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 GDA †. 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. However, current 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., 2020; 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 Fusion Information and multi-granularity Logical Reasoning. 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.
Indomethacin-induced renal insufficiency: recurrence on rechallenge. We have reported a case of acute oliguric renal failure with hyperkalemia in a patient with cirrhosis, ascites, and cor pulmonale after indomethacin therapy. Our case supports the hypothesis that endogenous renal prostaglandins play a role... Since nonsteroidal... cause acute renal failure, they should be used with caution in such patients.

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 $(e_i, e_j)$. 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 $D_T = [w_1, w_2, ..., w_N]$ 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:

$$X_T = [x_1, x_2, ..., x_i, ..., x_{N'}], x_i \in \mathbb{R}^{d_{emb}},$$

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 $e_i$:

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

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 $A \in \mathbb{R}^{H \times N' \times N'}$ 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 $A^{m}_{i} \in \mathbb{R}^{1 \times N'}$ by averaging all attention heads. For entity $e$, we get the entity-context attention matrix $A^{e}_{i} \in \mathbb{R}^{1 \times N'}$ by averaging all related mention-context attention. Then we obtain the entity representations $E_{c}$ and mention representations $M_{c}$ with context information by:

$$A^{m}_{i} = \frac{A^{m}}{\sum_{l} A^{m}}, A^{m} \in \mathbb{R}^{N_{m} \times N'}, \quad (3)$$

$$M_{c} = A^{m}_{i} \cdot X_{T}, M_{c} \in \mathbb{R}^{N_{m} \times d_{emb}}, \quad (4)$$

$$A^{e}_{i} = \frac{A^{e}}{\sum_{l} A^{e}}, A^{e} \in \mathbb{R}^{N_{e} \times N'}, \quad (5)$$

$$E_{c} = A^{e}_{i} \cdot X_{T}, E_{c} \in \mathbb{R}^{N_{e} \times d_{emb}}, \quad (6)$$

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 $n_{b_{i}} = [b_{i}; t_{b}], n_{b_{i}} \in \mathbb{R}^{d_{emb} + d_{l}}$ 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:

$$b_{i} = \frac{1}{|\{x_{k} \in \{s_{i}, s_{i+1}\}\}|} \sum_{x_{k} \in \{s_{i}, s_{i+1}\}} x_{k}. \quad \text{Meanwhile, we let} \quad n_{m_{ij}} = [m_{ij}; t_{m}] \quad \text{and} \quad n_{e} = [e; t_{e}] \quad \text{as the representations of mention and entity nodes, respectively.}$$

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:

$$n_{i}^{l+1} = \sigma \left( \sum_{r \in R} \sum_{j \in N_{r}^{*}} \frac{1}{|N_{r}^{*}|} W_{r}^{l} n_{j} + W_{0}^{l+1} \right), \quad (7)$$

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. $N_{r}^{*}$ 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 $l^{th}$ node representation resulted from the $l^{th}$ R-GCN layer.

After $L$-layer R-GCN is stacked, mention representations $M_{s} \in \mathbb{R}^{N_{m} \times d_{hid}}$ and entity representations $E_{s} \in \mathbb{R}^{N_{e} \times d_{hid}}$ 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
Context feature and structure features are fused by the channel-aware mechanism. The channel-aware weight is calculated by a sigmoid function. $W^e_{sc}$ and $W^m_{sc}$ 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 3: The structure of MDIF. \( \oplus \) denotes addition and \( \otimes \) denotes the element-wise multiplication.](image)

![Figure 4: The structure of reasoning block.](image)
where \( E_{rs} = \text{Conv}^E_n(E_{sc}) \),

\[ (17) \]

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

### 2.6 Classifier Layer

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

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

\[ r^m_{rs} = \frac{1}{|E|} \sum_{e \in E_h} \left( \frac{1}{|E|} \sum_{e' \in E_l} (M_{rs}[r; e'; e]) \right). \]

Then, \( r^e_{rs} \) and \( r^m_{rs} \) are concatenated as the reasoning representation of \( r \):

\[ r_{rs} = [r^e_{rs}; r^m_{rs}], \quad r_{rs} \in \mathbb{R}^{2d_{emb}}. \]

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:

\[ r = [r_{rs}; r_{ht}; r_f], \]

where \( r_{ht} = W_e [e_h^e; e_l^e] + b_e \) and \( r_f \in \mathbb{R}^{d_{emb}} \) denotes the entity-pair representation of \( r \) from \( E_{sc} \). \( e_h^e \) and \( e_l^e \) are the entity representations of \( e_h \) and \( e_l \) from \( E_s \). \( W_e \in \mathbb{R}^{d_{emb} \times 2d_{emb}} \), and \( b_e \) are learnable parameters.

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

\[ P(r|e_h, e_l) = \delta(W_b \sigma(W_a r + b_a) + b_b), \]

where \( W_a, b_a \) 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:

\[ L = -\sum_{r \in P_T} \log \left( \frac{\exp(\logit_r)}{\sum_{r' \in P_T \cup \{\text{TH}\}} \exp(\logit_r')} \right) \]

\[ -\log \left( \frac{\exp(\logit_{r_{TH}})}{\sum_{r' \in P_T \cup \{\text{TH}\}} \exp(\logit_{r'})} \right), \]

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>
<thead>
<tr>
<th>Dataset</th>
<th>Docs</th>
<th>relations</th>
<th>N/A</th>
<th>Intra-</th>
<th>Inter-</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR</td>
<td>Train</td>
<td>500</td>
<td>1038</td>
<td>4280</td>
<td>30199</td>
</tr>
<tr>
<td></td>
<td>Dev</td>
<td>500</td>
<td>1012</td>
<td>4136</td>
<td>766</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>500</td>
<td>1066</td>
<td>4270</td>
<td>763</td>
</tr>
<tr>
<td>GDA</td>
<td>Train</td>
<td>23353</td>
<td>36079</td>
<td>96299</td>
<td>30199</td>
</tr>
<tr>
<td></td>
<td>Dev</td>
<td>5839</td>
<td>8762</td>
<td>24362</td>
<td>7408</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>1000</td>
<td>1502</td>
<td>3720</td>
<td>1273</td>
</tr>
</tbody>
</table>

#### 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.

<table>
<thead>
<tr>
<th>Model</th>
<th>CDR F1(%)</th>
<th>GDA F1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Intra-</td>
</tr>
<tr>
<td>non-reasoning model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN+ME+PP (Gu et al., 2017)</td>
<td>61.3</td>
<td>57.2</td>
</tr>
<tr>
<td>BRAN (Verga et al., 2018) (Verga et al., 2018)</td>
<td>62.1</td>
<td>-</td>
</tr>
<tr>
<td>EncAttAgg (Jiang et al., 2020)</td>
<td>64.9</td>
<td>-</td>
</tr>
<tr>
<td>EgGANE (Tran et al., 2020) (Tran et al., 2020)</td>
<td>66.1</td>
<td>70.7</td>
</tr>
<tr>
<td>GLRE (Wang et al., 2020) (Wang et al., 2020)</td>
<td>68.5</td>
<td>-</td>
</tr>
<tr>
<td>DAM-GAN (Li et al., 2021) (Li et al., 2021b)</td>
<td>68.6</td>
<td>78.8</td>
</tr>
<tr>
<td>ATLOP (Zhou et al., 2021) (Zhou et al., 2021)</td>
<td>69.4</td>
<td>-</td>
</tr>
<tr>
<td>SSAN (Xu et al., 2021) (Xu et al., 2021)</td>
<td>68.7</td>
<td>74.5</td>
</tr>
<tr>
<td>reasoning model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRN (Li et al., 2021) (Li et al., 2021a)</td>
<td>65.9</td>
<td>70.4</td>
</tr>
<tr>
<td>EgG (Christopoulou et al., 2019) (Christopoulou et al., 2019)</td>
<td>63.6</td>
<td>68.2</td>
</tr>
<tr>
<td>LSR (Nan et al., 2020) (Nan et al., 2020)</td>
<td>64.8</td>
<td>68.9</td>
</tr>
<tr>
<td>DHG (Zhang et al., 2020) (Zhang et al., 2020b)</td>
<td>65.9</td>
<td>70.1</td>
</tr>
<tr>
<td>SIRE (Zeng et al., 2021) (Zeng et al., 2021)</td>
<td>70.8</td>
<td>-</td>
</tr>
<tr>
<td>CGM2IR (Zhao et al., 2022) (Zhao et al., 2022)</td>
<td>73.8</td>
<td>79.2</td>
</tr>
<tr>
<td>FILR (ours)</td>
<td>85.7</td>
<td>89.1</td>
</tr>
</tbody>
</table>

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-MDF-cat” model, MDF module is replaced with concat operation. “o-e-r” refers to the model removing entity-pair-based reasoning block.

<table>
<thead>
<tr>
<th>Model</th>
<th>CDR F1(%)</th>
<th>GDA F1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Intra-</td>
</tr>
<tr>
<td>FILR</td>
<td>85.7</td>
<td>89.1</td>
</tr>
<tr>
<td>o-context</td>
<td>81.4</td>
<td>87.1</td>
</tr>
<tr>
<td>o-structure</td>
<td>84.6</td>
<td>88.8</td>
</tr>
<tr>
<td>o-MDF-cat</td>
<td>83.6</td>
<td>88.4</td>
</tr>
<tr>
<td>o-e-r</td>
<td>79.1</td>
<td>85.2</td>
</tr>
<tr>
<td>o-MGLR</td>
<td>70.1</td>
<td>74.8</td>
</tr>
</tbody>
</table>

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.

We conduct a thorough ablation study as shown in Table 3 and Fig. 5 to study the effectiveness of our contributions: MDF, 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 MDF 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-
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

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 <methylprednisolone, neutropenia> and <methylprednisolone, thrombocytopenia> with the prediction scores -1.55 and -1.6, respectively. And the prediction scores of <GEM, neutropenia> and <GEM, thrombocytopenia> 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


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.


