Journal of Universal Computer Science, vol. 30, no. 8 (2024), 1112-1136
submitted: 11/1/2024, accepted: 2/6/2024, appeared: 28/8/2024

Interaction and Fusion of Rich Textual Information Network for Document-level Relation Extraction

Yu Zhong
(Beijing Jiaotong University, Beijing, China
https://orcid.org/0000-0001-5623-1076, 20111030@bjtu.edu.cn)

Bo Shen
(Beijing Jiaotong University, Beijing, China
https://orcid.org/0000-0002-1040-1575, bshen@bjtu.edu.cn)

Tao Wang
(Beijing Jiaotong University, Beijing, China
https://orcid.org/0009-0004-2160-517X, 20111027@bjtu.edu.cn)

Jinglin Zhang
(Beijing Jiaotong University, Beijing, China
https://orcid.org/0009-0009-6541-2069, 19111049@bjtu.edu.cn)

Yun Liu
(Beijing Jiaotong University, Beijing, China
https://orcid.org/0000-0002-2861-0316, liuyun@bjtu.edu.cn)

Abstract: Detecting relations between entities across multiple sentences in a document, referred to as document-level relation extraction, poses a challenge in natural language processing. Graph networks have gained widespread application for their ability to capture long-range contextual dependencies in documents. However, previous studies have often been limited to using only two to three types of nodes to construct document graphs. This leads to insufficient utilization of the rich information within the documents and inadequate aggregation of contextual information. Additionally, relevant relationship labels often co-occur in documents, yet existing methods rarely model the dependencies of relationship labels. In this paper, we propose the Interaction and Fusion of Rich Textual Information Network (IFRTIN) that simultaneously considers multiple types of nodes. First, we utilize the structural, syntactic, and discourse information in the document to construct a document graph, capturing global dependency relationships. Next, we design a regularizer to encourage the model to capture dependencies of relationship labels. Furthermore, we design an Adaptive Encouraging Loss, which encourages well-classified instances to contribute more to the overall loss, thereby enhancing the effectiveness of the model. Experimental results demonstrate that our approach achieves a significant improvement on three document-level relation extraction datasets. Specifically, IFRTIN outperforms existing models by achieving an F1 score improvement of 0.67% on Dataset DocRED, 1.2% on Dataset CDR, and 1.3% on Dataset GDA. These results highlight the effectiveness of our approach in leveraging rich textual information and modeling label dependencies for document-level relation extraction.

Keywords: Natural language processing, Document-level relation extraction, Graph convolutional network
Categories: I.2.7
DOI: 10.3897/jucs.130588

1 Introduction

Relation extraction is an essential task in natural language processing, providing support for downstream tasks such as knowledge graphs [Ji et al., 2021, Zhou et al., 2021], search engines [Seymour et al., 2011], social media data analysis [Anitha et al., 2020], text networks [Belfin et al., 2020], and question answering [Kolomiyets and Moens, 2011]. Predicting relationships between entities from multiple sentences is referred to as document-level relation extraction, which is more challenging and better suited for practical applications compared to sentence-level relation extraction.

Figure 1 illustrates an example from the document-level relation extraction dataset, DocRED. The entities " Fort Sumter" and "Charleston" appear in the first sentence, and their close proximity, along with the key information "in," makes it easy to recognize that "Fort Sumter" is located in the area of "Charleston." However, the entities "American Civil War" and "Confederate" appear in the first and eighth sentences, respectively. Identifying their relationship requires integrating contextual information across sentences. It involves analyzing that the "two battles" of the "American Civil War" occurred at "Fort Sumter." Additionally, it requires recognizing that the "fort" mentioned in the eighth sentence refers to "Fort Sumter" in the first sentence, and that "it" refers to the "fort," indicating that "Fort Sumter" was controlled by the "Confederate," as in "it remained in Confederate hands."

[1] **Fort Sumter** is a sea fort in **Charleston**, South Carolina, notable for two battles of the **American Civil War**.

...

[7] The **Second Battle of Fort Sumter** (**September 8, 1863**) was a failed attempt by the Union to retake the fort, dogged by a rivalry between army and navy commanders.

[8] Although the fort was reduced to rubble , it remained in **Confederate** hands until it was evacuated as General Sherman marched through South Carolina in February 1865.

● **Intra-sentence Relation**	● **Inter-sentence Relation**
Head Entity: **Fort Sumter** Tail Entity: **Charleston** Relation: **located in**	Head Entity: **Second Battle of Fort Sumter** Tail Entity: **American Civil War** Relation: **part of**
Head Entity: **Second Battle of Fort Sumter** Tail Entity: **September 8, 1863** Relation: **point in time**	Head Entity: **Confederate** Tail Entity: **American Civil War** Relation: **participant of**, **conflict**

Figure 1: An example from DocRED

Therefore, the complete relation support chain from the head entity "Confederate" to the tail entity "American Civil War" is "Confederate" - "remained in ... hands" - "it" - "fort" - "Fort Sumter" - "two battles" - "American Civil War." This example demonstrates that document-level relation extraction requires models to have the capability to integrate contextual information.

Document-level relation extraction methods can be summarized into sequence-based methods and graph-based methods. Sequence-based methods directly utilize neural architectures or pre-training models to model the entire document. For instance, local context pooling was proposed by [Zhou et al., 2021] to make different entity pairs pay attention to different context information. They introduced an adaptive threshold to reduce decision errors to tackle the multi-label problem. [Xu et al., 2021a] summarized various kinds of mention dependencies, and they integrate these dependencies into self-attention mechanisms and throughout the overall encoding stage. However, these models often ignore the structural information of documents, which can be used as prior knowledge to help the model understand relationships between entities.

Graph-based methods capture structural information through multiple iterations and establish long-distance dependencies. However, previous methods have modeled in a coarse-grained manner, leading to the loss of important information that contributes to identifying relationships. For example, [Sahu et al., 2019] and [Nan et al., 2020] constructed document graphs using syntactic dependency relationships but focused primarily on local dependencies. To overcome these limitations, we propose the Interaction and Fusion of Rich Textual Information Network (IFRTIN), which considers multiple types of nodes to construct a comprehensive document graph.

To tackle these problems, we propose an Interaction and Fusion of Rich Textual Information network to simultaneously consider multiple types of information. First, we construct a document graph based on the structural, syntactic, and discourse information in the document, modeling the global interaction between words, mentions, entities, elementary discourse units, sentences, and documents. Then, we build a label dependency graph based on the relationships appearing in the document. This graph consists of intra-sentence and inter-sentence labels, and the dependencies between labels are learned under the influence of a regularizer. Additionally, to address the common problem of decreased model performance due to a focus on more challenging classifications in document-level relation extraction models, we propose an Adaptive Encouraging Loss.

Our contributions can be summarized as follows:

• Proposing a document-level relation extraction model based on the Interaction and Fusion of Rich Textual Information, which effectively models interactions between entities using rich information within the document, enabling simultaneous extraction of inter-sentence and intra-sentence relations.

• We introduce a new regularizer to constrain the representation vectors learned by GCN in predicting the graph and the gold dependency graph, capturing dependencies between relationship labels.

• We design an Adaptive Encouraging Loss, which encourages well-classified instances to make a greater contribution to the overall loss in order to enhance the effectiveness of the model.

• Demonstrating the superior performance of our model on three dataset through extensive experiments, validating the effectiveness of our proposed approach.

2 Motivation and Related Work

2.1 Sentence-level Relation Extraction

The goal of relation extraction is to identify the relationships between entities within specific texts. Early approaches, such as those by [Agichtein and Gravano, 2000, Rink and Harabagiu, 2010], relied on statistical machine learning techniques to determine relationships within single sentences. These methods heavily depended on manually crafted features, resulting in performance limitations and the potential for error propagation. With the advent of neural networks, researchers like [Zhang et al., 2017, Guo et al., 2021] shifted towards automatic feature extraction using these advanced models. [Wang et al., 2016] introduced a multi-level attention convolutional neural network (CNN), which utilized entity-specific attention in the input layer and relation-specific pooling attention in the pooling layer to identify patterns within diverse contexts. [Zhou et al., 2016] utilized attention mechanisms alongside Bidirectional Long Short-Term Memory (BiLSTM) networks to pinpoint crucial word information for relation classification. [Luo et al., 2017] implemented a curriculum learning approach to train adaptable transition matrices. [Ji et al., 2017] introduced a sentence-level attention model that selected multiple relevant sentences from a document and enriched entity representations with detailed descriptions. [Zhu et al., 2019] established fully connected graphs to support multi-hop relation reasoning. [Guo et al., 2019] leveraged dependency structures to form sentence graphs, which were then refined using attention-based pruning.

The relation between entities is usually inferred from multiple sentences in the real world. Therefore, more and more work has begun to attach importance to document-level relationship extraction [Han and Wang, 2020, Kuang et al., 2022, Wang et al., 2020b, Li et al., 2022, Liu et al., 2020].

2.2 Document-level Relation Extraction

Sequence-based methods and graph-based methods are the main methods of document-level relation extraction.

The sequence-based methods directly utilize neural architectures or pre-training models to model the entire document. [Ye et al., 2020] introduced mention reference prediction to do pre-training tasks. In this way, the model could learn much coreference information. A multi-task learning method based on reference parsing was proposed by [Eberts and Ulges, 2021] to achieve end-to-end joint relation extraction. They extracted entity sets in documents and used multi-instance learning to predict relationships between entities by combining global-level entity representation with local-level mention representation. [Xu et al., 2021a] summarized various kinds of mention dependencies, and they integrate these dependencies into self-attention mechanisms and throughout the overall encoding stage. [Zhou et al., 2021] introduced local context pooling to make different entity pairs pay attention to the different context information and an adaptive threshold to reduce decision errors. [Dong and Xu, 2023] improved the model's understanding of coreference information by implementing mention replacement and used contrastive learning to better perceive relation distances.

Because graph neural networks can simulate the interaction between nodes, a large number of document-level relation extraction models [Zeng et al., 2021, Xu et al., 2021b, Sun et al., 2022] are based on graph structure. [Sahu et al., 2019] constructed a document graph, using syntactic parsing to model local dependencies, and utilizing coreference resolution and other semantic dependencies to model non-local dependencies. Various types of nodes and edges were introduced by [Christopoulou et al., 2019] to create a document graph that expresses entity relations through edge representations formed as paths between nodes. To overcome the problem of the document graph modeling a large number of entity pairs with no relationship, [Xu et al., 2021c] proposed a re-constructor to reconstruct the ground-truth path dependencies from the document graph. In this way, the model would focus on related entity pairs. [Wang et al., 2020a] created a graph with heuristic rules to learn global entity representations and used multi-head attention to learn local entity representations. Two graphs were constructed by [Zeng et al., 2021] for reasoning. One was the mention-level graph, which is responsible for simulating the interaction between different mentions. The other was the entity-level graph, which inferred the relationship between entities through path attention. [Nan et al., 2020] utilized structural attention to capture global dependency. Wang et al. [Wang et al., 2021] utilized discourse information to construct a document-level graph for capturing semantic dependencies between text units. Wan [Wan et al., 2023] reconstructed the document by region and introduced bridge entities to construct a dependency structure, aiming to improve the efficiency of relation extraction. [Liu et al., 2023] captured semantic information within documents by constructing document-level graphs and modeled long-distance relations between entities by creating entity-level graphs.

2.3 Research Questions

Despite the promising results achieved by sequence-based methods and graph-based methods in the field of document-level relation extraction, they still have certain limitations. On one hand, sequence-based methods often ignore the structural information of documents, which can serve as prior knowledge to help the model understand relationships between entities. On the other hand, graph-based methods tend to model in a coarse-grained manner, leading to the loss of important information that contributes to identifying relationships.

Based on the gaps in the existing research, we formulate the following research questions:

RQ1: How can we effectively model the rich structural, syntactic, and discourse information in a document to improve document-level relation extraction?

RQ2: Can the integration of multiple types of nodes (e.g., words, mentions, entities, discourse units) into a single document graph enhance the performance of relation extraction models?

RQ3: How does modeling the dependencies of relationship labels impact the accuracy of document-level relation extraction?

2.4 Contribution

The IFRTIN model addresses aforementioned several research questions. By constructing a more comprehensive document graph and utilizing advanced regularization techniques, our model achieves superior performance in document-level relation extraction tasks. Specifically, our experiments demonstrate significant improvements on three datasets: DocRED, CDR, and GDA, highlighting the effectiveness of our approach.

Rich Representation of Document Information: By incorporating multiple types of nodes, IFRTIN captures a broader range of information within documents. This allows for more nuanced reasoning between entities, especially those that are not directly connected.

Regularizer for Relationship Label Dependencies: The regularizer helps in learning dependencies between relationship labels, which is often overlooked in traditional methods. This improves the model's ability to predict complex relationships accurately.

Adaptive Encouraging Loss: This novel loss function encourages instances that are well-classified to contribute more to the overall loss, enhancing the model's effectiveness in handling diverse and challenging classifications.

Experimental Validation: Our extensive experiments validate the superior performance of IFRTIN over existing models. The F1 score improvements on the DocRED, CDR, and GDA datasets underscore the practical applicability and robustness of our approach.

In conclusion, the IFRTIN model represents a significant advancement in document-level relation extraction by leveraging rich textual information and advanced regularization techniques. Our contributions pave the way for more accurate and efficient extraction of relationships in complex documents, addressing key challenges in the field and setting a new benchmark for future research.

3 Study Design

The study aims to enhance document-level relation extraction by constructing a comprehensive document graph and capturing label dependencies through innovative network design. The proposed Interaction and Fusion of Rich Textual Information Network (IFRTIN) integrates multiple types of nodes and utilizes advanced regularization techniques to achieve superior performance.

3.1 Methodology

As shown in Figure 2, our IFRTIN framework consists of four parts: (i) the Encoder; (ii) the Document Hierarchical GCN; (iii) the Classifier (iv) and the Relation Type Dependency-based Regularizer.

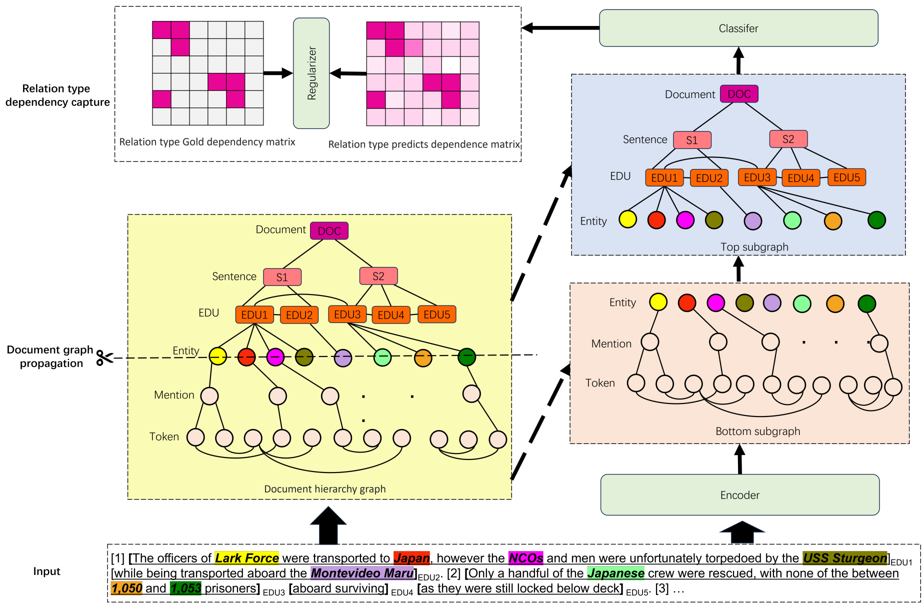

Figure 2: The architecture of our model IFRTIN

3.1.1 Task Formulation

Given a document $d = \{x_i\}_{i=1}^{N_d}$ containing multiple sentences $s = \{s_i\}_{i=1}^{N_s}$ and entities $e = \{e_i\}_{i=1}^{N_e}$, the aim is to extract relation r from R between head and tail entities

(e_h, e_t), where R is a pre-defined set of relations and r may have one or more. In the document, each entity ei may occur multiple times by mentions $m_i = \{m_{ij}\}_{j=1}^{N_{m\,i}}$.

3.1.2 Encoder

Given the promising performance of pre-trained language models in various downstream tasks, we use BERT to encode document d:

$$[h_1, h_2, ..., h_{N_d}], A = \text{Encoder}([x_1, x_2, ..., x_{N_d}]), \tag{1}$$

where A is a multi-head attention matrix output by the encoder.

Following [Soares et al., 2019] and [Shi and Lin, 2019], we insert a symbol ”*” at the beginning and end of each mention to mark the entity positions. We take the embedding of ”*” before each mention as the mention embedding:

$$m_{ij} = h_{a_{ij}}, \tag{2}$$

where m_{ij} is j-th mention of i-th entity, a_{ij} is the position of ”*” before the m_{ij}.

For an entity e_i, we leverage a smooth version of max pooling – logsumexp pooling [Jia et al., 2019] to obtain the entity embedding:

$$e_i = log \sum_{j=1}^{N_{e_i}} exp(m_{ij}), \tag{3}$$

where $e_i \epsilon R^d$. In this way, we aggregate different mentions information belonging to the same entity.

Similarly, we also apply logsumexp pooling to obtain the sentence embedding:

$$s_i = log \sum_{j=1}^{N_{s_i}} exp(h_{ij}), \tag{4}$$

where $s_i \epsilon R^d$.

We then use the embedding of [CLS] as document embedding. Besides, for an entity pair (e_h, e_t), we compute its context representation $c_{h,t}$ based on well-learned dependencies from the pre-training language model:

$$c_{h,t} = H^T \frac{A_h \cdot A_t}{1^T (A_h \cdot A_t)}, \tag{5}$$

where A_h is attention from the i-th entity to all tokens in the document. Similar for A_t.

3.1.3 Document Hierarchical GCN

As shown in Figure 2, we can construct a document-level graph based on the syntactic, structural, and discourse information implied in the document. The document-level graph consists of six types of nodes and seven types of edges. To capture the rich contextual information in documents, we constructed a document graph incorporating various node types:

Word Nodes: Represent individual words in the document. Mention Nodes: Represent mentions of entities. Entity Nodes: Represent unique entities linked by coreference resolution. Discourse Units: Represent discourse elements derived from discourse parsing. Edges between nodes were established based on syntactic dependencies, coreference links, and discourse relations, resulting in a global graph structure that captures long-range dependencies.

However, in this graph, interactions between the lower-order word nodes and the higher-order document nodes require passing through at least four other types of nodes. To address the issue of long interaction distances between nodes, we use entity nodes as intermediate nodes to partition the document-level graph into lower-order and higher-order subgraphs, respectively, to learn useful information.

Lower-order Subgraph

The lower-order subgraph comprises three types of nodes: word nodes, mention nodes, and entity nodes. It contains edges of three types, which are detailed as follows:

Syntactic Edge: To capture syntactic information in the document, we take syntax dependencies as input, enabling the construction of the lower-order adjacency matrix A^{lower}. Specifically, if x_i and x_j are two words in the same sentence and they are connected in the corresponding dependency tree, we set A_{ij}^{lower} to 1; otherwise, it is set to 0. Thus, if two words are adjacent in the dependency tree, it is considered important for representation learning in document-level relation extraction.

Mnetion-word Edge:If a word falls within the span of a mention, we use an Mention-word Edge to connect this word with this mention.

Entity-Mention Edge:To capture global information, we connect each mention with its corresponding entity using Entity-Mention Edges.

Based on the above rules, we construct the lower-order graph matrix A^{lower}. For graph-structured data, GCN can encode local information with convolutional operations and aggregate global information by message passing in multiple convolutional layers. For the i-th node at the l-th layer, its hidden state representation, denoted as h_i^l, is updated by the following equation:

$$h_i^l = \sigma(\sum_{g \epsilon G} \sum_{j \epsilon N_g(i)} W_g^{l-1} h_j^{l-1} + b_g^{l-1}), \tag{6}$$

where W_g^{l-1} is a weight matrix, b_g^{l-1} is a bias term and G are different types of edges. $N_g(i)$ denotes neighbors for node i connected in g-th type edge. Next, we concatenate the hidden states of each layer:

$$e_i^{'} = \sigma(W_{lower}[h_i^0 : h_i^1 : ...h_i^N]), \tag{7}$$

Higher-order Subgraph

In order to leverage structural and discourse information, we construct a higher-order graph A^{higher} for each document. Specifically, we use entity, elementary discourse unit,sentence and document nodes to build A^{higher}. The description is as follows:

• Discourse-Entity Edge

If an entity is within a discourse unit, then we use a Discourse-Entity Edge to connect this entity with the unit.

• Sentence-Discourse Edge

A Sentence can be divided into one or more discourse units. We connect each discourse unit with its corresponding sentence using a Sentence-Discourse Edge to capture discourse information.

• Document-Sentence Edge

We connect all sentence nodes to the document node through Document-Sentence Edges.

Similar to the lower-order subgraph module, the higher-order subgraph module obtains the final representation of node i by concatenating hidden states of each layer:

$$e_i^{''} = \sigma(W_{higher}[h_i^0 : h_i^1 : ...h_i^N]). \tag{8}$$

3.1.4 Classification Module

We initialize all relationship labels r_i as vectors randomly. We use the context vector obtained from Eq (5) as the query vector, and compute the corresponding relationship representation for each entity pair:

$$r_{(h,t)} = \sum_i \alpha_i r_i \tag{9}$$

$$\alpha_i = \frac{exp(Wc_{(h,t)} \cdot Wr_i)}{\sum_o exp(Wc_{(h,t)} \cdot Wr_o)} \tag{10}$$

$$r_{(h,t)} = \sum_i \alpha_i r_i \tag{11}$$

Finally, we fuse context representations to get the final representations of the head and tail entities:

$$z_h^{(h,t)} = tahn(W_h \cdot e_h + W_{h'} \cdot e_h^{''} + W_{c_1} \cdot c_{h,t} + W_{o1} \cdot r_{(h,t)}), \tag{12}$$

$$z_t^{(h,t)} = tahn(W_t \cdot e_t + W_{t'} \cdot e_t^{''} + W_{c_2} \cdot c_{h,t} + W_{o2} \cdot r_{(h,t)}), \tag{13}$$

where W_h, W_t, $W_{h'}$, $W_{t'}$, W_{c_1} and W_{c_2}, W_{o_1} and W_{o_2} are the weight matrix. We then employ the group bilinear [Tang et al., 2020b] to reduce the number of parameters in the bilinear classifier. To be specific, z_h and z_t will be split into k equal-sized groups:

$$z_h = [z_h^1; ...; z_h^k], \tag{14}$$

$$z_t = [z_t^1; ...; z_t^k]. \tag{15}$$

We apply bilinear to obtain the representation of entity pair (e_h, e_t):

$$b_{h,t} = \sigma(\sum_{i=1}^{k}(z_h^i)^T W_r^i z_t^i + b_r), \tag{16}$$

where W_r^i and b_r are the weight matrix and the bias term.

We then use a linear layer to predict the relationship between e_h and e_t:

$$P(r|e_h, e_t) = \text{FFN}(b_{h,t}). \tag{17}$$

Adaptive Encouraging Loss

To overcome the limitation of the global threshold, [Zhou et al., 2021] introduce a unique class as the adaptive threshold value for each example. However, their approach overlooks well-classified examples that are far from the decision boundary. In order to emphasize these well-classified examples, this paper proposes an adaptive reward loss function. Our loss function consists of two parts: the first part is used to compute the loss between positive classes and the threshold, while the second part is used to compute the loss between negative classes and the threshold. The probability calculation for each positive class is as follows:

$$P_r = -\frac{exp(logit_r)}{\sum_{r' \epsilon P_T U\{TH\}} exp(logit_{r'})} \tag{18}$$

where P_T denote set of positive classes and TH denote TH class.

The first part of the loss function can be represented as:

$$L_1 = \begin{cases} \sum_{r \epsilon P_T}(-logP_r + log(1 - P_r)), \text{ if } P_r \leq LE_1 \\ \sum_{r \epsilon P_T}(-logP_r - \frac{P_r - LE_1}{1 - LE_1} + log(1 - LE_1)), \text{ if } P_r > LE_1 \end{cases} \tag{19}$$

For the negative classes, their probability can be computed as follows:

$$P_{TH} = -\log(\frac{exp(logit_{TH})}{\sum_{r' \epsilon N_T U\{TH\}} exp(logit_{r'})} \tag{20}$$

The second part of the loss function can be represented as:

$$L_2 = \begin{cases} -logP_{TH} + log(1 - P_{TH}), \text{ if } P_{TH} \leq LE_2 \\ -logP_{TH} - \frac{P_{TH} - LE_2}{1 - LE_2} + log(1 - LE_2), \text{ if } P_{TH} > LE_2 \end{cases} \tag{21}$$

The L_{RE} can be represented as:

$$L_{RE} = L_1 + L_2 \tag{22}$$

3.1.5 Relation Type Dependency-based Regularizer

To capture the dependencies between labels, we construct the relation type probability vector $l^{pred} \in R^k$based on the entity pair relation probabilities output by Eqs. 2-4. $k$is equal to the number of relation label types, and the value of the $i$row of the vector is equal to the maximum probability of all pairs of entities in the document having relation type r_i. Therefore, the relational type prediction dependency matrix A^{pred}can be computed as follows.

$$A^{pred} = l^{pred} \cdot (l^{pred})^T \tag{23}$$

Then, the relation type gold vector $l^{gold} \in R^K$is constructed, and the value of the $i$row of the vector is equal to 1 if the relation type $r_i$occurs in the document, and 0 otherwise. The relational type gold dependency matrix A^{gold}can be computed as follows.

$$A^{gold} = l^{gold} \cdot (l^{gold})^T \tag{24}$$

Next, we feed these two matrices and the initial relation type embedding R_0into the same GCN, Generate tags embedded relationship between $R_{pred} = GCN(R_0, A^{pred})$and $R_{gold} = GCN(R_0, A^{gold})$.

Our goal is to minimize the difference between the relational type prediction dependency matrix $A^{pred}$and the relational type gold dependency matrix A^{gold}. If the difference between these two matrices is small, then the difference between the outputs of the GCN using these two matrices as adjacency matrices should also be small. The difference value is calculated as follows:

$$L_{REG} = \frac{1}{K}||R_{pred} - R_{gold}||_2^2 \tag{25}$$

where Kdenotes the number of relation labels.

The loss function can be calculated as follows:

$$L = L_{RE} + \eta L_{REG} \tag{26}$$

where ηdenotes the coefficient of the regularization term.

3.2 Instruments

We employed various tools and libraries to facilitate our experiments: Spacy for tokenization and dependency parsing. PyTorch for building and training the neural network models. NetworkX for constructing and managing graph data structures. We run our IFRTIN-BERT-base on one RTX A5000 GPU and IFRTIN-RoBERTa-large on one Tesla A40 GPU.

The preprocessing steps included tokenization, part-of-speech tagging, and dependency parsing to extract structural information. Additionally, coreference resolution was performed to identify and link entity mentions across sentences. We implement our IFRTIN with the PyTorch version of the Huggingface Transformers and use BERT-base [Devlin et al., 2018] and SciBERT [Beltagy et al., 2019] as encoders on DocRED. We optimize IFRTIN with AdamW using learning rates 3e-5 with a linear warmup for the first 6% of steps. During the fine-tuning stage, we set the batch size to 4 and train the model for 30 epochs. Early stopping was used based on the dev F1 score to avoid over-fitting.

4 Experiment

4.1 Datasets

We evaluate our IFRTIN model on three public document-level relation extraction datasets.

DocRED is a new large-scale document-level relation extraction dataset constructed by [Yao et al., 2019]. The dataset provides 3053 articles in the training set, 1000 articles in the development set, 1000 articles in the test set, and 101873 articles in the distantly supervised set. It should be noted that we do not use the distantly supervised set.

CDR is a relation extraction dataset in the biomedical domain constructed by [Li et al., 2016]. The dataset provides 500 articles in the training set, 500 articles in the development set, 500 articles in the test set.

GDA is a dataset constructed by [Wu et al., 2019]. The dataset provides 23353 articles in the training set, 5839 articles in the development set, 1000 articles in the test set.

4.2 Comparison models

In this section, we describe variants of IFRTIN and some Baselines for the document-level relation extraction task.

4.2.1 Variants of IFRTIN model

In IFRTIN model, document graph propagation module, adaptive incentive loss and relation type dependency capture module are very important components of IFRTIN. In order to verify the validity of these components and modules, some variant models of IFRTIN are proposed. Among these variants, IFRTIN-v1 and IFRTIN-v2 are used to verify the validity of the document graph propagation module model. IANN-v3 is used to verify the effectiveness of the adaptive excitation loss. IANN-v4 is used to verify the validity of the relational type dependency capture module.

- **IFRTIN-v1 :** The first variant of IFRTIN removes the lower-order subgraph from the document graph propagation module based on IFRTIN. In this variant model, nodes containing rich local information such as mention nodes and sub-word nodes are removed, resulting in IFRTIN-v1 lacking certain ability to capture local information.
- **IFRTIN-v2:** The second variant of IFRTIN removes the higher-order subgraph from the document graph propagation module. In this variant model, EDU nodes, sentence nodes and document nodes are all removed, which leads to the blocking of the propagation of higher-order semantic information.
- **IFRTIN-v3:** The third variant of IFRTIN adds an adaptive incentive loss to IFRTIN. In this variant, the adaptive incentive loss is replaced by an adaptive threshold loss, and excessive attention to difficult instances that are difficult to optimize leads to performance degradation.
- **IFRTIN-v4 :** The fourth variant of IFRTIN does not focus on relation type dependencies, which makes it difficult to guide relation classification by the co-occurrence of relation types.

4.2.2 Baselines

We compare IFRTIN with sequence-based methods and graph-based methods on the DocRED dataset. Sequence-based models, which use a pre-trained language model to directly model the entire document without using graph structures, include BRAN [Verga et al., 2018], CNN [Yao et al., 2019], LSTM [Yao et al., 2019], BiLSTM [Yao et al., 2019], Context-Aware [Yao et al., 2019], BERT [Wang et al., 2019], Coref-BERT [Ye et al., 2020], HIN-BERT [Tang et al., 2020a], SSAN-BERT [Xu et al., 2021a], and ATLOP-BERT [Zhou et al., 2021]. Graph-based models, which use graph structures to model the document, include DHG[Zhang et al., 2019], GAT [Veličković et al., 2017], GCNN [Sahu et al., 2019], EoG [Christopoulou et al., 2019], AGGCN [Guo et al., 2019], HeterGSAN [Xu et al., 2021c], GEDA [Li et al., 2020], LSR [Nan et al., 2020], GAIN [Zeng et al., 2020], DRN [Xu et al., 2021b], DISCO [Wang et al., 2021], MPCA [Ding et al., 2023], and HDT-BP [Wan et al., 2023].

5 Results

5.1 Main Results

5.1.1 Results on the DocRED Dataset

The experimental results in Table 1 show that the proposed IFRTIN model consistently outperforms both sequence-based and graph-based methods on the DocRED dataset. Specifically, IFRTIN achieves an F1 score of 61.81% on the development set and 61.97% on the test set. The F1 value on the development set and test set is 2.62% and 3.81% higher than that of SSAN, and 2.44% and 3.20% higher than that of HDT-BP. The main reason for the improvement of IFRTIN is that the document graph is constructed in a fine-grained way, and the context information can be effectively aggregated by serial propagation on the two subgraphs.

In addition, in order to verify the effectiveness of each key component in IFRTIN, four variants of IFRTIN model are compared with the complete IFRTIN model. The results are

Model	Dev		Test	
	Ign F1	F1	Ign F1	F1
Sequence-based Models				
CNN [Yao et al., 2019]	41.58	43.45	40.33	42.26
LSTM [Yao et al., 2019]	48.44	50.68	47.71	50.07
BiLSTM [Yao et al., 2019]	48.87	50.94	48.78	51.06
Context-Aware [Yao et al., 2019]	48.94	51.09	48.40	50.70
BERT [Wang et al., 2019]	-	54.16	-	53.20
Coref-BERT [Ye et al., 2020]	55.32	57.51	54.54	56.96
HIN-BERT [Tang et al., 2020a]	54.29	56.31	53.70	55.60
ATLOP-BERT [Zhou et al., 2021]	59.22	61.09	59.31	61.30
SSAN-BERT [Xu et al., 2021a]	57.03	59.19	55.84	58.16
Graph-based Models				
AGGCN [Guo et al., 2019]	46.29	52.47	48.89	51.45
GAT [Veličković et al., 2017]	45.17	51.44	47.36	49.51
GCNN [Sahu et al., 2019]	46.22	51.52	49.59	51.62
EOG [Christopoulou et al., 2019]	45.94	52.15	49.48	51.82
HeterGSAN [Xu et al., 2021c]	54.27	56.22	53.27	55.23
GLRE-BERT [Wang et al., 2020a]	-	-	55.40	57.40
LSR-BERT [Nan et al., 2020]	52.43	59.00	56.97	59.05
GAIN-BERT [Zeng et al., 2020]	59.14	61.22	59.00	61.24
GEDA-BERT [Li et al., 2020]	51.03	53.60	51.22	52.97
HeterGSAN-BERT [Xu et al., 2021c]	58.13	60.18	57.12	59.45
DISCO-BERT [Wang et al., 2021]	55.91	57.78	55.01	55.70
HDT-BP-BERT [Wan et al., 2023]	57.17	59.37	56.28	58.77
IFRTIN-v1-BERT	59.36	61.29	59.28	61.22
IFRTIN-v2-BERT	59.41	61.36	59.46	61.39
IFRTIN-v3-BERT	59.67	61.64	59.73	61.71
IFRTIN-v4-BERT	59.56	61.51	59.64	61.66
IFRTIN-BERT	**59.92**	**61.81**	**60.03**	**61.97**

Table 1: Main results (%) on DocRED

shown in Table 1, and the performance of the full IFRTIN model is better than that of all variant models, indicating that the four key components of the IFRTIN model contribute to its performance improvement. DDGSN-v1 performs the worst among the four variants, which indicates that the high-order information contained in the higher-order subgraph is crucial for the model to perform effective relation extraction.

5.2 Results on the CDR and GDA Datasets

In order to verify the effectiveness of the IFRTIN model on different datasets, further experiments are conducted on CDR and GDA datasets in this section. The experimental results are shown in Table 2, where the IFRTIN model achieves an F1 score of 70.6% on CDR and 85.2% on GDA. The IFRTIN model improves the performance of the previous best model ATLOP by 1.2%F1 on the CDR dataset and 1.3%F1 on the GDA dataset. The excellent performance on document-level relation extraction tasks is attributed to effective aggregation of contextual information, effective attention to well-classified instances, and effective capture of relation type dependencies.

Model	CDR	GDA
BRAN	62.1	-
EoG	63.6	81.5
SciBERTbase	65.1	82.5
LSR	64.8	82.2
DHG	65.9	83.1
GLRE	68.5	-
ATLOP	69.4	83.9
IFRTIN	**70.6**	**85.2**

Table 2: Main results (%) on CDR and GDA. We use F1 as evaluation metric

5.3 Performance analysis of intra-sentence and inter-sentence relation extraction

Due to differences in the pattern and complexity of relations, the performance of relation extraction models will vary greatly within and between sentences. To investigate the performance variation of IFRTIN in intra-sentence and inter-sentence relation extraction, we present the statistical results in Figure 3.

It can be seen from the figure that the performance of all models in inter-sentence relation extraction (Inter-F1) is far worse than intra-sentence relation extraction (Intra-F1), highlighting that since long-distance context information needs to be considered in inter-sentence relation extraction, relation recognition is more difficult than intra-sentence relation extraction. IFRTIN achieves the best performance in both intra-sentence and inter-sentence relation extraction, and the performance in inter-sentence relation extraction is more significantly improved. This shows that the model effectively captures not only local but also global information. This is attributed to our document graph propagation module, which effectively integrates the contextual information scattered in multiple sentences in a document in a serial propagation manner by dividing the document graph with deep hierarchical structure into two different subgraphs.

5.4 Analysis of the impact of the number of entities on model performance

In general, the more entities there are in a document, the more difficult it is to infer relationships between them. To evaluate the effectiveness of IFRTIN proposed in this section, the documents in the DocRED development set are grouped according to the number of entities. IFRTIN and the strong baseline model ATLOP are compared in these document groups.

The experimental results are shown in Figure 4, IFRTIN consistently outperforms ATLOP, especially for documents containing a small number of entities (1 to 5,6 to 10) and a large number of entities (26 to 30,31 to 35). Moreover, the performance of all these models degrades across the table as the number of document entities increases. However,

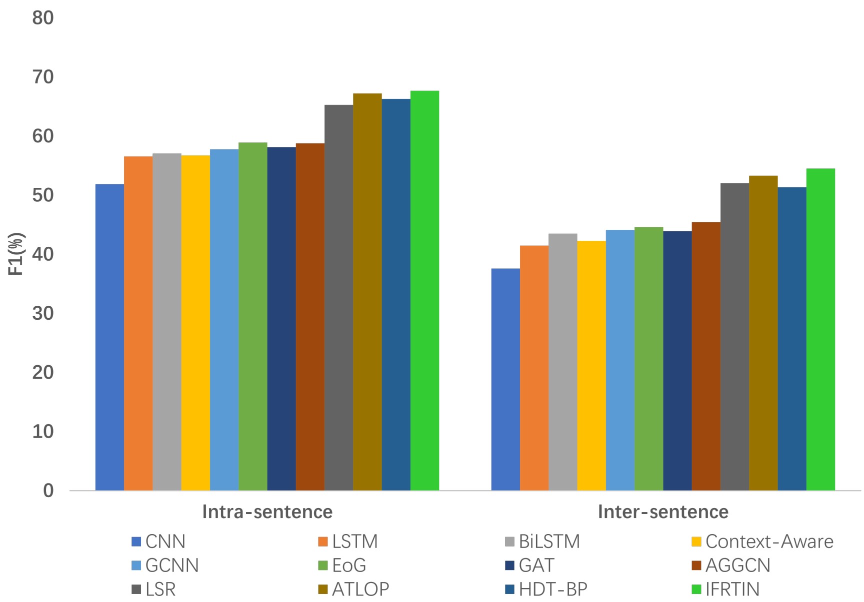

Figure 3: Performance comparison of intra-sentence and inter-sentence relation extraction on DocRED development set

IFRTIN performance degrades less than ATLOP, which indicates that it can effectively model complex entity interactions. This robustness shows that IFRTIN can better handle complex documents with different entity densities and maintain higher accuracy and reliability even when the document complexity increases. This demonstrates the superior ability of the model to handle complex semantic relationships in different document scenarios.

5.5 Analysis of the effect of the number of sentences on the performance of the model

In the document-level relation extraction task, the number of sentences has a significant impact on the model performance. Grouping documents with different number of sentences, the experimental results on different ancestors are shown in Table 3.

By analyzing the experimental results, we find that the extraction performance of the model shows a certain trend with the increase of the number of sentences. In general, the model is able to maintain high accuracy and stability when dealing with documents containing a small number of sentences. However, when the number of sentences in a document increases significantly, the model needs to deal with more complex contexts and more entity interactions, which may lead to fluctuations in performance. Nevertheless, our experimental results show that the model is stable across documents with different number of sentences, and can achieve a satisfactory F1 score especially for documents with a large number of sentences (above 57% in all documents except for documents with 15 - 16 sentences). This shows that the model can effectively capture and utilize complex relation features in documents, adapt to documents of different lengths, and thus

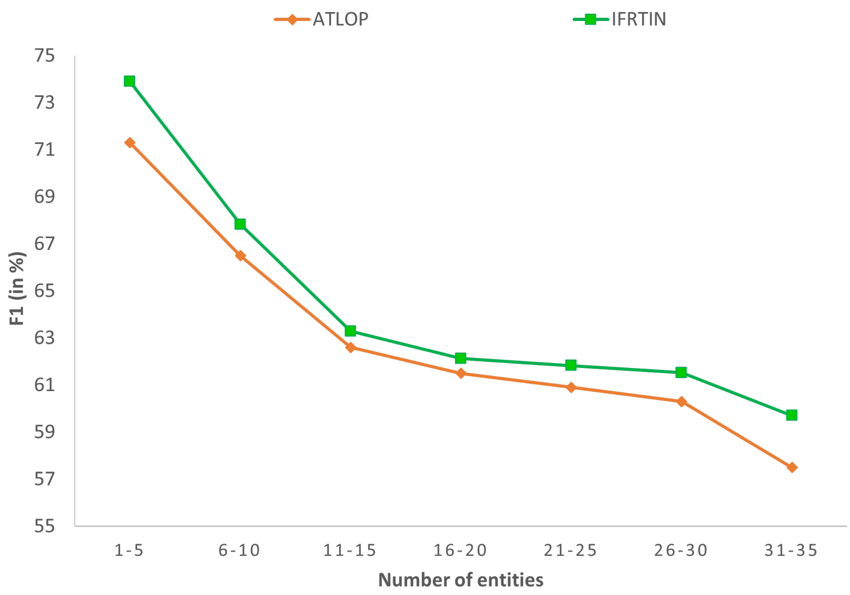

Figure 4: Impact of the number of entities on the DocRED development set on model performance

Sentence number	IgnF1	F1	P	R	Intra-F1	Inter-F1
<5	59.59	60.77	75.60	50.80	64.16	53.96
5, 6	61.37	63.24	69.47	58.03	68.55	55.87
7, 8	60.93	62.89	67.57	58.82	69.10	55.82
9, 10	60.41	62.21	66.86	58.16	68.14	55.15
11, 12	57.55	59.79	65.47	55.02	65.82	53.65
13, 14	57.69	59.75	69.05	52.65	69.44	50.73
15, 16	38.36	41.27	51.32	34.51	47.67	34.59
>16	61.62	62.98	75.00	54.29	72.73	43.33

Table 3: Analysis of the impact of the number of sentences on model performance on the DocRED development set

maintain good performance in document-level relation extraction with multi-sentence structures. In addition, the F1 value of relation extraction of the model in the documents with 15 - 16 sentences is only 41.27%, which may be because the documents with 15 - 16 sentences only account for 2% of the total number of documents in the validation set, and the difficult instances are mainly concentrated in these 2% documents, resulting in poor performance of the model. It can also be observed that as the number of sentences changes, intra-sentence F1 is almost not affected by the number of sentences, while increasing the number of sentences leads to a significant decrease in inter-sentence F1.

5.6 Analysis of the influence of relation type dependent regularization term coefficient on model performance

The coefficient of the regularization term in this chapter is an important parameter that controls how much the model pays attention to relational dependencies, and by adding a regularization term to the loss function, we can force the model to learn relational dependencies. In this paper, we provide a detailed analysis of the effect of the regularization term coefficients on the model performance. The experimental results are shown in Figure 5.

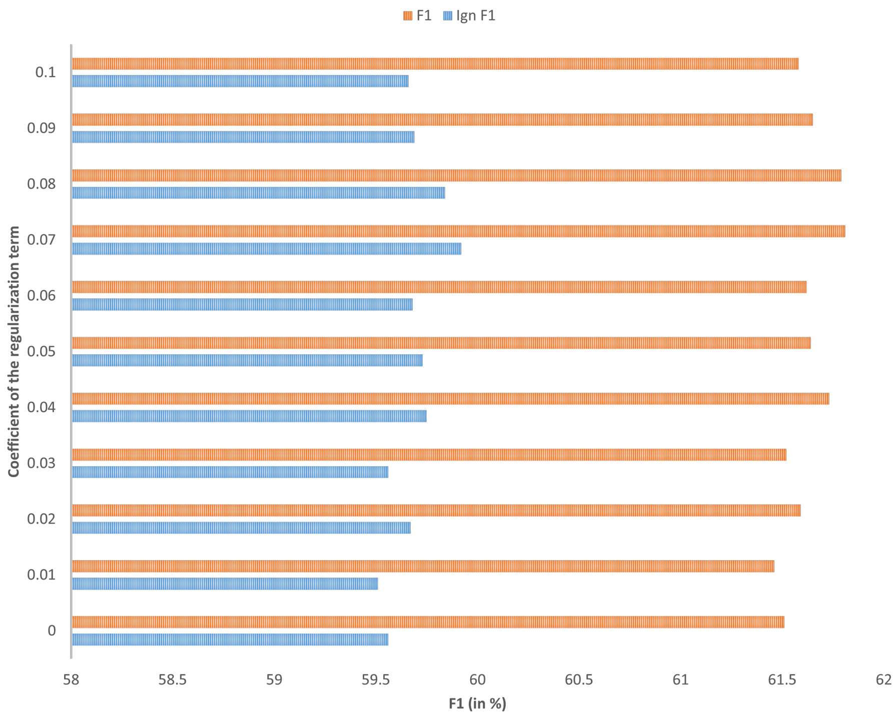

Figure 5: Effect of relation type dependence regularization term coefficients on model performance on DocRED development set

First, we selected a set of different regularization coefficient values (the weight coefficient of the regularization term is increased from 0 to 0.1 in steps of 0.01) to train the model, and recorded the model performance corresponding to each coefficient. By comparing these performance metrics, it can be found that the adjustment of the regularization term coefficient has a significant impact on the accuracy of the model. When the regularization factor is too small ($\leq$0.03), the model may struggle to capture the relation type dependency, resulting in poor performance on the validation set. An appropriate regularization coefficient (0.04-0.08) can effectively balance the model capturing relation type dependence and relation recognition, and the performance of the model in this case is relatively ideal. When the regularization factor is too large (≥0.9), the regularization effect is too strong, and the model focuses too much on the co-occurrence between relation types, resulting in underfitting of relation extraction. At this point, the error of the model on the dataset increases and it cannot effectively capture the complex patterns in the data. The experimental results show that the relationship type dependence regularization coefficient has an important impact on the performance of the model, and the model performs best when the regularization coefficient is 0.07.

5.7 Case Study

We conduct a case study to further illustrate the effectiveness of our model IFRTIN-BERT-base. Figure 6 shows an example from the DocRED dataset. We present the inference results of IFRTIN. The document consists of 3 sentences and 13 entities, with entities of different types displayed in different colors. The Golden Label below the figure represents the true relationships between entities, while the Prediction Label represents the predictions made by IFRTIN. Black solid arrows in the figure represent correctly predicted relationships, while red solid arrows indicate incorrectly predicted relationships, and red dashed arrows represent relationships that were not predicted. We follow [Yao et al., 2019] for the definition of a relation, such as P607 and P166.

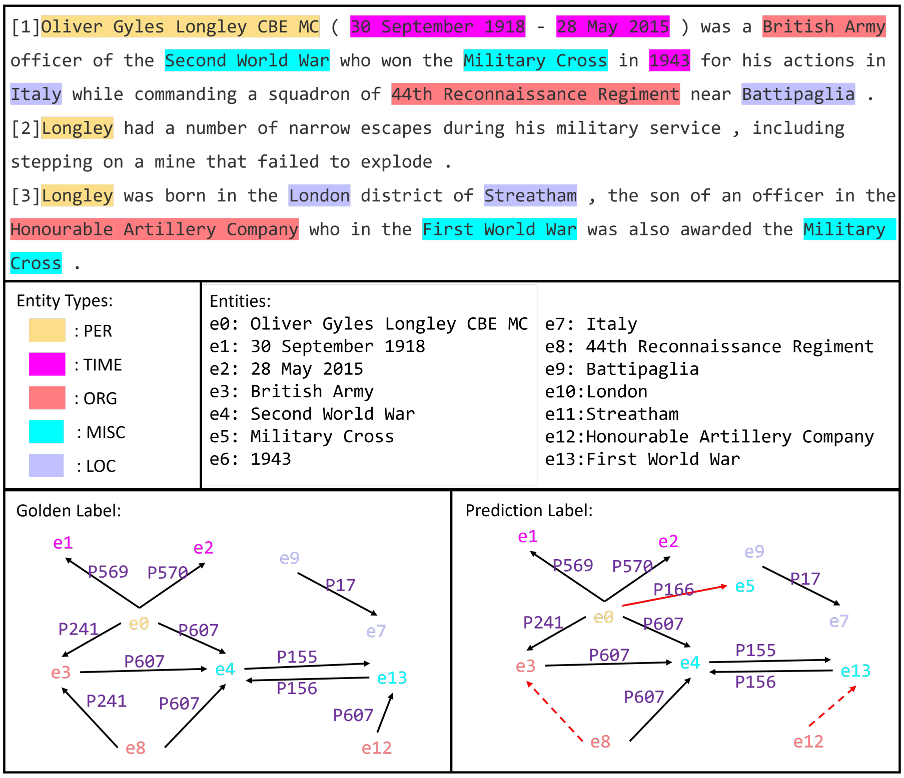

Figure 6: Case study of an example from the development set of DocRED

From the predicted results in the figure, we can observe that there are a total of 11 relationships within the document, out of which 9 were correctly predicted by our model, demonstrating its excellent performance. Specifically, the model successfully predicted all relationships between entity e0 and entities e1, e2, e3, and e4. However, there are still some incorrect predictions, such as the model inferring a relationship between e0 and e5, indicating that document-level relation extraction is a highly challenging task.

6 Discussion

Document-level relation extraction has seen significant advancements through the development of sequence-based and graph-based methods. These approaches [Zeng et al., 2021, Xu et al., 2021b, Sun et al., 2022] have set the foundation for extracting complex relationships between entities spread across multiple sentences in a document. Sequence-based methods primarily leverage pre-trained language models and neural architectures to process entire documents sequentially. For instance, [Ye et al., 2020] introduced mention reference prediction tasks during pre-training, which enabled models to learn coreference information effectively. Eberts and Ulges [Eberts and Ulges, 2021] proposed a multi-task learning approach based on reference parsing for end-to-end joint relation extraction, which involves extracting entity sets and learning their interactions within documents. These methods have demonstrated strong performance in capturing local context and entity interactions. Graph-based methods have gained prominence due to their ability to model long-range dependencies within a document. [Nan et al., 2020] utilized structural attention to capture global dependencies, while [Wang et al., 2021] employed discourse information to construct document-level graphs. These graphs facilitate the modeling of semantic dependencies between text units, improving reasoning between entities that are not directly connected within the text. Despite their success, both sequence-based and graph-based methods have inherent limitations that need to be addressed to further advance the field. Sequence-Based methods often overlook the structural information of documents, which can serve as valuable prior knowledge for understanding relationships between entities. The linear nature of sequence-based models may lead to the loss of contextual information spread across distant parts of the document, making it challenging to capture complex inter-sentence relationships. Graph-based methods excel at capturing global dependencies, they typically model documents in a coarse-grained manner. This approach can result in the loss of fine-grained information that is crucial for identifying relationships. The challenge lies in balancing the granularity of the graph representation to ensure that essential details are preserved without overwhelming the model with excessive complexity.

To overcome these limitations, future research in document-level relation extraction should focus on the following areas:

Interpretable Models: Developing models that can explain their extraction processes and results, allowing users to understand and trust the model's decisions.

User Controllability: Designing models that allow user intervention and correction during the extraction process, enhancing flexibility and accuracy in practical applications.

Integration of Structural Information: Enhancing sequence-based models with structural information, such as document layout and discourse markers, can improve the understanding of entity relationships. Hybrid models that combine the strengths of sequence-based and graph-based approaches may offer a more comprehensive solution.

Fine-Grained Graph Representations: Developing more sophisticated graph representations that capture both fine-grained and coarse-grained information is essential.

Techniques that dynamically adjust the granularity based on the context and importance of the information can help preserve critical details while maintaining model efficiency.

Advanced Regularization Techniques: Incorporating advanced regularization methods to capture relationship label dependencies and improve generalization is crucial. For example, our IFRTIN model introduces a regularizer for relationship label dependencies and an Adaptive Encouraging Loss to focus on well-classified instances, which significantly enhances performance.

Benchmarking and Evaluation: Establishing standardized benchmarks and evaluation protocols for document-level relation extraction can facilitate more consistent and comparable assessments of different models. This can drive the development of more robust and generalizable methods.

Efficient Model Architectures: Designing more efficient model architectures to reduce computational resource consumption and improve model performance in real-world applications.

7 Conclusion and Future Work

In this paper, we propose a IFRTIN model for document-level relation extraction. We consider syntax information, structural properties and discourse features simultaneously. Specifically, we design a document hierarchical GCN to facilitate cooperative learning by introducing distinct sorts of edges into the network's underlying document-level graph. Adaptive encoraging loss is introduced to detect possible relations in document. Besides, we employ a relation type dependency-based regularizer to help model capture the label dependencies information. We evaluate the performance of our model on three document-level relation extraction datasets. Specifically, IFRTIN outperforms existing models by achieving an F1 score improvement of 0.67% on Dataset DocRED, 1.2% on Dataset CDR, and 1.3% on Dataset GDA. The results show that our model can integrate helpful information related to entity pairs. In the future, we plan to extend IFRTIN to few-shot relation extraction.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 62376018).

References

[Agichtein and Gravano, 2000] Agichtein, E. and Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In Proceedings of the fifth ACM conference on Digital libraries, pages 85–94. https://doi.org/10.1145/336597.336644.

[Anitha et al., 2020] Anitha, J., Ting, I.-H., Agnes, S. A., Pandian, S. I. A., and Belfin, R. (2020). Social media data analytics using feature engineering. In Systems Simulation and Modeling for Cloud Computing and Big Data Applications, pages 29–59. Elsevier.

[Belfin et al., 2020] Belfin, R., Grace Mary Kanaga, E., and Kundu, S. (2020). Application of machine learning in the social network. Recent advances in hybrid metaheuristics for data clustering, pages 61–83.

[Beltagy et al., 2019] Beltagy, I., Lo, K., and Cohan, A. (2019). Scibert: A pretrained language model for scientific text. arXiv preprint arXiv:1903.10676.

[Christopoulou et al., 2019] Christopoulou, F., Miwa, M., and 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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4925–4936. https://doi.org/10.18653/v1/D19-1498.

[Devlin et al., 2018] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. Retrieved from https://arxiv.org/abs/1810.04805.

[Ding et al., 2023] Ding, X., Zhou, G., and Zhu, T. (2023). Multi-perspective context aggregation for document-level relation extraction. Applied Intelligence, 53(6):6926–6935.

[Dong and Xu, 2023] Dong, Y. and Xu, X. (2023). Relational distance and document-level contrastive pre-training based relation extraction model. Pattern Recognition Letters, 167:132–140.

[Eberts and Ulges, 2021] Eberts, M. and Ulges, A. (2021). An end-to-end model for entity-level relation extraction using multi-instance learning. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 3650–3660. http://doi.org/10.18653/v1/2021.eacl-main.319.

[Guo et al., 2021] Guo, Z., Nan, G., Lu, W., and Cohen, S. B. (2021). Learning latent forests for medical relation extraction. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, pages 3651–3657. http://doi.org/10.24963/ijcai.2020/505.

[Guo et al., 2019] Guo, Z., Zhang, Y., and Lu, W. (2019). Attention guided graph convolutional networks for relation extraction. arXiv preprint arXiv:1906.07510.

[Han and Wang, 2020] Han, X. and Wang, L. (2020). A novel document-level relation extraction method based on bert and entity information. IEEE Access, 8:96912–96919. https://doi.org/10.1109/ACCESS.2020.2996642.

[Ji et al., 2017] Ji, G., Liu, K., He, S., and Zhao, J. (2017). Distant supervision for relation extraction with sentence-level attention and entity descriptions. In Proceedings of the AAAI conference on artificial intelligence, volume 31.

[Ji et al., 2021] Ji, S., Pan, S., Cambria, E., Marttinen, P., and Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 33(2):494–514. https://doi.org/10.1109/TNNLS.2021.3070843.

[Jia et al., 2019] Jia, R., Wong, C., and Poon, H. (2019). Document-level n-ary relation extraction with multiscale 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. https://doi.org/10.18653/v1/N19-1370.

[Kolomiyets and Moens, 2011] Kolomiyets, O. and Moens, M.-F. (2011). A survey on question answering technology from an information retrieval perspective. Information Sciences, 181(24):5412–5434. https://doi.org/10.1016/j.ins.2011.07.047.

[Kuang et al., 2022] Kuang, H., Chen, H., Ma, X., and Liu, X. (2022). A keyword detection and context filtering method for document level relation extraction. Applied Sciences, 12(3):1599. https://doi.org/10.3390/app12031599.

[Li et al., 2020] Li, B., Ye, W., Sheng, Z., Xie, R., Xi, X., and Zhang, S. (2020). Graph enhanced dual attention network for document-level relation extraction. In Proceedings of the 28th international conference on computational linguistics, pages 1551–1560.

[Li et al., 2016] Li, J., Sun, Y., Johnson, R. J., Sciaky, D., Wei, C.-H., Leaman, R., Davis, A. P., Mattingly, C. J., Wiegers, T. C., and Lu, Z. (2016). Biocreative v cdr task corpus: a resource for chemical disease relation extraction. Database, 2016.

[Li et al., 2022] Li, R., Zhong, J., Xue, Z., Dai, Q., and Li, X. (2022). Heterogenous affinity graph inference network for document-level relation extraction. Knowledge-Based Systems, page 109146. https://doi.org/10.1016/j.knosys.2022.109146.

[Liu et al., 2023] Liu, H., Kang, Z., Zhang, L., Tian, L., and Hua, F. (2023). Document-level relation extraction with cross-sentence reasoning graph. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 316–328. Springer.

[Liu et al., 2020] Liu, X., Fan, J., Dong, S., 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. https://doi.org/10.2196/17644.

[Luo et al., 2017] Luo, B., Feng, Y., Wang, Z., Zhu, Z., Huang, S., Yan, R., and Zhao, D. (2017). Learning with noise: Enhance distantly supervised relation extraction with dynamic transition matrix. arXiv preprint arXiv:1705.03995.

[Nan et al., 2020] Nan, G., Guo, Z., Sekulić, I., and 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, pages 1546–1557. http://doi.org/10.18653/v1/2020.acl-main.141.

[Rink and Harabagiu, 2010] Rink, B. and Harabagiu, S. (2010). Utd: Classifying semantic relations by combining lexical and semantic resources. In Proceedings of the 5th international workshop on semantic evaluation, pages 256–259. Retrieved from https://aclanthology.org/S10-1057.pdf.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer. http://doi.org/10.1007/978-3-319-24574-4_28.

[Sahu et al., 2019] Sahu, S. K., Christopoulou, F., Miwa, M., and Ananiadou, S. (2019). Inter-sentence relation extraction with document-level graph convolutional neural network. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4309–4316. http://doi.org/10.18653/v1/P19-1423.

[Seymour et al., 2011] Seymour, T., Frantsvog, D., Kumar, S., et al. (2011). History of search engines. International Journal of Management & Information Systems (IJMIS), 15(4):47–58. https://doi.org/10.19030/ijmis.v15i4.5799.

[Shi and Lin, 2019] Shi, P. and Lin, J. (2019). Simple bert models for relation extraction and semantic role labeling. arXiv preprint arXiv:1904.05255. Retrieved from https://arxiv.org/abs/1904.05255.

[Soares et al., 2019] Soares, L. B., Fitzgerald, N., Ling, J., and Kwiatkowski, T. (2019). Matching the blanks: Distributional similarity for relation learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2895–2905. http://doi.org/10.18653/v1/P19-1279.

[Sun et al., 2022] Sun, Q., Zhang, K., Huang, K., Li, X., Zhang, T., and Xu, T. (2022). Enhanced graph convolutional network based on node importance for document-level relation extraction. Neural Computing and Applications, pages 1–11. https://doi.org/10.1007/s00521-022-07223-3.

[Tan et al., 2022] Tan, Q., He, R., Bing, L., and Ng, H. T. (2022). Document-level relation extraction with adaptive focal loss and knowledge distillation. In Findings of the Association for Computational Linguistics: ACL 2022, pages 1672–1681. http://doi.org/10.18653/v1/2022.findings-acl.132.

[Tang et al., 2020a] Tang, H., Cao, Y., Zhang, Z., Cao, J., Fang, F., Wang, S., and Yin, P. (2020a). Hin: Hierarchical inference network for document-level relation extraction. In Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part I 24, pages 197–209. Springer.

[Tang et al., 2020b] Tang, Y., Huang, J., Wang, G., He, X., and Zhou, B. (2020b). Orthogonal relation transforms with graph context modeling for knowledge graph embedding. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 2713–2722. http://doi.org/10.18653/v1/2020.acl-main.241.

[Vaswani et al., 2021] Vaswani, A., Ramachandran, P., Srinivas, A., Parmar, N., Hechtman, B., and Shlens, J. (2021). Scaling local self-attention for parameter efficient visual backbones. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12894–12904. http://doi.org/10.1109/CVPR46437.2021.01270.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. Retrieved from https://arxiv.org/abs/1810.04805.

[Veličković et al., 2017] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

[Verga et al., 2018] Verga, P., Strubell, E., and McCallum, A. (2018). Simultaneously self-attending to all mentions for full-abstract biological relation extraction. arXiv preprint arXiv:1802.10569.

[Wan et al., 2023] Wan, Q., Du, S., Liu, Y., Fang, J., Wei, L., and Liu, S. (2023). Document-level relation extraction with hierarchical dependency tree and bridge path. Knowledge-Based Systems, 278:110873.

[Wang et al., 2020a] Wang, D., Hu, W., Cao, E., and Sun, W. (2020a). 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. http://doi.org/10.18653/v1/2020.emnlp-main.303.

[Wang et al., 2019] Wang, H., Focke, C., Sylvester, R., Mishra, N., and Wang, W. (2019). Finetune bert for docred with two-step process. arXiv preprint arXiv:1909.11898. Retrieved from https://arxiv.org/abs/1909.11898.

[Wang et al., 2021] Wang, H., Qin, K., Lu, G., Yin, J., Zakari, R. Y., and Owusu, J. W. (2021). Document-level relation extraction using evidence reasoning on rst-graph. Knowledge-Based Systems, 228:107274.

[Wang et al., 2020b] Wang, J., Chen, X., Zhang, Y., Zhang, Y., Wen, J., Lin, H., Yang, Z., Wang, X., et al. (2020b). Document-level biomedical relation extraction using graph convolutional network and multihead attention: algorithm development and validation. JMIR Medical Informatics, 8(7):e17638. https://doi.org/10.2196/17638.

[Wang et al., 2016] Wang, L., Cao, Z., De Melo, G., and Liu, Z. (2016). Relation classification via multi-level attention cnns. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1298–1307. http://doi.org/10.18653/v1/P16-1123.

[Wu et al., 2019] Wu, Y., Luo, R., Leung, H. C., Ting, H.-F., and Lam, T.-W. (2019). Renet: A deep learning approach for extracting gene-disease associations from literature. In Research in Computational Molecular Biology: 23rd Annual International Conference, RECOMB 2019, Washington, DC, USA, May 5-8, 2019, Proceedings 23, pages 272–284. Springer.

[Xu et al., 2021a] Xu, B., Wang, Q., Lyu, Y., Zhu, Y., and Mao, Z. (2021a). 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. http://doi.org/10.1609/aaai.v35i16.17665.

[Xu et al., 2021b] Xu, W., Chen, K., and Zhao, T. (2021b). Discriminative reasoning for document-level relation extraction. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 1653–1663. http://doi.org/10.18653/v1/2021.findings-acl.144.

[Xu et al., 2021c] Xu, W., Chen, K., and Zhao, T. (2021c). Document-level relation extraction with reconstruction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 14167–14175. http://doi.org/10.1609/aaai.v35i16.17667.

[Yang et al., 2016] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pages 1480–1489. http://doi.org/10.18653/v1/N16-1174.

[Yao et al., 2019] Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., and 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, pages 764–777. http://doi.org/10.18653/v1/P19-1074.

[Ye et al., 2020] Ye, D., Lin, Y., Du, J., Liu, Z., Li, P., Sun, M., and Liu, Z. (2020). Coreferential reasoning learning for language representation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 7170–7186. http://doi.org/10.18653/v1/2020.emnlp-main.582.

[Zeng et al., 2021] Zeng, S., Wu, Y., and Chang, B. (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. http://doi.org/10.18653/v1/2021.findings-acl.47.

[Zeng et al., 2020] Zeng, S., Xu, R., Chang, B., and 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), pages 1630–1640. http://doi.org/10.18653/v1/2020.emnlp-main.127.

[Zhang et al., 2019] Zhang, N., Deng, S., Sun, Z., Wang, G., Chen, X., Zhang, W., and Chen, H. (2019). Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. arXiv preprint arXiv:1903.01306.

[Zhang et al., 2017] Zhang, Y., Zhong, V., Chen, D., Angeli, G., and Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In Conference on Empirical Methods in Natural Language Processing. http://doi.org/10.18653/v1/D17-1004.

[Zhou et al., 2016] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., and Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers), pages 207–212.

[Zhou et al., 2021] Zhou, W., Huang, K., Ma, T., and Huang, J. (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. http://doi.org/10.1609/aaai.v35i16.17717.

[Zhu et al., 2019] Zhu, H., Lin, Y., Liu, Z., Fu, J., Chua, T.-S., and Sun, M. (2019). Graph neural networks with generated parameters for relation extraction. arXiv preprint arXiv:1902.00756.