LLM with Relation Classifier for Document-Level Relation Extraction

Anonymous ACL submission

Abstract

Large language models (LLMs) create a new paradigm for natural language processing. Despite their advancement, LLM-based methods still lag behind traditional approaches in document-level relation extraction (DocRE), a critical task for understanding complex entity relations. This paper investigates the causes of this performance gap, identifying the dispersion of attention by LLMs due to entity pairs without relations as a primary factor. To address this challenge, we introduce a novel classifier-LLM approach to DocRE. The proposed method (LMRC) begins with a classifier specifically designed to filter and select entity pair candidates exhibiting potential relations, and thereby feeds them to LLM for the extraction of final relations. This method ensures that during inference, the LLM's focus is directed primarily at entity pairs with relations. Experiments on DocRE benchmark reveal that our method significantly outperforms recent LLM-based DocRE models and achieves competitive performance with several leading traditional DocRE models.

1 Introduction

Document-level Relation Extraction (DocRE) aimes at extracting semantic relations between entity pairs from documents. Prior DocRE models emulate the process of reading and reasoning on entity pairs throughout the entire document using advanced neural network architectures, including self-attention mechanisms (Xu et al., 2021a; Tan et al., 2022a), and GNNs (Li et al., 2020; Christopoulou et al., 2019). Recently, (Sun et al., 2024) utilizes chain-of-retrieval prompt and denoising strategy to enhance the performance of LLMs in DocRE.

However, LLM-based methods still lag behind traditional approaches in document-level relation extraction. We observe that following the definition of the DocRE task, all possible entity pairs (which we refer to as candidate space) are constructed and fed into models, and within this extensive array of entity pairs, only a select few actually harbor relations. We define this phenomenon as the imbalance in the candidate space, which may lead to LLMs focusing more on NA entity pairs that do not express any relation. Consequently, this might be the root cause of the performance deficiencies observed in LLMs on DocRE.

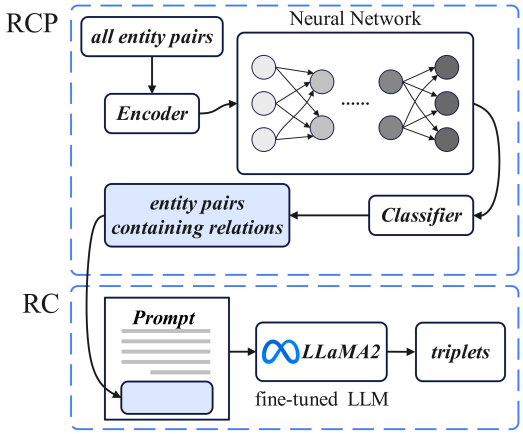


Figure 1: **Illustration of LMRC**. Relation Candidate Proposal(RCP) leverages localized context pooling (Zhou et al., 2021) in the construction of a preprocessing classifier, focusing on selecting entity pairs that contain relations. Relation Classification(RC) takes the results from the previous stage to create prompt that guides fine-tuned LLaMA2 to accomplish multi-classification task.

The statistical analysis of the DocRED (Yao et al., 2019) dataset, along with our results from fine-tuning LLaMA2-13B-Chat (Touvron et al., 2023) for DocRE, further corroborate our hypothesis that the integration of NA entity pairs expands the candidate space, thereby leading to confusion of LLMs. In response to this problem, we introduce a novel method LMRC (shown in Figure 1), which conceptualizes DocRE as a workflow comprising two key stages: **Relation Candidate Proposal**(RCP) and **Relation Classification**(RC). The

former constructs a pre-processing classifier that explicitly leverages attention mechanism to filter out NA entity pairs. The latter uses LLaMA2-13B-Chat to accomplish multi-classification on the reduced candidate space.

Our evaluation of LMRC on the DocRED dataset shows that adopting a pre-processing classifier significantly enhances performance over our baseline, suggesting its viability as a strategy for future DocRE.

2 Preliminary Experiment

In this section, we fine-tune LLaMA2-13B-Chat for document-level relation extraction and report the results realized by this approach as well as the statistical findings on the DocRED dataset.

2.1 Fine-tuning LLaMA2

To construct prompts for this task, we use the instruction: *Your task is to determine whether there are relations between the entity pairs based on the information in the text. If there exists relations, select relations for the entity pairs from the relation set; if there is no relation, return None.*, followed by an input consisting of a predefined relation set, the text corresponding to the document, and several entity pairs. To prevent semantic ambiguity and reduce token usage, we employ *None* as a representation for NA, requiring the model to label entity pairs devoid of relations with *None*. The complete prompt format is provided in Appendix.

DocRED has many tokens, often exceeding max token length, unlike sentence-level tasks. For this, we conduct relation extraction for each document D via $\frac{n \times (n-1)}{k}$ inputs, where n denotes the number of entities in document D, and the variable k represents the max entity pair limit per input. We construct inputs based on DocRED to perform LoRA (Hu et al., 2022) fine-tuning and testing[1] on LLaMA2-13B-Chat.

2.2 Results

Statistics of DocRED are shown in Table 1, Entity pairs containing relations occupy only 2.7% of the candidate space on DocRED's dev set. The predominance of NA entity pairs results in a imbalance in the candidate space. Figure 2 illustrates that the performance of LLaMA2-13B-Chat can be improved during fine-tuning. However, as shown in Table 2, LLaMa2-13B-Chat only generates 4925 triplets on the dev set, significantly less than the 12275 annotated in DocRED. This indicates that LLaMA2-13B-Chat tends to label entity pairs as NA, resulting in a lower recall, there by decreasing the F_1. The foregoing outcomes demonstrate that the imbalanced candidate space has adversely affected the performance of LLM on DocRE.

[1]Entities in the triplets returned by LLaMA2-13B-Chat are aligned to the dataset using thefuzz, and the relations generated not in the predefined relation set are considered incorrect.

Description	**Dev**	**Test**
Candidate Space	395,572	392,158
NA Entity Pairs	384,949	-
Relation Entity Pairs	10,623	-
Annotated Triples	12,275	-

Table 1: Statistics on DocRED

Metrics	**Dev**	**Test**
Precision	69.00	-
Recall	27.43	-
F_1	39.25	38.66
Ign F_1	38.62	38.09
Extracted Triples	4,925	4,932

Table 2: Number of valid triplets in golden label and those generated by different methods.

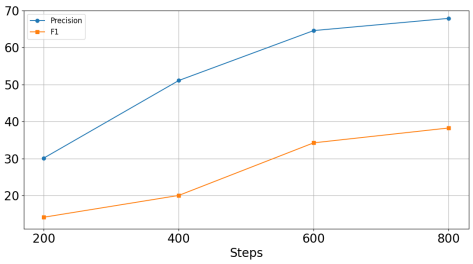


Figure 2: Learning curve of LLaMA2-13B-Chat regarding to training steps

3 LMRC

To prevent LLaMA2-13B-Chat from prioritizing NA entity pairs, LMRC initially uses traditional neural networks for **Relation Candidate Proposal** to identify entity pairs with relations. Then, LLaMA2-13B-Chat uses these proposals for **Relation Classification**.

3.1 Relation Candidate Proposal

In this stage, we build a simple model to conduct a binary classification task, with the outcome being entity pairs containing relations. As prior works (Tan et al., 2022b; Ma et al., 2023) have shown that contextual information is indispensable for the

relation extraction task, our model adapts localized context pooling from Zhou et al. (2021).

Entity Representation Following the entity marker technique (Zhang et al., 2017; Shi and Lin, 2019), a special token "*" is inserted at the start and end position of each entity mention. Then, tokens $T = \{t_i\}_{i=1}^{l}$ within document D are encoded by a Transformer-based(Vaswani et al., 2017) pretrained language model (PLM) to generate contextualized embeddings $\mathbf{H}$ along with their attentions $\mathbf{A}$:

$$\mathbf{H}, \mathbf{A} = PLM(T) \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{l \times d}$, $\mathbf{A} \in \mathbb{R}^{H \times l \times l}$, d is the hidden dimension of the PLM and H is the number of attention heads. We take the embedding of "*" at the start of mentions as their embeddings. The entity embedding $h_{e_i} \in \mathbb{R}^d$ for each entity e_i with mentions $M_{e_i} = \{m_j^i\}_{j=1}^{N_{e_i}}$ is computed by logsumexp pooling (Jia et al., 2019):

$$h_{e_i} = log \sum_{j=1}^{N_{e_i}} exp(h_{m_j^i}) \tag{2}$$

Localized Context Representation For each entity e_i, we aggregate the attention output for its mentions by mean pooling $A_{e_i} = \sum_{j=1}^{N_{e_i}} (a_{m_j^i})$, where $a_{m_j^i} \in \mathbb{R}^{H \times l}$ is the attention weight at the position of mention m_j^i from the last layer. Then given an entity pair (e_s, e_o), its localized context embedding $c^{(s,o)} \in \mathbb{R}^d$ can be obtained by:

$$q^{(s,o)} = \sum_{i=1}^{H} (A_{e_s}^i \circ A_{e_o}^i) \tag{3}$$

$$c^{(s,o)} = \mathbf{H}^\top q^{(s,o)} \tag{4}$$

where $q^{(s,o)} \in \mathbb{R}^l$ is the mean-pooled attention weight for entity pair (e_s, e_o) and $\mathbf{H}$ is the contextualized embedding in Eq.(1).

Binary Classification To predict whether entity pair (e_s, e_o) expresses relation, we first generate context-enhanced entity representations:

$$z_s^{(s,o)} = tanh(\mathbf{W}_s h_{e_s} + \mathbf{W}_c c^{(s,o)}) \tag{5}$$

where $\mathbf{W}_s, \mathbf{W}_c \in \mathbb{R}^{d \times d}$ are trainable parameters. We obtain the object representation $z_o^{(s,o)}$ in the same manner. Then, a bilinear classifier is applied on the representations to compute the probability:

$$P(NA|e_s, e_o) = \sigma(z_s^{(s,o)\top} \mathbf{W} z_o^{(s,o)} + b) \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{d \times d}$ is a trainable parameter matrix, σ is the sigmoid function, $P(NA|e_s, e_o)$ is the probability that entity pair (e_s, e_o) does not express any relation. We choose Binary Cross Entropy as our loss function.

3.2 Relation Classification

After identifying entity pairs with relations in the RCP stage, we apply the method from Section 2.1, driving LLaMA2-13B-Chat to complete relation classification with supervised fine-tuning. We slightly modify the previous prompt by removing the *None* category and changing some expressions. These changes aim to sharpen the model's focus on the classification task. Additionally, the number of inputs for each document is greatly reduced ($\frac{n \times (n-1)}{k} \rightarrow 1$) due to the elimination of *None*. Detailed changes can be found in Appendix A.

4 Experiments

4.1 Settings

Dataset We conduct experiment on DocRED (Yao et al., 2019), a large-scale crowd-sourced dataset for document-level RE. DocRED consists of 3053 documents for training, 998 for development, and 1000 for test. Over 40.7% of DocRED's relational facts require multi-sentence extraction.

Configuration In the RCP stage , we select RoBERTa$_{\text{large}}$ (Liu et al., 2019) as the PLM encoder. We use AdamW (Loshchilov and Hutter, 2019) as the optimizer, with learning rates $\in \{1e-5, 3e-5, 5e-5\}$. We implement early stopping based on the F_1 obtained from the dev set. In the RC stage, we LoRA fine-tune LLaMA2-13B-Chat with the RC-specific prompt.

Evaluation We employ the standard evaluation metrics of DocRED (Yao et al., 2019): F_1 and Ign F_1. Ign F_1 is measured by removing triplets present in the training set from the dev and test set.

4.2 Performance of LMRC

We compare our LMRC with sequence-based, BERT-based and LLM-based methods on DocRED. The experimental results are shown in Table 3. Sequence-based methods use neural networks to encode the entire document, while BERT-based methods adapt BERT family pretrained models as encoders. LLM-based methods, recently proposed, employ in-context learning or retrieval augmented generation (RAG, Lewis et al. (2020)) to improve the performance of LLMs on DocRE.

	Method	Dev		Test	
		Ign F_1	F_1	Ign F_1	F_1
Sequence-based	CNN(Yao et al., 2019)	38.0	43.5	36.4	42.3
	BiLSTM(Yao et al., 2019)	45.1	51.0	44.7	51.1
BERT-based	HIN-BERT$_{base}$(Tang et al., 2020)	54.3	56.3	53.7	55.6
	CorefBERT$_{base}$(Ye et al., 2020)	55.3	57.5	54.5	57.0
	CorefRoBERTa$_{large}$(Ye et al., 2020)	57.4	59.4	57.9	60.3
	SSAN-RoBERTa$_{large}$(Xu et al., 2021b)	60.3	62.1	59.5	61.4
LLM-based	CoR(Sun et al., 2024)	-	38.4 ± 10.6	-	38.5 ± 9.1
	GenRDK(Sun et al., 2024)	-	42.5 ± 10.6	-	41.5 ± 8.7
Our Methods	LoRA FT LLaMA2-13B-Chat†	38.6	39.3	38.1	38.7
	LMRC	58.2	60.0	58.5	60.5

Table 3: Evaluation results on the development and test set of DocRED. The scores of prior methods are borrowed from corresponding papers. Result marked with † is our baseline.

Method	Intra	Inter
CNN†	51.87	37.58
BiLSTM†	57.05	43.49
BERT-RE$^{\dagger}_{base}$	61.61	47.15
RoBERTa-RE$^{*}_{base}$	65.65	50.09
LSR-BERT$^{\dagger}_{base}$	65.26	52.05
GAIN-BERT$^{*}_{base}$	67.10	53.90
LoRA FT LLaMA2-13B-Chat	45.43	31.67
LMRC	65.88	52.66

Table 4: Intra- and Inter-F_1 on dev set of DocRED. † denotes results from Nan et al. (2020), and * denotes results from Zeng et al. (2020).

Metrics	RC Stage
Precision	88.65
Recall	84.94
F_1	86.75
Ign F_1	86.09

Table 5: Result of the RC stage with entity pairs from golden labels of dev set as inputs.

Fine-tuned LLaMA2-13B-Chat and similar LLM methods perform akin to early CNNs, indicating challenges in using LLMs for DocRE. After dividing the task, our LMRC significantly improves, increasing F_1 by 20.7 and 21.8 on the dev and test sets over LoRA fine-tuning. Besides, LMRC achieves an F_1 of 60.5, which exceeds other LLM-based methods and approaches the performance of one of the state-of-the-art BERT-based models, SSAN-RoBERTa$_{large}$.

We also report Intra-F_1/Inter-F_1, which only consider either intra- or inter-sentence relations respectively. LSR (Nan et al., 2020) and GAIN (Zeng et al., 2020) are both graph-based models that perform inference through graph neural network. As Table 4 shows, LMRC not only surpasses our baseline in Intra- and Inter-F_1 but also remains competitive when contrasted with graph-based models, underlining the potential of our method.

4.3 Analysis of RC Stage

To further clarify the specific effect of LLaMA2-13B-Chat in the RC stage, we isolate this part: take entity pairs annotated in the golden labels of dev set as inputs, mask their relation tags, and then task the fine-tuned LLM to classify these entity pairs into predefined relation set. The result, presented in Table 5, illustrates that the fine-tuned LLM excels in relation classification based on the context of the documents. Moreover, The subtle difference between F_1 and Ign F_1 reveals that the extensive world knowledge possessed by LLMs may help bridge the gap between them.

5 Conclusion

In this work, we investigate the underlying reasons for LLM's limited effectiveness in document-level relation extraction and introduce a new approach, the LLM with Relation Classifier (LMRC), for DocRE. Our method comprises two main stages: relation candidate proposal and relation classification. Through experiments conducted on the DocRED dataset, we demonstrate the effectiveness of our proposed LMRC approach. The results further reveal that LMRC holds strong competitive advantages over other existing fully supervised models, indicating its potential as a viable framework for future DocRE research.

6 Limitations

Despite our efforts, there are some limitations in this study:

LLMs: We only tested our method with LLaMA2-13B-Chat because of time limits. This choice limits how well our results might apply to other models. In the future, we plan to test our method with different LLaMA2 models and other freely available LLMs like Chat-GLM. This will help us understand how well our method works across different models.

Datasets: We have only used the DocRED dataset to evaluate our method. Next, we will try it on other datasets like RE-DocRED and CDR to see if our method is effective in different settings.

Other Methods: Right now, our approach needs a special pre-processing step before the main analysis. In the future, we want to use LLMs to do this pre-processing. This could make our method simpler and more direct.

These limitations point out where we need to improve. By working on these areas, we hope to make our method better and more useful for document analysis and relation extraction.

References

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4925–4936, Hong Kong, China. Association for Computational Linguistics.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Robin Jia, Cliff Wong, and Hoifung Poon. 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.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474. Curran Associates, Inc.

Bo Li, Wei Ye, Zhonghao Sheng, Rui Xie, Xiangyu Xi, and Shikun Zhang. 2020. Graph enhanced dual attention network for document-level relation extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1551–1560, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Youmi Ma, An Wang, and Naoaki Okazaki. 2023. DREEAM: Guiding attention with evidence for improving document-level relation extraction. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1971–1983, Dubrovnik, Croatia. Association for Computational Linguistics.

Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1546–1557, Online. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling. *arXiv preprint arXiv:1904.05255*.

Qi Sun, Kun Huang, Xiaocui Yang, Rong Tong, Kun Zhang, and Soujanya Poria. 2024. Consistency guided knowledge retrieval and denoising in llms for zero-shot document-level relation triplet extraction. *arXiv preprint arXiv:2401.13598*.

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

Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. 2022b. Document-level relation extraction with adaptive focal loss and knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1672–1681, Dublin, Ireland. Association for Computational Linguistics.

Hengzhu Tang, Yanan Cao, Zhenyu Zhang, Jiangxia Cao, Fang Fang, Shi Wang, and Pengfei Yin. 2020. 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*, page 197–209.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Somin Wadhwa, Silvio Amir, and Byron Wallace. 2023. Revisiting relation extraction in the era of large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15566–15589, Toronto, Canada. Association for Computational Linguistics.

Benfeng Xu, Quan Wang, Yajuan Lyu, Yong Zhu, and Zhendong Mao. 2021a. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14149–14157.

Benfeng Xu, Quan Wang, Yajuan Lyu, Yong Zhu, and Zhendong Mao. 2021b. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14149–14157.

Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 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, Florence, Italy. Association for Computational Linguistics.

Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Peng Li, Maosong Sun, and Zhiyuan Liu. 2020. Coreferential Reasoning Learning for Language Representation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7170–7186, Online. Association for Computational Linguistics.

Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 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, Online. Association for Computational Linguistics.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

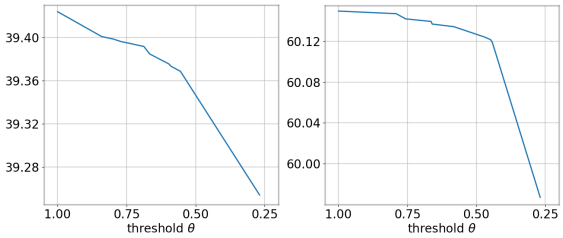
Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-level relation extraction with adaptive thresholding and localized context pooling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14612–14620.

A Prompts

Table 6 shows the prompt designed for document-level relation extraction task, and Table 7 shows the prompt for relation classification task. The primary distinction between the two lies in "Instruction" and "Entity Pairs". The former encompasses all constructible entity pairs, while the latter's entity pairs are obtained through the RCP stage.

B Out-of-Domain Relations Studying

In the aforementioned evaluation, we simply categorize all relations generated by LLaMA2-13B-Chat that do not fall within the predefined relation set as erroneous outcomes. However, previous work (Wadhwa et al., 2023) has pointed out that evaluating LLM-based models cannot entirely rely on exact matches to targets. For example, although "works at" from the result is semantically similar to "work for" in the target, strict evaluation criteria would count it as a misclassification.



(a) Our baseline (b) LMRC

Figure 3: The impact of threshold θ on F_1

In order to delve into this phenomenon more thoroughly, we revisit the out-of-domain relations that emerged during generation. We leverage SBERT[2] (Reimers and Gurevych, 2019) to align out-of-domain relations into the predefined relation set R. This process involves the computation of cosine similarity. For each out-of-domain relation r_i,

[2]https://www.sbert.net

we choose the relation in R with the highest similarity s^i_{max} as the final result. Intuitively, some of the out-of-domain relations may be meaningless, and aligning all of them to R is not appropriate. Therefore, we introduce a heuristic threshold θ, where alignment is performed when $s^i_{\text{max}} \geq \theta$; otherwise, the triplets containing r_i are discarded. After alignment, we recalculate the F_1 of our methods on the dev set.

Figure 2 shows that, in both methods, F_1 attains its maximum value when threshold θ is set to 1. It means that even with semantic alignment, all out-of-domain relations generated by the fine-tuned LLaMA2-13B-Chat remain incorrect, contradicting our idea.

Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

Instruction:
Your task is to determine whether there are relations between the entity pairs based on the information in the text. If there exists relations, select relations for the entity pairs from the relation set; if there is no relation, return None.
The format of the input entity pair is '(head entity| -| tail entity)'.
Your output format is '(head entity| relation/None| tail entity)'.

Relation set:
{predefined relation set}

Text:
{text}

{number of entity pairs} Entity pairs:
{entity pairs}

Response:

Table 6: Prompt for document-level relation extraction

Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

Instruction:
This is a relation classification task. we will provide entity pairs that require relation classification. Your task is to select relations for each entity pair from the given relation set based on the information in the text. There may be multiple relations between an entity pair.
The format of the input entity pair is '(head entity| -| tail entity)'.
Your output format is '(head entity| relation| tail entity)'.

Relation set:
{predefined relation set}

Text:
{text}

{number of entity pairs} Entity pairs:
{entity pairs}

Response:

Table 7: Prompt for relation classification