Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

KnowRA: Knowledge Retrieval Augmented Method for Document-level Relation Extraction with Comprehensive Reasoning Abilities

Chengcheng Mai[1,2,3], **Yuxiang Wang**[2,3], **Ziyu Gong**[2,3], **Hanxiang Wang**[2,3] and **Yihua Huang**[2,3]
[1] Nanjing Normal University, School of Computer Science and Electronic Informatics / School of Artificial Intelligence
[2] State Key Laboratory for Novel Software Technology at Nanjing University
[3] Nanjing University, School of Computer Science
maicc@njnu.edu.cn, yuxiangwang@smail.nju.edu.cn, ziyugong@smail.nju.edu.cn, wanghanxiang@smail.nju.edu.cn, yhuang@nju.edu.cn

Abstract

Document-level relation extraction (Doc-RE) aims to extract relations between entities across multiple sentences. Therefore, Doc-RE requires more comprehensive reasoning abilities like humans, involving complex cross-sentence interactions between entities, contexts, and external general knowledge, compared to the sentence-level RE. However, most existing Doc-RE methods focus on optimizing single reasoning ability, but lack the ability to utilize external knowledge for comprehensive reasoning on long documents. To solve these problems, a knowledge retrieval augmented method, named KnowRA, was proposed with comprehensive reasoning to autonomously determine whether to accept external knowledge to assist Doc-RE. Firstly, we constructed a document graph for semantic encoding and integrated the co-reference resolution model to augment the co-reference reasoning ability. Then, we expanded the document graph into a document knowledge graph by retrieving the external knowledge base for common-sense reasoning and a novel knowledge filtration method was presented to filter out irrelevant knowledge. Finally, we proposed the axis attention mechanism to build direct and indirect associations with intermediary entities for achieving cross-sentence logical reasoning. Extensive experiments conducted on two datasets verified the effectiveness of our method compared to the state-of-the-art baselines. Our code is available at https://anonymous.4open.science/r/KnowRA.

1 Introduction

The document-level relation extraction (Doc-RE) task aims to extract pre-defined relation triples from documents containing multiple sentences. Compared with the existing sentence-level RE task, Doc-RE is not only more complicated but also more fundamental for real-world applications, like retrieval augmented generation (RAG) method for large language model (LLM) [Dai *et al.*, 2022; Yang *et al.*, 2023; Xu *et al.*, 2024], automatic question and answering [Ding *et al.*, 2022; Orogat and El-Roby, 2022], and event extraction [Man *et al.*, 2022; Huang *et al.*, 2023; Yuan *et al.*, 2023].

Obviously, Doc-RE models need to have comprehensive reasoning abilities for Doc-RE. Previous studies divided these abilities into four categories [Yao *et al.*, 2019]: pattern recognition, co-reference reasoning, common-sense reasoning, and logical reasoning, as shown in Figure 1.

Reasoning Type	Examples
Pattern recognition	[S1] **Niklas Bergqvist** (born **6 October 1962** in Stockholm) is a Swedish songwriter ... [S2] ... **Relation:** < **Niklas Bergqvist**, ***date_of_birth***, **6 October 1962** >
Co-reference reasoning	[S1] **Dwight Tillery** is an American politician of the Democratic Party ...[S2] ... [S3] ***He*** also holds a law degree from the **University of Michigan Law School**. **Dwight Tillery** *Co-reference* ***He*** **R1:** ***educated_at*** **University of Michigan Law School** ***<head entity>*** **R2:** ***educated_at*** ***<tail entity>***
Common-sense reasoning	[S1] The news that British's **Prince Harry** is engaged to his partner **Meghan Markle** has attracted widespread attention from England, **America** and around the world. [S2] … [S10] **Meghan Markle's** parents **Thomas Markle** and Doria Ragland said in a statement: ... **Prince Harry** **R1:** ***spouse*** **Megan Markle** **R2:** ***parent*** **Thomas Markle** **R3:** ***country_of_citizenship*** ***Reasoned from external knowledge*** **America**
Logical reasoning	[S1] **Eminem Show** is the fourth studio album by American rapper Eminem, rel-eased on **May 26 2002** by …[S2] **It includes** the commercially successful singles "**Without Me**", …, [S3] … **Eminem Show** **R1:** ***publication_date*** **May 26 2002** ***Logic reasoned with intermediary words: {Eminem Show, It includes…}*** **R3:** ***publication_date*** **It** **R2:** ***part_of*** **Without Me**

Figure 1: Different reasoning abilities for Doc-RE. Relation R3 in common-sense reasoning can hardly be extracted from the original document but can be retrieved from external knowledge. Relation R3 in logical reasoning can be reasoned by intermediary words.

However, the existing Doc-RE models were mostly optimized for partial reasoning abilities but lacked comprehensive reasoning abilities. For pattern recognition, most recent methods constructed graph structures [Xu *et al.*, 2021c; Wei and Li, 2022; Peng *et al.*, 2022; Zhang *et al.*, 2023b] to establish long-distance associations between entities. For logical reasoning, several models used bridge/intermediary entities or evidence sentences to establish relations between two entities that are not directly connected [Zhang *et al.*, 2023a; Huang *et al.*, 2024; Xu *et al.*, 2022]. Considering that entity mentions often appear in the form of co-reference pronouns in the document, methods based on co-reference reasoning have been proposed [Ye *et al.*, 2020] and focused on identifying co-reference pronouns of entities, which were used as bridge entities to establish indirect relations between entities.

Also, the Doc-RE model requires the ability of commonsense reasoning because some relations cannot be directly inferred from the document itself, and external knowledge is needed to assist in Doc-RE. However, few existing Doc-RE models combined external knowledge for common sense reasoning [Wang *et al.*, 2022a].

To summarize, the main challenges for Doc-RE lie in: 1) How to integrate the comprehensive reasoning abilities required by Doc-RE. 2) How to represent and integrate external knowledge with the internal semantics of the document. 3) How to autonomously determine whether to accept external knowledge, considering that external knowledge may be lagging, one-sided, or even wrong.

To solve above mentioned problems, we proposed a comprehensive reasoning method for Doc-RE with knowledge retrieval augmentation and filtration, as shown in Figure 2. Firstly, a heterogeneous multi-level document graph was constructed for semantic encoding of entities, mentions, and sentences of the document. Then, a pre-trained co-reference resolution model was introduced to establish associations between entities and their corresponding pronouns for co-reference reasoning. Moreover, the document graph was extended with the retrieved external knowledge for commonsense reasoning. On this basis, a novel knowledge filtration method was proposed to determine whether to accept external knowledge. Finally, an axial attention method was proposed to use intermediary entities for realizing logical reasoning across multiple sentences.

The main contributions of our work are three folds:

- A comprehensive method for Doc-RE was proposed by achieving document semantic encoding with a multi-layer heterogeneous document graph, integrating the co-reference resolution model, injecting external knowledge, and introducing the axial attention mechanism.
- A knowledge augmentation and filtration method was proposed to judge whether to accept external knowledge, for filtering out irrelevant knowledge.
- Extensive experiments performed on two public datasets demonstrated the superiority of our method compared to the state-of-the-art (SOTA) baselines.

2 Our Methodology

2.1 Construction of document graph

Firstly, a pre-trained language model was employed to perform semantic encoding operations on the input document. Then, a multi-level heterogeneous document graph was present to model the connections among different entities, mentions, sentences in a document, defined as follows:

***Definition* 1. Multi-level Heterogeneous Document Graph (MHDG)**. MHDG=$< V, E >$, where $V = \{v | v \in V^M \bigcup V^S \bigcup V^D\}$ represents the node set and $E = \{< v_i, v_j > | v_i, v_j \in V, i \neq j\}$, which represents edges between nodes v_i and v_j.

According to definition 1, we defined three types of node in MHDG: Mention node (V^M), Sentence node (V^S), and Document node (V^D). Then, four types of edges were given:

1) **Document-Sentence Edge**: $E^{DS} = \{< v_i^D, v_j^S > | v_i^D \in V^D, v_j^S \in V^S\}$. The document node and all sentence nodes are connected through these edges.
2) **Sentence-Sentence Edge**: $E^{SS} = \{< v_i^S, v_j^S > | v_i^S, v_j^S \in V^S, i \neq j\}$. Two adjacent sentences are connected by sentence-sentence edges.
3) **Mention-Sentence Edge**: $E^{MS} = \{< v_i^M, v_j^S > | v_i^M \in V^M, v_j^S \in V^S\}$, connecting mention nodes and the sentence node appearing with the same one sentence.
4) **Mention-Mention Edge**: Two mention nodes are connected by the Mention-Mention Edge (MME), which can be further divided into two categories: Co-Occurrence Mention-Mention Edge (CO-MME) means different mentions appear in the same sentence, formalized as: $E^{CO-MME} = \{< v_i^M, v_j^M > | v_i^M, v_j^M \in V^M\}$. Co-Reference Mention-Mention Edge (CR-MME) means that different mentions refer to the same entity, formalized as: $E^{CR-MME} = \{< v_i^M, v_j^M > | v_i^M, v_j^M \in V^M\}$.

Then, the semantic representation was performed on the MHDG with a graph-based neural network, denoted as:

$$
\begin{aligned}
H_{\{t_1,\ldots,t_L\}} &= Encoder(\mathcal{T}_{input}) \\
&= Encoder(t_1, \ldots, t_L) = [h_1, \ldots, h_L]
\end{aligned} \tag{1}
$$

where $\mathcal{T}_{input}$ represents the input tokenized sequence, L represents the length of the document. The representations of three kinds of nodes are as follows:

$$H^D = h_{[CLS]} \tag{2}$$

$$H^{m_j^i} = h_{P_e^{i,j}} = h_{``*"} \tag{3}$$

$$H^{S_i} = \log \sum_{j=1}^{|S_i|} \exp(h_j) \tag{4}$$

where $H^D \in R^d$ represents the d dimension semantic representation of the document node. $H^{m_j^i}$ represents the semantic embedding of mention node for the *j*-th mention of the *i*-th entity, and $P_e^{i,j}$ represents the position of the *j*-th mention of the *i*-th entity, and "*" denotes for the special symbol placed before the mention m_j^i. H^{S_i} represents the semantic representation of sentence node S_i.

Then, the Graph Attention Network (GAT) was used to calculate association score $AS(e_{i,j})$ for each edge between node *i* and *j*:

$$AS(e_{i,j}) = LeakyReLU(W_\alpha[W_{\beta_1} h_i \oplus W_{\beta_2} h_j]) \tag{5}$$

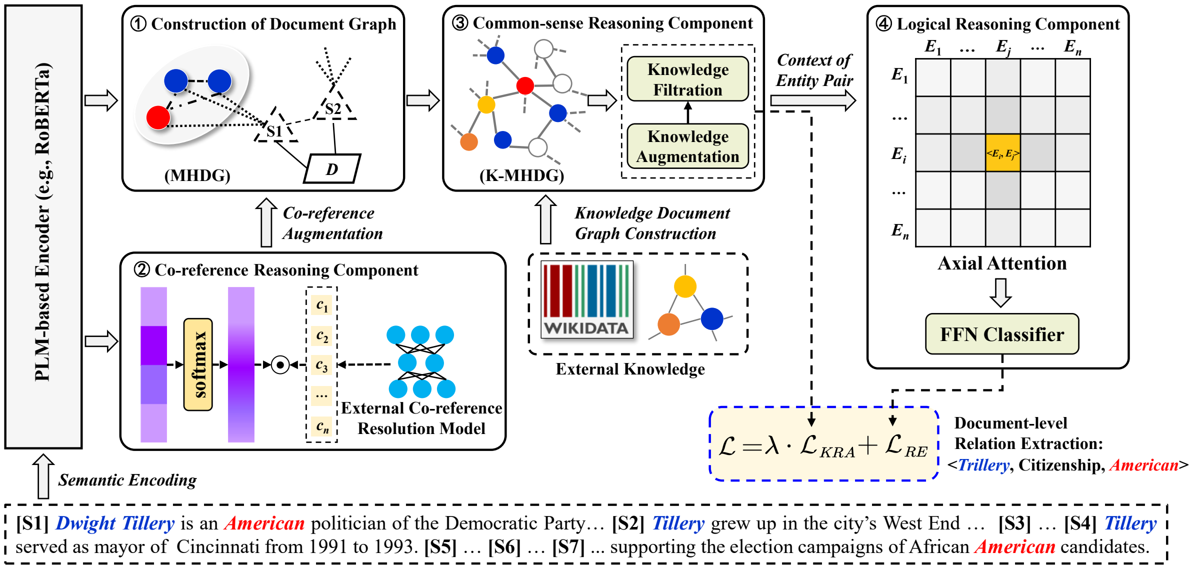

Figure 2: Overview of the proposed comprehensive reasoning method with knowledge augmentation and filtration for Doc-RE.

where $\oplus$ represents the concatenation operation. The updated node representations based on MHDG are as follows:

$$h_i = \sum_{j \in N(i)} a_{i,j}(W_\beta h_j) \tag{6}$$

$$a_{i,j} = softmax(AS(e_{i,j})) \tag{7}$$

where $W_\alpha \in R^d$ and $W_\beta \in R^{d\times d}$ are trainable parameters. $N(i)$ represents the number of adjacent nodes to node i.

2.2 Co-reference reasoning component for Doc-RE

To solve the co-reference reasoning problem, we firstly pre-trained a co-reference resolution model, M_{coref}, and then used it to identify the co-reference pronouns, as follows:

$$C_i = M_{coref}(\mathcal{T}_{input}, E_i) = \{c_k^i\}_{k=1}^{n_i^c} \tag{8}$$

where C_i represents the set of identified co-reference pronouns and n_i^c represents the number of the co-reference pronouns. Then, we used attention matrix between each input token and recognized co-reference pronouns to update semantic representations of co-reference pronouns, as follows:

$$A = MultiHeadAttention(Encoder(\mathcal{T}_{input})) \tag{9}$$

$$Q_i = \sum_{h=1}^{H}\left(\frac{1}{n_i^c}\sum_{k=1}^{n_i^c} A_i^{h,k}\right) \tag{10}$$

$$h_{c_k^i} = Q_i'[P_c^{i,k}] \cdot h_{P_c^{i,k}} \tag{11}$$

where $A \in R^{H\times L\times L}$ is the attention matrix of all input tokens, H represents the number of attention heads, and L represents the length of input tokens. $Q_i \in R^L$, represents the averaged vector of attention value of entity E_i to its co-reference pronouns. $A_i^{h,k}$ represents the h-th attention head of entity E_i to its k-th co-reference word. $P_c^{i,k}$ represents the position of each co-reference pronoun.

Finally, we obtained the semantic representation of c_k^i as shown in formula (11), where $Q_i'[P_c^{i,k}] \in R^{n_{i,k}^c}$ represents the normalized attention value of entity E_i to its k-th co-reference pronoun. $h_{P_c^{i,k}}$ represents the semantic embedding of the co-reference pronoun $c_{i,k}$.

2.3 Common-sense reasoning with knowledge augmentation and filtration

Construction of knowledge-augmented document graph

To introduce external knowledge for common-sense reasoning, we further constructed a knowledge-augmented document graph, denoted as K-MHDG.

***Definition* 2. Knowledge-retrieval-augmented Multi - level heterogeneous document graph (K-MHDG).** K-MHDG=$< V, E \cup E_{know} >$, where $V = \{v | v \in V^{\{M,S,D\}}\}$ represents the entity node set, $E \subseteq \{< v_i, v_j > | v_{i,j} \in V, i \neq j\}$ represents the relation edges between node v_i and v_j in MHDG, and E_{know} represents the newly added edges, consisting of relations between entities, denoted as r_{know}, retrieved from the external knowledge base, formalized as: $E_{know} \subseteq \{< v_m, v_n > | \exists e^{r_{know}} =< v_m, v_n >\in hasRel(Q_{id}^{E_i}, r_{know}, Q_{id}^{E_j}), v_{m,n} \in V^M\}$.

We selected Wikidata as the external knowledge base to construct the K-MHDG. By utilizing the interface $getQid(EntityName)$ provided by Wikidata, all possible entity identifiers, i.e., $Qids$, were obtained based on the corresponding entity name. Then, based on the $hasRel(Q_{id}^t, r_{know}, Q_{id}^h)$ interface in Wikidata, we can query all relation triples with the provided entity pairs. Finally, the newly retrieved relation $e^{r_{new}} =< v_{Ent_i}^M, v_{Ent_j}^M >$ would be added to E_{know} and the MHDG was augmented to K-MHDG.

Semantic encoding for the external knowledge

Based on K-MHDG, the representation of entity node E_i was updated as follows:

$$h_{E_i} = \log\left(\frac{1}{n_i^m}\sum_{j=1}^{n_i^m}\exp\left(h_{m_j^i}\right) \oplus \frac{1}{n_i^c}\sum_{k=1}^{n_i^c}\exp\left(h_{c_k^i}\right)\right) \tag{12}$$

where $\oplus$ represents the concatenation operation, $h_{m_j^i}$ represents the embedding of mention m_j^i and n_i^m represents the number of mentions related to E_i. $h_{c_k^i}$ and n_i^c refer to the semantic embedding and number of the co-reference pronouns for entity E_i, respectively.

External knowledge filtration method

Considering that the external knowledge introduced is not always correct, we further proposed a new confidence-score-based knowledge filtration method to help our model autonomously determine whether to accept external knowledge.

The confidence score for each edge, i.e., $e_{i,j} =< E_i, r_k, E_j >$, was denoted as $\tau_{i,j}$ and calculated as:

$$\tau_{i,j}^k = f_{conf}(< E_i, r_k, E_j >) = h_{E_i}^T \cdot diag(r_k) \cdot h_{E_j} \tag{13}$$

where $diag(\cdot)$ function represents the diagonal matrix with the vector r_k as the diagonal element. The confidence score represents the likelihood of a relation existing between entity E_i and E_j. After calculating the confidence score, the entity representation of entity E_i was also updated as follows:

$$h'_{E_i} = \sum_{j\in N(i)}\sum_{k\in r_{i,j}} \sigma(\tau_{i,j}^k)(h_{E_j} \cdot r_k) \tag{14}$$

where $\sigma(\cdot)$ represents the sigmoid function, which is used to convert confidence score into probabilities with values between 0 and 1. $r_{i,j}$ represents the set of relations that exist between entities E_i and E_j.

Then, we used the correct relation label set provided by the annotated training set to train the model in a way that increases the confidence score for correct relations and decreases the confidence score for incorrect relations. The optimization function is:

$$\mathcal{L}_{KRA} = -\frac{1}{N}\sum_{n=1}^{N}(y_n \cdot \log(\sigma(\tau_{i,j}^k)) + (1-y_n)\cdot\log(\sigma(1-\tau_{i,j}^k))) \tag{15}$$

where N represents the edge number in K-MHDG, and y_n represents the annotated relation label in the training set, which value is 1 (correct relation) or 0 (incorrect relation).

2.4 Logical reasoning with axial attention

Semantic fusion based on common context of entity pair

Because Doc-RE is implemented based on the entity pairs, it is necessary to integrate context information into entity pairs for logical reasoning, not just context of individual entities.

Firstly, we used the attention mechanism proposed to obtain the context of the common concern of entity E_i and E_j, that is, the relevant context representation of entity pair $< E_i, E_j >$. The formula was given as follows:

$$C^{i,j} = \frac{Q_i \times Q_j}{Q_i^T \cdot Q_j} H^D \tag{16}$$

where symbol "$\times$" represents the outer product of embeddings, and "$\cdot$" represents the dot product of embeddings.

Then, the semantic embedding for head and tail entities fused with the context of entity pair $< E_i, E_j >$ was calculated as:

$$z_i = \tanh(W_i h_{e_i} + W_c C^{i,j}) \tag{17}$$

$$z_j = \tanh(W_j h_{e_j} + W_c C^{i,j}) \tag{18}$$

$$z_{i,j} = z_i^T W_b z_j + b \tag{19}$$

where z_i and z_j represent the entity embedding fused with the context information. $W_* \in R^d$ represents the trainable model parameters. Finally, the semantic representation of entity pair $< E_i, E_j >$, denoted as $z_{i,j}$, was obtained.

Axial-attention-based method for logical reasoning across sentences

We arranged all entity pair representations $z_{i,j}$ in the document into an entity pair matrix of $N \times N$, where $N = |V|$ represents the number of entities. If there is an intermediary entity E_k between entity pair $< E_i, E_j >$, the purpose of our proposed method is to establish a multi-hop logical reasoning model by integrating semantic representations of all intermediary entities and the corresponding entity pairs.

In the entity pair matrix, the semantic information of the horizontal intermediary entity pair $< E_i, E_k | k = 1, \ldots, N >$ in the same row was firstly fused into $< E_i, E_j >$, calculated as:

$$G_{<E_i,E_j>}^{row} = z_{i,j} + \sum_{k=1,\ldots,N} softmax_k(q_{i,j}^T k_{i,k}) v_{i,k} \tag{20}$$

where $q_{i,j}, k_{i,j}, v_{i,j} = W_q z_{i,j}, W_k z_{i,j}, W_v z_{i,j}$ is the query, key, and value vector obtained by linear transformation of $z_{i,j}$. $W_q, W_k, W_v \in R^{d\times d}$ are query, key, and value vectors obtained by linear transformation.

Then, the similar operation was performed on all vertical entity pairs in the same column and calculated as follows:

$$G_{<E_i,E_j>}^{col} = z_{i,j} + \sum_{k=1,\ldots,N} softmax_k(q_{i,j}^T k_{k,j}) v_{k,j} \tag{21}$$

Finally, all intermediary entity pairs $< E_i, E_k >$ and $< E_k, E_j >$ were fused and the final representation for $< E_i, E_j >$ was obtained by :

$$G_{<E_i,E_j>} = G_{<E_i,E_j>}^{row} + G_{<E_i,E_j>}^{col} \tag{22}$$

Adaptive Relation Extraction Loss for Doc-RE

Considering Doc-RE is a multi-label classification task, we adopted adaptive threshold loss function [Zhou *et al.*, 2021; Guo *et al.*, 2023] with a threshold class, denoted as TH, to adaptively separate positive ($\boldsymbol{R}^+$) and negative ($\boldsymbol{R}^-$) relations:

$$\mathcal{L}_{RE} = -\log\left(\frac{\exp\left(L_{i,j}^{TH}\right)}{\sum_{r' \in \boldsymbol{R}^- \cup \{TH\}} \exp\left(L'_{i,j}\right)}\right) - \sum_{i \neq j} \sum_{r \in \boldsymbol{R}^+} \log\left(\frac{\exp\left(L_{i,j}^{r}\right)}{\sum_{r' \in \boldsymbol{R}^+ \cup \{TH\}} \exp\left(L_{i,j}^{r'}\right)}\right) \tag{23}$$

where $L_{i,j}$ represents the probability that entity pair $< E_i, E_j >$ belongs to each predefined relation type:

$$L_{i,j} = W_l G_{<E_i,E_j>} + b_l \tag{24}$$

Through the joint training of the loss function $\mathcal{L}_{RE}$ and $\mathcal{L}_{KRA}$, the final optimization function for Doc-RE is:

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{KRA} + \mathcal{L}_{RE} \tag{25}$$

where λ is the pre-defined hyper-parameter.

3 Experiments and Analyses

3.1 Experimental Settings

Datasets. We evaluated our model on two public datasets for document-level RE. **Re-DocRED** [Tan *et al.*, 2022b] is a high-quality revised version of DocRED [Yao *et al.*, 2019]. Re-DocRED corrects the false negatives problem in dataset DocRED and contains 3,053 documents for training, 500 for development, and 500 for the test set. **DWIE** [Zaporojets *et al.*, 2021] is sampled and annotated from the news website Deutsche Welle, containing 602, 98, 99 documents for training, development, and testing, respectively, with 43,373 entities, 21,749 relational facts, and 65 relation types.

Evaluation metrics. We used the micro F1 (**F1**), ignore F1 (**Ign F1**), Intra F1, and Inter F1 as the metrics for model performance, following previous work [Wang *et al.*, 2022b]. **Ign F1** is a revised version of F1, which excludes the shared relations between the training and development/test set. **Intra F1** is used to evaluate F1 of relation triples that appear in the same sentence. **Inter F1** is used to evaluate F1 of cross-sentence relation triples.

3.2 Baselines

According to different model structures, the following models were selected for performance comparison.

- Sequence-based models: These models used different deep neural structures for relation extraction, including convolution neural network (CNN), Bi-LSTM, and Context-Aware LSTM [Yang *et al.*, 2023].
- Graph-based models: These models used graph neural network for Doc-RE, including GAIN [Zeng *et al.*, 2020], SIRE [Zeng *et al.*, 2021], DRN [Xu *et al.*, 2021b], SagDRE [Wei and Li, 2022].
- Transformer-based models: These Transformer-based models [Vaswani *et al.*, 2017] include SSAN [Xu *et al.*, 2021a], ATLOP [Zhou *et al.*, 2021], ATLOP-MILR [Fan *et al.*, 2022], DocuNET [Zhang *et al.*, 2021a], KD-DocRE [Tan *et al.*, 2022a], UGDRE [Sun *et al.*, 2023], and JMRL-DREEAM [Qi *et al.*, 2024]. It should be noted that documents in dataset DWIE are quite long, with over 50% of documents exceeding 768 tokens, and the maximum length reaches 2560 tokens. Therefore, for dataset DWIE, we replaced RoBERTa-*large* with LongFormer [Beltagy *et al.*, 2020] as the backbone model, which supports a maximum of 4096 tokens. For dataset Re-DocRED, baseline models used RoBERTa-*large* as the cornerstone model.
- LLM-based models: We used 13B LLAMA-2[1] model as the large language model for Doc-RE, finetuned with LoRA[2].

Considering that these baseline models used different datasets and cornerstone models in their original papers, we tried our best to rerun and fine-tune these models for fair comparison. For example, the rerun model JMRL-DREEAM even achieved better performance on dataset Re-DocRED compared to its original paper.

3.3 Main Results

Table 1 listed the performance of models on two datasets. We observed that: 1) Our KnowRA model outperformed other baseline models in almost metrics on two datasets. For dataset Re-DocRED and DWIE, KnowRA surpassed the existing SOTA model, JMRL-DREEAM, by 0.28 and 1.13 in F1 score, respectively. These experiments proved the superiority of our model for Doc-RE. 2) In terms of Intra- and Inter-F1, our model achieved the second-best results in dataset Re-DocRED and the best performance in dataset DWIE. These experimental results proved the advantages of our model in cross-sentence relation extraction. Meanwhile, for dataset DWIE with longer document length, our method achieved better performance compared to the SOTA models, i.e., KD-DocRE and JMRL-DREEAM, which indicates that our model has stronger semantic reasoning ability for extracting long-distance document-level relations. 3) For both datasets, our model outperformed the LLaMA-2-based model in all metrics, which implies that large language models need to improve their effectiveness through relevant optimization technologies when applied to downstream tasks.

3.4 Ablation Experiments

Ablation experiments were shown in Table 2. For dataset Re-DocRED, after removing the document graph (denoted as "*-w/o graph*") and axial attention (denoted as "*-w/o axial atten.*"), the model performance decreased by 1.07 and 1.01 in F1 score, respectively. When both the two components were removed (denoted as "*-w/o graph+axial*"), the performances on F1 decreased by 1.28. The removal of the co-reference reasoning component (denoted as "*-w/o coref*") also leads to a drop of 0.70 in terms of F1.

In addition, the common-sense reasoning component based on the knowledge-retrieval-augmentation (denoted as "*-w/o knowAug.*", representing the removal of the K-MHDG) is the most important component, as the F1 score decreased the most (-1.10 on Re-DocRED and -1.02 on DWIE) when it was

[1]https://github.com/meta-llama/llama.

[2]https://github.com/microsoft/LoRA.

Models	Re-DocRED				DWIE			
	Ign F1	F1	Intra-F1	Inter-F1	Ing F1	F1	Intra-F1	Inter-F1
CNN [Yao *et al.*, 2019]	54.28$_{\pm0.33}$	56.20$_{\pm0.35}$	59.61$_{\pm0.62}$	53.54$_{\pm0.67}$	40.24$_{\pm0.19}$	51.84$_{\pm0.19}$	51.84$_{\pm0.19}$	51.31$_{\pm0.63}$
BiLSTM [Yao *et al.*, 2019]	58.08$_{\pm0.38}$	60.03$_{\pm0.30}$	62.99$_{\pm0.07}$	57.70$_{\pm0.51}$	55.07$_{\pm0.18}$	66.21$_{\pm0.25}$	68.58$_{\pm0.28}$	64.78$_{\pm0.38}$
Context-Aware [Yao *et al.*, 2019]	58.29$_{\pm0.26}$	60.19$_{\pm0.21}$	63.18$_{\pm0.32}$	57.82$_{\pm0.17}$	56.80$_{\pm0.18}$	66.05$_{\pm0.17}$	69.24$_{\pm0.39}$	63.54$_{\pm0.43}$
GAIN [Zeng *et al.*, 2020]	73.55$_{\pm0.15}$	74.89$_{\pm0.21}$	77.46$_{\pm0.31}$	72.68$_{\pm0.15}$	63.49$_{\pm0.57}$	68.62$_{\pm0.32}$	68.97$_{\pm0.34}$	68.28$_{\pm0.43}$
SIRE [Zeng *et al.*, 2021]	73.10$_{\pm0.40}$	74.55$_{\pm0.38}$	77.28$_{\pm0.46}$	72.22$_{\pm0.63}$	63.01$_{\pm0.27}$	68.31$_{\pm0.22}$	68.07$_{\pm0.29}$	67.74$_{\pm0.37}$
DRN [Xu *et al.*, 2021b]	72.37$_{\pm0.23}$	73.28$_{\pm0.22}$	76.28$_{\pm0.15}$	70.58$_{\pm0.34}$	63.13$_{\pm0.32}$	69.32$_{\pm0.23}$	71.51$_{\pm0.23}$	67.07$_{\pm0.38}$
SagDRE [Wei and Li, 2022]	73.44$_{\pm0.29}$	74.56$_{\pm0.23}$	76.99$_{\pm0.18}$	72.46$_{\pm0.38}$	63.37$_{\pm0.27}$	69.61$_{\pm0.31}$	69.84$_{\pm0.26}$	68.98$_{\pm0.35}$
SSAN [Xu *et al.*, 2021a]	72.64$_{\pm0.32}$	73.88$_{\pm0.28}$	75.28$_{\pm0.38}$	72.20$_{\pm0.35}$	76.26$_{\pm0.24}$	81.06$_{\pm0.10}$	86.10$_{\pm0.24}$	77.09$_{\pm0.39}$
ATLOP [Zhou *et al.*, 2021]	76.85$_{\pm0.29}$	77.48$_{\pm0.30}$	79.54$_{\pm0.28}$	75.65$_{\pm0.34}$	78.67$_{\pm0.24}$	83.21$_{\pm0.19}$	87.25$_{\pm0.11}$	80.84$_{\pm0.32}$
ATLOP-MILR [Sun *et al.*, 2023]	75.99$_{\pm0.24}$	76.68$_{\pm0.17}$	78.95$_{\pm0.21}$	74.69$_{\pm0.19}$	79.95$_{\pm0.29}$	84.66$_{\pm0.41}$	87.04$_{\pm0.68}$	82.29$_{\pm0.14}$
DocuNET [Zhang *et al.*, 2021a]	77.19$_{\pm0.22}$	77.88$_{\pm0.26}$	79.89$_{\pm0.21}$	76.11$_{\pm0.41}$	79.41$_{\pm0.21}$	84.18$_{\pm0.13}$	87.88$_{\pm0.18}$	80.84$_{\pm0.32}$
KD-DocRE [Tan *et al.*, 2022a]	77.34$_{\pm0.33}$	78.12$_{\pm0.30}$	80.19$_{\pm0.29}$	76.31$_{\pm0.35}$	80.22$_{\pm0.24}$	84.87$_{\pm0.19}$	<u>88.01$_{\pm0.35}$</u>	81.57$_{\pm0.32}$
UGDRE [Sun *et al.*, 2023]	78.15$_{\pm0.30}$	78.87$_{\pm0.27}$	**81.05$_{\pm0.24}$**	76.93$_{\pm0.35}$	72.08$_{\pm0.59}$	76.35$_{\pm0.61}$	79.51$_{\pm0.72}$	73.44$_{\pm0.60}$
JMRL-DREEAM [Qi *et al.*, 2024]	**78.57$_{\pm0.06}$**	<u>78.97$_{\pm0.06}$</u>	80.13$_{\pm0.29}$	**77.96$_{\pm0.18}$**	<u>80.60$_{\pm0.55}$</u>	<u>85.27$_{\pm0.36}$</u>	87.88$_{\pm0.34}$	82.68$_{\pm0.72}$
LLaMA-2	46.84$_{\pm0.40}$	47.02$_{\pm0.43}$	45.66$_{\pm0.88}$	72.42$_{\pm0.29}$	80.57$_{\pm0.69}$	82.37$_{\pm0.58}$	79.74$_{\pm0.88}$	<u>82.84$_{\pm0.31}$</u>
KnowRA **(ours)**	<u>78.39$_{\pm0.32}$</u>	**79.25$_{\pm0.16}$**	<u>80.42$_{\pm0.33}$</u>	<u>76.99$_{\pm0.28}$</u>	**81.48$_{\pm0.32}$**	**86.40$_{\pm0.22}$**	**88.17$_{\pm0.09}$**	**83.83$_{\pm0.45}$**

Table 1: Model performance in two datasets. **Bold** denotes the best result, <u>underline</u> denotes the second best result. We ran each experiment five times, using different random seeds, and reported their performance and standard deviation.

Variant Models	Re-DocRED		DWIE	
	F1		F1	
KnowRA	**79.25**	Δ	**86.40**	Δ
-w/o graph	78.18	-1.07	85.65	-0.75
-w/o coref	78.55	-0.70	86.02	-0.38
-w/o knowAug.	78.15	-1.10	85.38	-1.02
-w/o knowFilt.	78.28	-0.97	85.61	-0.79
-w/o axial atten.	78.24	-1.01	85.57	-0.83
-w/o graph+axial	77.97	-1.28	85.25	-1.15

Table 2: Ablation experimental results in Re-DocRED and DWIE.

removed. Also, when the knowledge filtration method (denoted as "*-w/o knowFilt.*") was removed from our complete model, the F1 score decreased by 0.97 and 0.79 on dataset Re-DocRED and DWIE, respectively, which proves that the filtration of the external knowledge is necessary for filtering out noise information and boosting performance. For dataset DWIE, we observed similar phenomena, which verifies the effectiveness of our comprehensive reasoning components.

3.5 Effects of Intra- and Inter-sentence Reasoning

The number of sentence intervals was used to present the distance between the head and tail entities in a relation triple. According to the results shown in Figure 3, we found that: 1) When the head and tail entities are in the same sentence (intervals = 0), our model performed better than the baselines. 2) When the head and tail entities are from different sentences (intervals > 0), our model was also superior to other baselines. 3) The performance of all models gradually decreased with the increase of sentence intervals, which indicates that our model performed well in both intra- and inter-senetence reasoning.

3.6 Effects of Context Length

We conducted experiments to quantitatively evaluate the impact of the context length on the performance of Doc-RE. For dataset DWIE, the length of most documents (94.81%) exceeds the maximum length supported by the RoBERTa$_{large}$-based encoder. To this end, we replaced the original encoders with longFormer [Beltagy *et al.*, 2020].

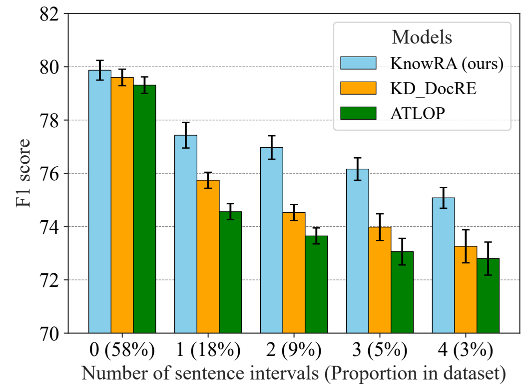

Figure 3: Model performance for Doc-RE changing with the length of sentence intervals on dataset Re-DocRED. For example, 1 (18%) represents the case where the length of interval sentences between entities in a relation is 1, and its proportion in the dataset is 18%.

The experimental results are shown in Figure 4 and we found that: 1) Longer context can greatly improve the performance of Doc-RE. 2) As the context length gradually increases, the performance of all models first improves, and then the growth rate gradually slows down. Similar trends were also observed on other baselines. These results proved that our model can outperform SOTA models in different context length, which indicates a good scalability for our model.

3.7 Impact of Knowledge-augmented Method

We conducted a hyper-parameter analysis on weight λ of the $\mathcal{L}_{KRA}$ loss function to quantitatively evaluate the impact of the proposed knowledge-retrieval-augmentation method.

As shown in Figure 5, it can be observed that: 1) For both datasets, the model performance shows a trend of first improving and then decreasing with the increase of parameter λ. 2) Introducing external knowledge through our proposed knowledge augmentation method can improve the model performance for Doc-RE. However, too high the weight value λ would lead to a decline for the model performance. One pos-

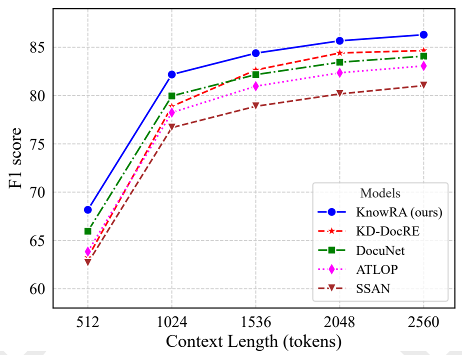

Figure 4: The trend of model performance changing with context length in the DWIE dataset

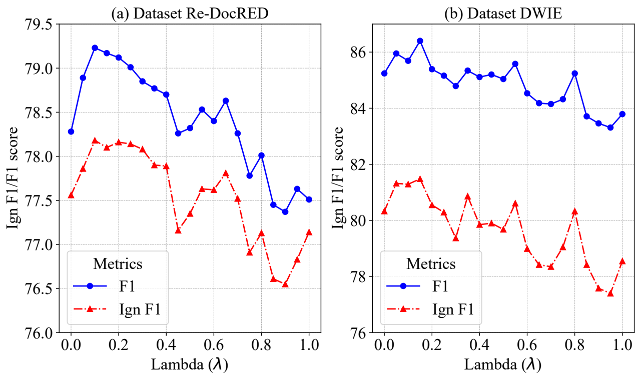

Figure 5: The trend of our KnowRA model performance changing with the weight λ in datasets Re-DocRED and DWIE.

sible explanation is that external knowledge may be wrong or inconsistent with the internal information of the document.

3.8 Case Study

Figure 6 shows the construction of K-MHDG by introducing the common-sense knowledge from the external knowledge base. The nodes in the figure represent entities, where the numbers indicate the identification index of the entities. The directed edges between nodes represent the relations between entities, the words on the edges represent the relation types, and numbers on edges represent the confidence score, i.e., τ.

Our knowledge augmentation method can extend the relation triples that were not annotated in the original dataset, identify the correct relation triples, denoted by the green ✓, and also filter out noise external knowledge and incorrect relation triples, which are denoted by the red ×.

4 Related Work

For pattern recognition and logical reasoning, existing models mainly used intermediary entities or evidence sentences to achieve logical reasoning [Yuan *et al.*, 2023; Zhang *et al.*, 2021b; Wei and Li, 2022]. However, evidence sentences are not always available. Peng et al. [2022] proposed a subgraph seasoning model for Doc-RE, which places emphasis on key

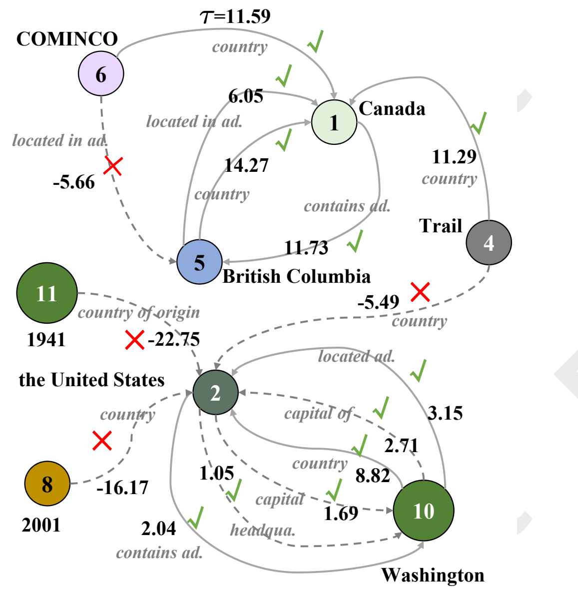

Figure 6: The illustration of K-MHDG. The solid lines denote relations constructed based on the original dataset, while dashed lines denote relations extended from external knowledge.

entities and integrates various paths between entity/mention pairs into a subgraph for relation reasoning.

For co-reference reasoning, this type of model has positive effects on improving the performance of modeling long-distance reasoning by reducing ambiguity between co-reference entities and mentions involving multiple sentences [Xue *et al.*, 2022], Ye et al. [2020] proposed mention reference prediction method to equip the language model with the capacity for capturing the co-reference relations. Wang et al. [Wang *et al.*, 2022a] distilled the co-reference reasoning ability into the relation extraction model.

However, for common-sense reasoning, using only the information in the current document makes it difficult to establish implicit associations and determine relation types between different entities/mentions in multiple sentences, without the help of common sense distilled from external knowledge [Li *et al.*, 2021]. Existing models rarely incorporate external knowledge into the Doc-RE model. To solve this problem, we introduced external knowledge into our model with the knowledge filtration method.

5 Conclusions

In this paper, we presented a comprehensive reasoning reasoning model for Doc-RE. Concretely, the proposed knowledge document graph and three different reasoning components, i.e., graph-based semantic encoding, co-reference resolution, knowledge augmentation and filtration, and axial attention method were integrated into our KnowRA model to enhance the comprehensive reasoning abilities. Experiments conducted on two benchmarks datasets demonstrated the superiority of the proposed model in Doc-RE.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61572250 and 62476135), Jiangsu Province Science & Tech Research Program (BE2021729), Open project of State Key Laboratory for Novel Software Technology, Nanjing University (KFKT2024B53), Jiangsu Province Frontier Technology Research and Development Program (BF2024005), Nanjing Science and Technology Research Project (202304016) and Collaborative Innovation Center of Novel Software Technology and Industrialization, Jiangsu, China.

References

[Beltagy *et al.*, 2020] Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. In *arXiv: 2004.05150*, 2020.

[Dai *et al.*, 2022] Qin Dai, Benjamin Heinzerling, and Kentaro Inui. Cross-stitching text and knowledge graph encoders for distantly supervised relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6947–6958, 2022.

[Ding *et al.*, 2022] Yang Ding, Jing Yu, Bang Liu, Yue Hu, Mingxin Cui, and Qi Wu. Mukea: Multimodal knowledge extraction and accumulation for knowledge-based visual question answering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5079–5088, 2022.

[Fan *et al.*, 2022] Shengda Fan, Shasha Mo, and Jianwei Niu. Boosting document-level relation extraction by mining and injecting logical rules. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10311–10323, 2022.

[Guo *et al.*, 2023] Jia Guo, Stanley Kok, and Lidong Bing. Towards integration of discriminability and robustness for document-level relation extraction. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2598–2609, 2023.

[Huang *et al.*, 2023] Quzhe Huang, Yutong Hu, Shengqi Zhu, Yansong Feng, Chang Liu, and Dongyan Zhao. More than classification: A unified framework for event temporal relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 9631–9646, 2023.

[Huang *et al.*, 2024] Heyan Huang, Changsen Yuan, Qian Liu, and Yixin Cao. Document-level relation extraction via separate relation representation and logical reasoning. In *ACM Transactions on Information Systems*, pages 22:1–22:24, 2024.

[Li *et al.*, 2021] Bo Li, Wei Ye, Canming Huang, and Shikun Zhang. Multi-view inference for relation extraction with uncertain knowledge. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 13234–13242, 2021.

[Man *et al.*, 2022] Hieu Man, Nghia Trung Ngo, Linh Ngo Van, and Thien Huu Nguyen. Selecting optimal context sentences for event-event relation extraction. In *Thirty-Sixth AAAI Conference on Artificial Intelligence*, pages 11058–11066, 2022.

[Orogat and El-Roby, 2022] Abdelghny Orogat and Ahmed El-Roby. Smartbench: Demonstrating automatic generation of comprehensive benchmarks for question answering over knowledge graphs. In *Proc. VLDB Endow.*, pages 3662–3665, 2022.

[Peng *et al.*, 2022] Xingyu Peng, Chong Zhang, and Ke Xu. Document-level relation extraction via subgraph reasoning. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 4331–4337, 2022.

[Qi *et al.*, 2024] Kunxun Qi, Jianfeng Du, and Hai Wan. End-to-end learning of logical rules for enhancing document-level relation extraction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistic*, pages 7247–7263, 2024.

[Sun *et al.*, 2023] Qi Sun, Kun Huang, Xiaocui Yang, Pengfei Hong, Kun Zhang, and Soujanya Poria. Uncertainty guided label denoising for document-level distant relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 15960–15973, 2023.

[Tan *et al.*, 2022a] Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. Document-level relation extraction with adaptive focal loss and knowledge distillation. In *Findings of the Association for Computational Linguistics*, pages 1672–1681, 2022.

[Tan *et al.*, 2022b] Qingyu Tan, Lu Xu, Lidong Bing, Hwee Tou Ng, and Sharifah Mahani Aljunied. Revisiting docred - addressing the false negative problem in relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in NaturalLanguage Processing*, pages 8472–8487, 2022.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 5998–6008, 2017.

[Wang *et al.*, 2022a] Xinyi Wang, Zitao Wang, Weijian Sun, and Wei Hu. Enhancing document-level relation extraction by entity knowledge injection. In *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference*, pages 39–56, 2022.

[Wang *et al.*, 2022b] Ye Wang, Xinxin Liu, Wenxin Hu, and Tao Zhang. A unified positive-unlabeled learning framework for document-level relation extraction with different levels of labeling. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4123–4135, 2022.

[Wei and Li, 2022] Ying Wei and Qi Li. Sagdre: Sequence-aware graph-based document-level relation extraction with adaptive margin loss. In *The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2000–2008, 2022.

[Xu *et al.*, 2021a] Benfeng Xu, Quan Wang, Yajuan Lyu, Yong Zhu, and Zhendong Mao. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 14149–14157, 2021.

[Xu *et al.*, 2021b] Wang Xu, Kehai Chen, and Tiejun Zhao. Discriminative reasoning for document-level relation extraction. In *Findings of the Association for Computational Linguistics*, pages 1653–1663, 2021.

[Xu *et al.*, 2021c] Wang Xu, Kehai Chen, and Tiejun Zhao. Document-level relation extraction with reconstruction. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 14167–14175, 2021.

[Xu *et al.*, 2022] Wang Xu, Kehai Chen, Lili Mou, and Tiejun Zhao. Document-level relation extraction with sentences importance estimation and focusing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2920–2929, 2022.

[Xu *et al.*, 2024] Zhentao Xu, Mark Jerome Cruz, Matthew Guevara, Tie Wang, Manasi Deshpande, Xiaofeng Wang, and Zheng Li. Retrieval-augmented generation with knowledge graphs for customer service question answering. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2905–2909, 2024.

[Xue *et al.*, 2022] Zhongxuan Xue, Rongzhen Li, Qizhu Dai, and Zhong Jiang. Corefdre: Document-level relation extraction with coreference resolution. In *arXiv: 2202.10744*, 2022.

[Yang *et al.*, 2023] Zhe Yang, Yi Huang, and Junlan Feng. Learning to leverage high-order medical knowledge graph for joint entity and relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9023–9035, 2023.

[Yao *et al.*, 2019] Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. Docred: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 764–777, 2019.

[Ye *et al.*, 2020] Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Peng Li, Maosong Sun, and Zhiyuan Liu. Coreferential reasoning learning for language representation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 7170–7186, 2020.

[Yuan *et al.*, 2023] Changsen Yuan, Heyan Huang, Yixin Cao, and Yonggang Wen. Discriminative reasoning with sparse event representation for document-level event-event relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 16222–16234, 2023.

[Zaporojets *et al.*, 2021] Klim Zaporojets, Johannes Deleu, Chris Develder, and Thomas Demeester. Dwie: an entity-centric dataset for multi-task document-level information extraction. In *Inf. Process. Manag.*, page 102563, 2021.

[Zeng *et al.*, 2020] Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. Double graph based reasoning for document-level relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 1630–1640, 2020.

[Zeng *et al.*, 2021] Shuang Zeng, Yuting Wu, and Baobao Chang. Sire: Separate intra- and inter-sentential reasoning for document-level relation extraction. In *Findings of the Association for Computational Linguistics*, pages 524–534, 2021.

[Zhang *et al.*, 2021a] Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. Document-level relation extraction as semantic segmentation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3999–4006, 2021.

[Zhang *et al.*, 2021b] Zhenyu Zhang, Bowen Yu, Xiaobo Shu, and Tingwen Liu. Na-aware machine reading comprehension for document-level relation extraction. In *Research Track - European Conference ECML/ PKDD*, pages 580–595, 2021.

[Zhang *et al.*, 2023a] Liang Zhang, Jinsong Su, Zijun Min, Zhongjian Miao, Qingguo Hu, Biao Fu, Xiaodong Shi, and Yidong Chen. Exploring self-distillation based relational reasoning training for document-level relation extraction. In *Thirty-Seventh AAAI Conference on Artificial Intelligence*, pages 13967–13975, 2023.

[Zhang *et al.*, 2023b] Ruoyu Zhang, Yanzeng Li, and Lei Zou. A novel table-to-graph generation approach for document-level joint entity and relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 10853–10865, 2023.

[Zhou *et al.*, 2021] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. Document-level relation extraction with adaptive thresholding and localized context pooling. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 14612–14620, 2021.