and 0.5% improvements in terms of the CDR and GDA dataset, respectively.

(2) Coarse-grained entity-pair-based reasoning can get the macro entity-relation reasoning information, which is necessary for the relationship inference of long-distance entity-pairs. At the same time, fine-grained mention-pair-based reasoning as the beneficial supplement to entity-pair-based rea-

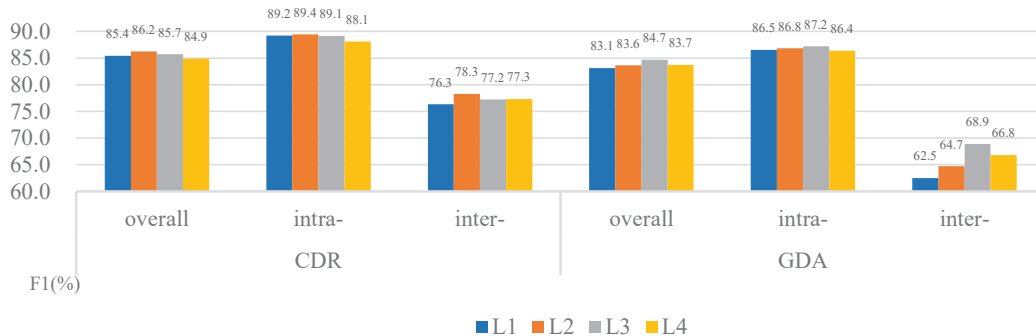


Figure 5: Ablation study of the effect of R-GCN layers.

soning, contributes evidently to both the intra- and inter-sentence relation extraction.

(3) As shown in Fig. 5, the best extraction results are achieved on CDR and GDA dataset with the R-GCN layers of 2 and 3, respectively. And with the increase of the number of RGCN layers, the decline of all evaluation metrics is relatively gentle. This result shows that the model finally achieves optimal performance after the convolution of two layers or three GCN layers, alleviating the over-smoothing problem as well as leveraging the advantages of GCN.

3.5 Case Study

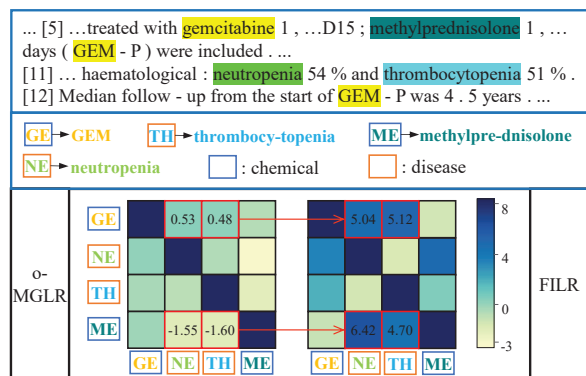


Figure 6: Case study of the multi-granularity reasoning module MGLR effectiveness on CDR dataset.

As shown in Fig. 6, we conduct a case study to further illustrate the effectiveness of the MGLR module. All values in Fig. 6 are the prediction scores of the given relation instances. In this example, *GEM* and *methylprednisolone* are chemicals, *neutropenia* and *thrombocytopenia* are diseases. The “o-MGLR” model incorrectly predicts the classification of relation instance $\langle \text{methylprednisolone}, \text{neutropenia} \rangle$ and $\langle \text{methylprednisolone}, \text{thrombocytopenia} \rangle$ with the prediction scores -1.55 and -1.6, respectively. And the prediction scores of $\langle \text{GEM}, \text{neutropenia} \rangle$ and $\langle \text{GEM}, \text{thrombocy-$

topenia} \rangle are 0.53 and 0.48, which indicates the model has insufficient confidence. Whereas, FILR predicts all relation instances correctly. We can see from Fig. 6, the confidence of FILR has been improved evidently with the help of MGLR module. In summary, the MGLR module can capture the global interactions among all entity-pairs and mention-pairs through convolution operation, even if there is no explicit connection between them, which effectively improves the performance of Bio-DocuRE.

4 Related Work

Non-reasoning methods. Previous non-reasoning Bio-DocuRE approaches can be divided into: sequence-based, graph-based, and transformer-based methods. Sequence-based methods (Gu et al., 2017; Verga et al., 2018; Jiang et al., 2020) used CNN or LSTM to extract document-level relations. However, these sequence-based methods do not capture the long-distance dependencies of documents and have limited effect. In order to break through the limitations, researchers tried to construct document graphs (Tran et al., 2020; Wang et al., 2020; Li et al., 2021b). Then, with the development of pre-trained language models, Transformer showed powerful capabilities in capturing context information, and is increasingly used in Bio-DocuRE task (Zhou et al., 2021; Xu et al., 2021).

Reasoning methods. Many researchers have tried to introduce various reasoning mechanisms to reinforce the ability of model in intra- and inter-sentence relation extraction. Li et al. (2021a) proposed a mention-based module for multi-hop reasoning, and introduced a co-attention mechanism for global mention reasoning. Christopoulou et al. (2019) implemented inference with an iterative algorithm. Nan et al. (2020) and Zhang et al. (2020b)

constructed a document graph and used GCN on the graph to employ multi-hop reasoning. Zeng et al. (2021) designed a logical reasoning module that can cover more logical reasoning chains. Zhao et al. (2022) executed entity-pair level logical reasoning using GNN-based methods.

5 Conclusion

In this paper, we propose a novel Bio-DocuRE model called FILR. FILR first acquires the global context and structure information of documents. Then, the MDIF module is proposed to fuse the two document information effectively. Next, we propose the MGLR module to conduct entity-pair-based and mention-pair-based reasoning parallelly. Last, we evaluate our FILR model on two widely used biomedical datasets CDR and GDA and the experimental results show that FILR achieves state-of-the-art on two datasets.

Acknowledgements

This work is supported by grant from the National Natural Science Foundation of China (No. 62076048), the Science and Technology Innovation Foundation of Dalian (2020JJ26GX035).

References

- Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4925–4936.
- Yimian Dai, Fabian Gieseke, Stefan Oehmcke, Yiquan Wu, and Kobus Barnard. 2021. Attentional feature fusion. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3560–3569.
- Jinghang Gu, Fuqing Sun, Longhua Qian, and Guodong Zhou. 2017. Chemical-induced disease relation extraction via convolutional neural network. *Database: The Journal of Biological Databases and Curation*, 2017:bax024.
- Xiyue Guo, Tingting He, and Jie Yuan. 2015. Relation dictionary construction and rule learning for ppi extraction from biomedical literatures. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1133–1140. IEEE.
- Robin Jia, Cliff Wong, and Hoifung Poon. 2019. Document-level n-ary relation extraction with multi-scale representation learning. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3693–3704.
- Ping Jiang, Xian-Ling Mao, Binbin Bian, and Heyan Huang. 2020. Improving document-level relation extraction via contextualizing mention representations and weighting mention pairs. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 305–312. IEEE.
- Zhenchao Jiang, Lishuang Li, and Degen Huang. 2016. A general protein-protein interaction extraction architecture based on word representation and feature selection. *International Journal of Data Mining and Bioinformatics*, 14(3):276–291.
- Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1106–1114.
- Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.
- Jiao Li, Yueping Sun, Robin J Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Thomas C Wieggers, and Zhiyong Lu. 2016. Biocreative v cdr task corpus: a resource for chemical disease relation extraction. *Database: The Journal of Biological Databases and Curation*, 2016:baw068.
- Jingye Li, Kang Xu, Fei Li, Hao Fei, Yafeng Ren, and Donghong Ji. 2021a. Mrn: A locally and globally mention-based reasoning network for document-level relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1359–1370.
- Lishuang Li, Ruiyuan Lian, and Hongbin Lu. 2021b. Document-level biomedical relation extraction with generative adversarial network and dual-attention multi-instance learning. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 438–443. IEEE.
- Qimai Li, Zhichao Han, and Xiao-Ming Wu. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 3538–3545. Association for the Advancement of Artificial Intelligence.
- Zhiheng Li, Zhihao Yang, Yang Xiang, Ling Luo, Yuanyuan Sun, and Hongfei Lin. 2020. Exploiting sequence labeling framework to extract document-level relations from biomedical texts. *BMC bioinformatics*, 21(1):1–14.

- Xiaofeng Liu, Jianye Fan, Shoubin Dong, et al. 2020. Document-level biomedical relation extraction leveraging pretrained self-attention structure and entity replacement: Algorithm and pretreatment method validation study. *JMIR Medical Informatics*, 8(5):e17644–e17644.
- Sitao Luan, Mingde Zhao, Xiao-Wen Chang, and Doina Precup. 2019. Break the ceiling: stronger multi-scale deep graph convolutional networks. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 10945–10955.
- Guoshun Nan, Zhijiang Guo, Ivan Sekulić, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1546–1557.
- Dat Quoc Nguyen and Karin Verspoor. 2018. Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings. *Proceedings of the BioNLP 2018 workshop*, pages 129–136.
- Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.
- Hieu Minh Tran, Minh Trung Nguyen, and Thien Huu Nguyen. 2020. The dots have their values: exploiting the node-edge connections in graph-based neural models for document-level relation extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4561–4567.
- Patrick Verga, Emma Strubell, and Andrew McCallum. 2018. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 872–884.
- Difeng Wang, Wei Hu, Ermei Cao, and Weijian Sun. 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)*, pages 3711–3721.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.
- Ye Wu, Ruibang Luo, Henry Leung, Hing-Fung Ting, and Tak-Wah Lam. 2019. Renet: A deep learning approach for extracting gene-disease associations from literature. In *International Conference on Research in Computational Molecular Biology*, pages 272–284. Springer.
- Wuti Xiong, Fei Li, Hong Yu, and Donghong Ji. 2019. Extracting drug-drug interactions with a dependency-based graph convolution neural network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 755–759. IEEE.
- Benfeng Xu, Quan Wang, Yajuan Lyu, Yong Zhu, and Zhendong Mao. 2021. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14149–14157.
- Shuang Zeng, Yuting Wu, and Baobao Chang. 2021. Sire: Separate intra-and inter-sentential reasoning for document-level relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 524–534.
- Jinyong Zhang, Weizhong Zhao, Jincan Yang, Xingpeng Jiang, and Tingting He. 2020a. An effective framework for document-level chemical-induced disease relation extraction via fine-grained interaction between contexts. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 391–396. IEEE.
- Zhenyu Zhang, Bowen Yu, Xiaobo Shu, Tingwen Liu, Hengzhu Tang, Wang Yubin, and Li Guo. 2020b. Document-level relation extraction with dual-tier heterogeneous graph. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1630–1641.
- Chao Zhao, Daojian Zeng, Lu Xu, and Jianhua Dai. 2022. Document-level relation extraction with context guided mention integration and inter-pair reasoning. *arXiv preprint arXiv:2201.04826*.
- Lingxiao Zhao and Leman Akoglu. 2019. Pairnorm: Tackling oversmoothing in gnns. *arXiv preprint arXiv:1909.12223*.
- Wei Zheng, Hongfei Lin, Ling Luo, Zhehuan Zhao, Zhengguang Li, Yijia Zhang, Zhihao Yang, and Jian Wang. 2017. An attention-based effective neural model for drug-drug interactions extraction. *BMC bioinformatics*, 18(1):1–11.
- Huiwei Zhou, Huijie Deng, Long Chen, Yunlong Yang, Chen Jia, and Degen Huang. 2016. Exploiting syntactic and semantics information for chemical–disease relation extraction. *Database: The Journal of Biological Databases and Curation*, 2016:baw048.
- Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-level relation extraction with adaptive thresholding and localized context pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14612–14620.