Knowledge Base-enhanced Multilingual Relation Extraction with Large Language Models

Tong Chen[1,2], Procheta Sen[2], Zimu Wang[2,3], Zhengyong Jiang[1,*] and Jionglong Su[1,*]

[1] School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China
[2] Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, UK
[3] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China

Abstract

Relation Extraction (RE) is an essential task that involves comprehending relational facts between entities from natural language texts. However, existing research in RE, particularly those based on large language models (LLMs), is proven to fall short in the task due to their *context unawareness* (lack of fine-grained understanding), *schema misalignment* (misaligned with human-defined schema), and *world knowledge ignorance* (relying solely on internal parametric knowledge). In this paper, we propose a novel framework to address the aforementioned challenges. The framework consists of two stages, including 1) entity linking and 2) relation inference, by fully leveraging the efficacy of external knowledge bases (KBs) and LLMs in this task. We conduct extensive experiments in a multilingual setting and achieve state-of-the-art performance on the experimented datasets. The LLMs with external knowledge can typically outperform those without knowledge by a significant margin, indicating the effectiveness of our proposed framework.

Keywords

Multilingual, Relation Extraction, Knowledge Bases, Large Language Models, Natural Language Processing

1. Introduction

Relation Extraction (RE) is an essential task in information extraction (IE) that aims to comprehend relational facts between entities in natural language texts [1]. For the first example in Table 1, given an original input and an entity pair of interest (*Apple Inc.*, *iPhone*), an RE model should be able to predict the relationship between them, i.e., *Apple Inc.* $\xrightarrow{\textit{product produced}}$ *iPhone*. The structured knowledge obtained from RE models can support a variety of downstream applications, such as knowledge graph construction or completion [1], question answering [2], and dialogue systems [3].

Previous research usually formulates RE as a pairwise classification task with pre-trained language models (PLMs), in which novel methods have been proposed [4, 5]. Recently, large language models (LLMs) demonstrate promising performance in a variety of downstream tasks [6, 7] across several paradigms, such as in-context learning (ICL) [8], chain-of-thought (CoT) prompting [9], and fine-tuning. However, they fall short in multiple specification-heavy tasks, including RE, whose performance under particularly ICL is much behind state-of-the-art PLM-based methods [10]. Table 1 gives some examples of mispredicted entity relationships using LLMs. Overall, the reasons why LLMs cannot perform well in RE include their *context unawareness*, *schema misalignment*, and *world knowledge ignorance*:

1. ***Context Unawareness.*** The completion of RE requires a thorough and fine-grained comprehension of the information in given contexts. However, LLMs with ICL usually lack fine-grained context awareness, which results in disregarded or erroneous relation prediction [10]. In the first example in Table 1, LLMs should first thoroughly appreciate the context and the connection between "*Apple Inc.*", "*device*", and "*iPhone*"; otherwise, they are unable to determine the relationship between "*Apple Inc.*" and "*iPhone*".

The First international OpenKG Workshop: Large Knowledge-Enhanced Models, August 03, 2024, Jeju Island, South Korea

[*] Corresponding author.

Zhengyong.Jiang02@xjtlu.edu.cn (Z. Jiang); Jionglong.Su@xjtlu.edu.cn (J. Su)
0000-0001-8873-4073 (Z. Jiang); 0000-0001-5360-6493 (J. Su)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Table 1

Examples of mispredicted relationships by large language models (LLMs), consisting of three categories: *context unawareness*, *schema misalignment*, and *world knowledge ignorance*.

Error Type	Example
Context Unawareness	**Input:** Apple Inc. is an American multinational corporation [...] Devices include the iPhone, iPad, Mac, [...]. **Entities:** Apple Inc., iPhone **Prediction:** N/A **Label:** product produced
Schema Misalignment	**Input:** Armstrong joined the NASA Astronaut Corps in the second group, which was selected in 1962. **Entities:** Armstrong, NASA Astronaut Corps **Prediction:** work for **Label:** part of
Knowledge Ignorance	**Input:** The theory of relativity usually encompasses two interrelated physics theories by Albert Einstein. **Entities:** theory of relativity, Albert Einstein **Prediction:** inventor **Label:** discoverer

2. ***Schema Misalignment.*** RE models are required to predict the relationships between entities from a human-labeled, pre-defined schema. However, the number of candidate relationships is typically lengthy, and some relation types are misaligned between LLMs and human expectations [10, 11]. In the second example in Table 1, LLMs may confuse the two relation types, "*work for*" and "*part of*", and make incorrect predictions on the relationship between "*Armstrong*" and "*NASA Astronaut Corps*".
3. ***World Knowledge Ignorance.*** World knowledge usually plays a vital role in RE, particularly in understanding implicit relationships [12] and domain-specific knowledge [13]. However, LLMs suffer in tasks that require rich world knowledge [14] and solely rely on their internal parametric knowledge [10]. In the third example in Table 1, LLMs may predict the relationship as "*inventor*" rather than "*discoverer*" without thoroughly understanding the knowledge of "*Albert Einstein*" and the "*theory of relativity*".

Knowledge bases (KBs) have been extensively employed in previous RE research. For example, researchers leverage the relationships obtained from Freebase [15] and Wikipedia infoboxes [16] to classify the relationships between entities in texts. However, such relationships are typically noisy and are not faithful to what is described in the given contexts [17]. The following research focuses on denoising and learning context-dependent relationships, such as utilizing natural language inference (NLI) with entailment prediction [18]. Nevertheless, as LLMs have demonstrated their abilities in NLI [19] and natural language reasoning [20], the capability of the combination of KBs and LLMs requires further exploration to design contextual, aligned, and knowledgeable RE models. Moreover, previous research on knowledge-enhanced RE primarily focuses on the English corpus, which limits the adaptability of RE models to different linguistic contexts. This shortage hinders the development of comprehensive IE systems in the multilingual setting.

In this paper, we propose a novel framework for RE to address the aforementioned challenges by making the process *contextually aware*, *schema-aligned*, *world knowledge-considered*. The framework consists of two stages, entity linking and relation inference, that fully leverage the efficacy of KBs and LLMs in this task. As shown in Figure 1, given an original document and two entities of interest, we first link the entities to Wikidata [21], a large-scale multilingual KB, to ascertain the relationship between the entities in the world knowledge and regard it as the candidate relationship in the document. Subsequently, in the second stage, we use the ICL strategy on LLMs to determine whether the candidate relationship actually takes place in the given context.

We conduct extensive experiments in a multilingual setting using three widely used RE datasets: DocRED [22], REBEL [18], and REDFM [23], with three LLMs: GPT-3.5, Llama 2 [24], and Flan-T5-XL [25]. Experimental results demonstrate the effectiveness of our framework on all datasets, where the performance of zero-shot RE on the models outperforms the cases without knowledge by a significant margin. Additionally, it also achieves state-of-the-art performance on all three datasets and outperforms fine-tuned PLM-based methods, validating the efficacy of our proposed framework. We also conduct additional analysis on the effectiveness of knowledge, the impact of scaling up model parameters, and the coverage of knowledge in multilingualism to further demonstrate the effectiveness and generalizability of our proposed method.

The key contributions of this work are summarized as follows:

- We review the key literature on LLM-based RE thoroughly, and we argue that well-behaved RE models should be *contextually aware*, *schema-aligned*, and *world knowledge-considered.*
- We propose a novel framework for RE, consisting of two stages: entity linking and relation inference, to fully leverage the efficacy of KBs and LLMs in the RE task.
- Experimental results under a multilingual setting demonstrate the effectiveness and generalizability of our method across diverse linguistic contexts with substantial improvements over state-of-the-art baselines.

2. Related Work

2.1. Relation Extraction

RE has been extensively studied over the past years due to its potency in various downstream applications. Early research in RE focuses on sentence-level RE [26, 27], while some later approaches shift to the document level, aiming to comprehend the relationships between entities across multiple sentences [22]. The most commonly used methods for RE are sequence-based techniques, which essentially rely on LSTM- or Transformer-based architectures [4, 5], modeling complicated interactions between entities while implicitly capturing long-distance relationships. Furthermore, graph neural networks (GNNs) are also employed in RE due to their efficacy in representing and interacting with structured data. In this process, researchers construct relevant graphs using words, mentions, entities, or sentences as nodes and predict relationships by reasoning on the graph [28, 29].

2.2. Knowledge-enhanced RE

Knowledge-enhanced RE incorporates external knowledge information to comprehensively understand the relations between entities. Some existing work utilizes external knowledge bases like Freebase and Wikidata to improve the representation by using entity and relation information. Liu et al. [30] injects triples from knowledge graphs into texts, transforming sentences into knowledge-enhanced sentence trees. Chen et al. [31] proposes a knowledge-aware prompt-tuning approach with synergistic optimization that incorporates knowledge from relation labels into RE. External knowledge can bridge the gap between general domain data and domain-specific data, while general domain RE methods are applied in specific domains. Roy and Pan [32] uses an entity-level knowledge graph in pre-trained BERT for clinical RE, integrating medical information.

2.3. LLM-based RE

LLM-based RE has also been studied by researchers motivated by the generalized intelligence of LLMs in various downstream tasks, such as information extraction [33], machine translation [7], and adversarial attacks [6]. However, previous research concludes that LLMs typically fall short in the RE task, whose performance is much behind PLM-based approaches [10, 34, 35]. To overcome this, Zhang et al. [36] proposes QA4RE, a framework to improve the performance of LLM by aligning RE with question

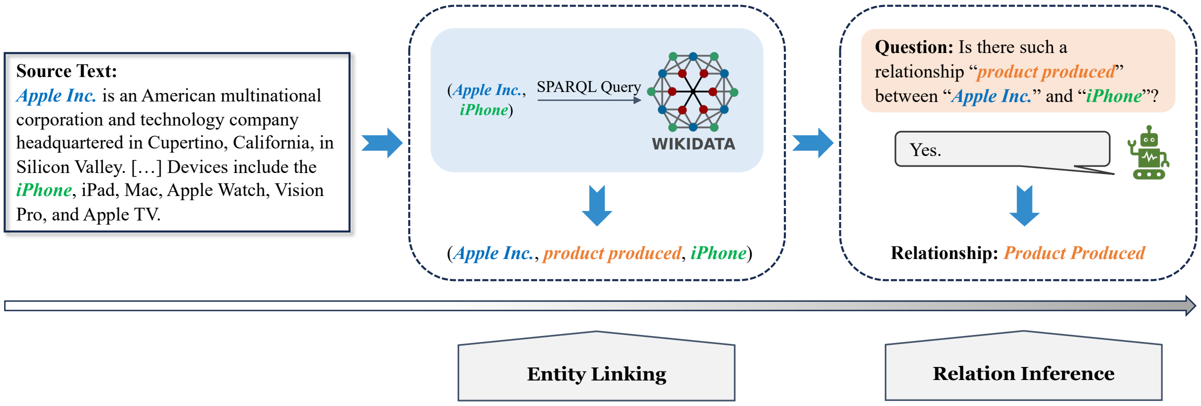
Source Text:
Apple Inc. is an American multinational corporation and technology company headquartered in Cupertino, California, in Silicon Valley. […] Devices include the iPhone, iPad, Mac, Apple Watch, Vision Pro, and Apple TV.
(Apple Inc., iPhone)
SPARQL Query
WIKIDATA
(Apple Inc., product produced, iPhone)
Question: Is there such a relationship "product produced" between "Apple Inc." and "iPhone"?
Yes.
Relationship: Product Produced
Entity Linking
Relation Inference

Figure 1: Overall framework of the proposed method, consisting of two stages: (1) entity linking and (2) relation inference using large language models (LLMs).

answering (QA) tasks. Wan et al. [37] proposes GPT-RE that utilizes task-aware representations and reasoning logic to improve entity-relationship relevance and the capability of explaining input-label mapping. Li et al. [38] suggests integrating LLM with an NLI module to construct relation triples in response to the abundance of pre-defined relation types and the uncontrollability of LLMs.

3. Methodology

3.1. Problem Formulation

We define our RE task as follows: Given a document D consisting of N sentences $\{s_1, s_2, ..., s_N\}$ (N is the number of sentences within the document, and $N = 1$ indicates sentence-level RE) and an entity pair of interest (e_h, e_t), in which e_h represents the head entity and e_t refers to the tail entity, the RE model aims to determine the potential relationship between e_h and e_t from a pre-defined schema. In our task, a KB $\mathcal{K}$ is leveraged with world knowledge, and an LLM is utilized to identify the existence of the relationship $r_{\mathcal{K}}$ retrieved form $\mathcal{K}$ between e_h and e_t in the given document.

3.2. Entity Linking and Querying

In the first stage of our proposed framework, we conduct entity linking and querying to obtain the candidate relationships between the entities of interest, which are regarded as supervision of world knowledge to the given entity pair. Entity linking is the process of linking recognized entity words to an entity in a KB, which is a pioneering step in extracting construction information from unstructured text [39]. In our framework, we link the labeled entity mentions to Wikidata [21], a large-scale multilingual KB. Once the entities are linked, we introduce a query based on SPARQL[1] to retrieve the relationships between the linked entities and regard it as the candidate relationship between them. For the datasets whose entities are annotated with coreference chains, we iterate the head and tail entities until a pair of entities can be linked to Wikidata.

3.3. Relation Inference using LLMs

After obtaining the candidate relationship between the entity pair of interest, in the second stage of our proposed framework, we adopt LLMs to identify whether the relationship actually occurs in the given context. Specifically, we leverage the ICL strategy [9] that conditions LLMs on a natural language instruction and formulate the task as a QA task due to the capacity of LLMs to answer natural questions. In accordance with the entity linking results in the first stage, we design separate prompts for the entity

[1] https://www.w3.org/TR/sparql12-query/

Table 2
Instruction and an example for relation inference for the entity pair with world knowledge retrieved from Wikidata.

Instruction:
Given information: `{source_text}`
Is there such a relationship `{relationship}` between `{head_entity}` and `{tail_entity}`?

Example:
Coburg Peak is the rocky peak rising to 783m in Erul Heights on **Trinity Peninsula** in **Graham Land**, Antarctica.
Head Entity: **Trinity Peninsula**
Tail Entity: **Graham Land**
Relationship: **part of**

Output:
Yes.

Answer:
(***Trinity Peninsula***, ***part of***, ***Graham Land***)

Table 3
Instruction and an example for relation inference for the entity pair without world knowledge retrieved from Wikidata.

Instruction:
Given information: `{source_text}`
Options of relations: `{relation_list}`
Which relationship between `{head_entity}` and `{tail_entity}` can be inferred from given options? (Please answer in English and only output the option)

Example:
Source Text: Utus Peak is the rocky peak rising to 1217m in **Trakiya Heights** on Trinity Peninsula in Graham Land, **Antarctica**. The peak is named after the ancient Roman town of Utus in Northern Bulgaria.
Relaiton List: head of government, country, place of death, sibling, [...]
Head Entity: **Trakiya Heights**
Tail Entity: **Antarctica**

Output:
continent

Answer:
(***Trakiya Heights***, ***continent***, ***Antarctica***)

pairs that have or have not been found potential relationships, and the prompts with separate examples are illustrated in Tables 2 and 3. For the entity pairs that have been found candidate relationships in the KB, we ask LLMs to determine whether they actually exist in the given context. Otherwise, we ask the LLMs to classify the relationships between the entities from the schemas directly.

This framework enables us to carry out a contextual, aligned, and knowledgeable RE process: it regards the knowledge in KBs as supervision, and the inference with LLMs makes the predictions with respect to the given contexts. Furthermore, since KBs are human-constructed world knowledge, their candidate knowledge also conforms to human-defined schemas.

4. Experiments and Analysis

4.1. Datasets

We conduct our experiments on the following three datasets, whose dataset statistics are organized in Table 4:

Table 4
Dataset statistics.

Dataset	#Types	Train	Val	Test
DocRED	96	$3,053$	$1,000$	$1,000$
REBEL	1146	3.13M	173K	174K
REDFM-EN	32	1.88K	449	446
REDFM-ES	32	1.87K	228	281
REDFM-DE	32	2.07K	252	285

- **DocRED** [22] is a document-level human-annotated RE dataset constructed from Wikipedia and Wikidata. Since at least 40.7% of relational facts in DocRED can only be extracted from multiple sentences, it requires models to comprehensively model the whole document to determine the relationships between entities.
- **REBEL** [18] is a distantly supervised dataset, hyperlinking with Wikidata and Wikipedia for relation extraction. It employs an NLI model to filter noise and address relations that are not entailed by the Wikipedia text through entailment prediction.
- **REDFM** [23] is constructed for multilingual RE that involves seven languages. Different from the REBEL dataset, REDFM not only applies NLI to filter noise but also conducts manual filtering to ensure the annotation quality. We select the English (EN), Spanish (ES), and German (DE) subsets to validate the performance of our framework in a multilingual setting.

Following the previous work in LLM-based RE [10, 34], we sample a subset from the validation set of DocRED and the test set of REBEL and REDFM to validate the performance of our method against baselines. We evaluate the performance of the experimented models using micro F1-score.

4.2. Baselines

We compare the performance of our proposed method on RE against the following baselines:

- **KD-DocRE** [4] is a semi-supervised framework for document-level RE that incorporates axial attention, adaptive focal loss, and knowledge distillation to capture the interdependency among entity-pairs. It addresses the class imbalance problem and the differences between human annotated and distantly supervised data in document-level RE.
- **DREEAM** [5] is a memory-efficient approach for improving document-level RE by incorporating evidence and offering a self-training strategy, addressing high memory consumption and limited annotated data availability in document-level RE.

4.3. Experimental Setup

We conduct our experiments on three commonly used multilingual LLMs: GPT-3.5, Llama 2 [24], and Flan-T5 [25], and we access the models with different approaches and settings. For GPT-3.5, we call the API by OpenAI[2] and select the `gpt-3.5-turbo-instruct` checkpoint due to its ability to interpret and execute human instructions seamlessly. For Llama 2 (`Llama-2-7b-chat-hf`) and Flan-T5 (`flan-t5-xl`), the models are retrieved from the HuggingFace repository[3]. To mimic the randomness of human reasoning and produce relatively stable outputs, we set the temperature of GPT-3.5 and Llama 2 as 0.2. All experiments are conducted on a single NVIDIA GeForce RTX 4090 graphics card.

[2] https://platform.openai.com/
[3] https://huggingface.co/

Table 5
Experimental results on F1-score of our proposed method under different large language models (LLMs) with and without external knowledge against baselines, in which the best and the second-best results are highlighted in **bold** and underlined, respectively.

Model	DocRED	REBEL	REDFM-EN	REDFM-ES	REDFM-DE
KD-DocRE	68.79	–	–	–	–
DREEAM	69.55	–	–	–	–
GPT-3.5	22.45	23.65	19.22	9.88	10.03
w/ Knowledge	62.52	56.68	68.17	60.27	69.39
LLaMA 2	0.00	0.72	0.00	0.00	0.00
w/ Knowledge	27.24	54.51	52.83	33.33	51.53
Flan-T5	62.79	60.65	70.76	61.05	67.86
w/ Knowledge	**73.90**	**70.40**	**79.32**	**73.84**	**81.97**

4.4. Main Results

The experimental results of our proposed framework under different LLMs with and without external knowledge are given in Table 5. From the table, we make the following observations:

First, without the incorporation of external knowledge, LLMs have been shown to fall short in the RE task and their performances are much behind those of the state-of-the-art baseline models. The results are also consistent with the previous work [10], indicating the correctness of our implementation. Among the three LLMs, Flan-T5 achieves the best performance and is remarkably close to the deliberated baseline models, indicating its excellent document-level understanding and relation reasoning ability. Llama 2 achieves the worst performance, with its results close to zero. We sample 50 outputs of Llama 2 and compare them with the ground truths. We conclude that this phenomenon is attributed to the excessively uncontrollable and flexible nature of its output compared to the rest of the models.

Second, after incorporating the external knowledge into the models, LLMs exhibit remarkable performance across all datasets, in which the average improvements of GPT-3.5, Llama 2, and Flan-T5 are 52.90, 45.90, and 11.82, respectively. Notably, the performance of Flan-T5 under the zero-shot setting achieves state-of-the-art results on all datasets, which is also better than the deliberated, fine-tuned PLM-based methods. GPT-3.5 improves the most among the models, but there is still room between the performance and the PLM-based methods. These results demonstrate that the performances of LLMs with external knowledge in a zero-shot setting can be comparable to or even surpass the fine-tuned PLM-based method on the RE task. They also underscore the effectiveness of our approach in multilingual settings, which is not limited to the English context.

Finally, the performance of the experimented models is consistent regardless of the language and the existence of external knowledge. Flan-T5 consistently achieve the best performance across all datasets, and Llama 2 exhibits comparatively lower performance, indicating that Flan-T5 has a better performance and a robust generalization advantage when dealing with the RE task and can be regarded as an ideal model in real-world application, while Llama 2 requires additional improvements for higher performance.

4.5. Additional Analysis

Effectiveness of External Knowledge First, we analyze the effectiveness of the external KB in our proposed method. Since not all entity pairs can be linked to Wikidata, we calculate the percentage of correct prediction of LLMs with and without the incorporation of external knowledge, denoted as $P_{w/know}$ and $P_{w/oknow}$, calculated as:

$$P_{w/know} = \frac{\text{\# of Correct Prediction}}{\text{\# of Entity Pairs Linked to Wikidata}}, \tag{1}$$

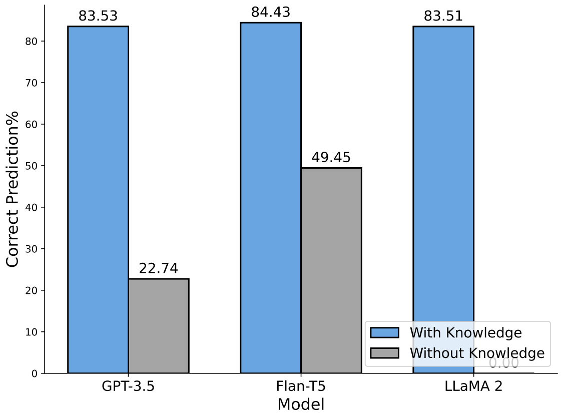

Figure 2: Percentage of correct relation prediction with and without external knowledge on the DocRED dataset.

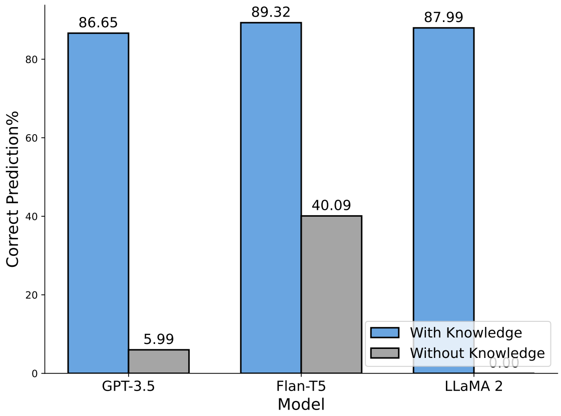

Figure 3: Percentage of correct relation prediction with and without external knowledge on the REBEL dataset.

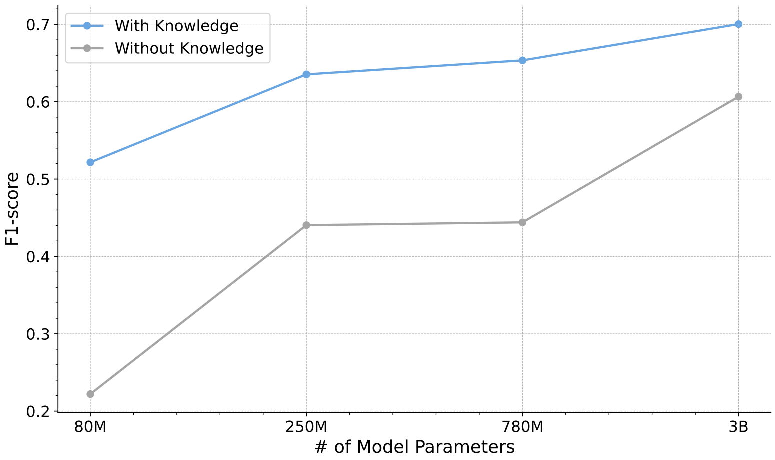

Figure 4: Scaling law of Flan-T5 on RE performance.

$$P_{w/oknow} = \frac{\text{\# of Correct Prediction}}{\text{\# of Entity Pairs Not Linked to Wikidata}}. \tag{2}$$

We visualize the calculation results in Figures 2 and 3, respectively. Our findings show a significant difference in performance with and without incorporating knowledge across all datasets and LLMs. Specifically, the correct prediction of LLMs with external knowledge is as high as more than 80%, while the results without knowledge are inferior, among which only Flan-T5 can exceed 40%. Because of the flexible and uncontrollable nature of Llama 2, its correct predictions without external knowledge are nearly zero, while after incorporating knowledge, its results improve to over 80%. The performance difference indicates that a performance gap exists across different models—while all models can achieve similar performance with external knowledge, their results are dominated by the relation classification result without external knowledge, indicating that LLMs are good *inferencers* but not *classifiers* for entity relationships. Moreover, although the performance of LLMs is better on the REBEL dataset with the incorporation of external knowledge, it becomes worse on the models without knowledge due to the large relation schema of the dataset. This remains a challenge for future research to design better methods to deal with the entity pairs that cannot link to the KBs.

Scaling Law We also analyze whether the performance of LLM-based RE can benefit from scaling up the model parameters. Specifically, we select the Flan-T5 series models with four different model sizes:

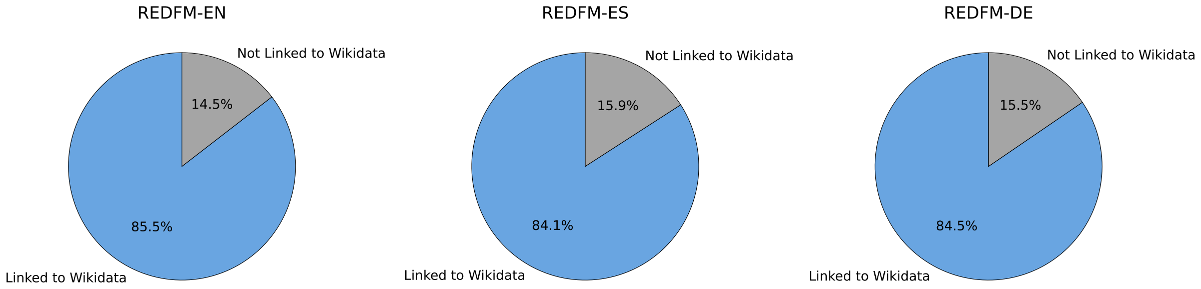

Figure 5: Knowledge coverage in the portion of the dataset we chose for the three languages.

Flan-T5-Small (80M), Flan-T5-Base (250M), Flan-T5-Large (780M), and Flan-T5-XL (3B) and evaluate the performance of the models with and without external knowledge. As shown in Figure 4, a clear positive scaling effect exists in LLM-based RE, i.e., fine-tuned larger models achieve better performance in the RE task. We can also observe the role of external knowledge. After incorporating external knowledge into the LLM, the increase in the number of model parameters has a smaller impact on the results. Moreover, with external knowledge, Flan-T5-Small can surpass Flan-T5-Large, and Flan-T5-Base can exceed Flan-T5-XL's performance without external knowledge. This validates the effectiveness of both LLMs and external knowledge when handling the RE task.

Coverage of Knowledge in Multilingualism Given the multilingual support of the chosen LLMs, we extend our investigation to include multilingual RE experiments using the REDFM dataset. The experimental outcomes, as summarized in Table 5, reveal subpar performance when the LLMs attempt multilingual RE tasks directly. However, integrating external knowledge significantly enhances performance, prompting us to explore the coverage of Wikidata for the selected multilingual dataset. To this end, we conduct supplementary experiments on REDFM-EN, REDFM-DE, and REDFM-ES to assess the percentage of samples that could be linked to Wikidata for external knowledge, as illustrated in Figure 5. The results indicate that a relatively high proportion of samples across the three languages could be covered by Wikidata, with coverages nearing or exceeding 85%, specifically 85.5%, 84.5%, and 84.1% for REDFM-EN, REDFM-DE, and REDFM-ES, respectively. The remaining 14.5%, 15.5%, and 15.9% are attributed to entries not indexed by Wikidata, with a small fraction being inaccessible due to unstable network connections.

5. Conclusion and Future Work

In this paper, we propose a novel framework to address the current challenges of LLMs falling short in RE tasks because of their *context-unawareness* and *schema-misalignment*, with *world knowledge ignorance*. It consists of two stages: entity linking and relation inference, fully leveraging the efficacy of KBs and LLMs in this task. We conduct experiments in a multilingual setting using three datasets and three LLMs to validate the effectiveness of our framework, where the zero-shot RE with world knowledge outperforms those without that by a significant margin and achieves state-of-the-art performance on all experimental datasets, even better than fine-tuned PLM-based methods, indicating the effectiveness of our proposed framework. We also conduct additional analysis on the effectiveness of knowledge, the impact of scaling up model parameters, and the coverage of knowledge in multilingualism to further demonstrate the effectiveness and generalizability of our proposed method. In the future, we will conduct more detailed analysis on other related tasks, such as event relation extraction, to further validate the effectiveness and generalizability of our proposed method.

6. Acknowledgments

This research is funded by the Postgraduate Research Scholarship (PGRS) at Xi’an Jiaotong-Liverpool University, contract number FOSSP221001.

References

[1] H. Peng, T. Gao, X. Han, Y. Lin, P. Li, Z. Liu, M. Sun, J. Zhou, Learning from Context or Names? An Empirical Study on Neural Relation Extraction, in: Proceedings of EMNLP, 2020, pp. 3661–3672. URL: https://aclanthology.org/2020.emnlp-main.298. doi:10.18653/v1/2020.emnlp-main.298.

[2] X. Li, F. Yin, Z. Sun, X. Li, A. Yuan, D. Chai, M. Zhou, J. Li, Entity-relation extraction as multi-turn question answering, in: Proceedings of ACL, 2019, pp. 1340–1350. URL: https://aclanthology.org/P19-1129. doi:10.18653/v1/P19-1129.

[3] A. Madotto, C.-S. Wu, P. Fung, Mem2Seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems, in: Proceedings of ACL, 2018, pp. 1468–1478. URL: https://aclanthology.org/P18-1136. doi:10.18653/v1/P18-1136.

[4] Q. Tan, R. He, L. Bing, H. T. Ng, Document-level relation extraction with adaptive focal loss and knowledge distillation, in: Findings of ACL, 2022, pp. 1672–1681. URL: https://aclanthology.org/2022.findings-acl.132. doi:10.18653/v1/2022.findings-acl.132.

[5] Y. Ma, A. Wang, N. Okazaki, DREEAM: Guiding attention with evidence for improving document-level relation extraction, in: Proceedings of EACL, 2023, pp. 1971–1983. URL: https://aclanthology.org/2023.eacl-main.145. doi:10.18653/v1/2023.eacl-main.145.

[6] Z. Wang, W. Wang, Q. Chen, Q. Wang, A. Nguyen, Generating valid and natural adversarial examples with large language models, 2023. arXiv:2311.11861.

[7] H. Na, Z. Wang, M. Maimaiti, T. Chen, W. Wang, T. Shen, L. Chen, Rethinking human-like translation strategy: Integrating drift-diffusion model with large language models for machine translation, 2024. arXiv:2402.10699.

[8] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: Proceedings of NeurIPS, volume 33, 2020, pp. 1877–1901. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

[9] J. Wei, X. Wang, D. Schuurmans, M. Bosma, b. ichter, F. Xia, E. Chi, Q. V. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, in: Proceedings of NeurIPS, volume 35, 2022, pp. 24824–24837. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf.

[10] H. Peng, X. Wang, J. Chen, W. Li, Y. Qi, Z. Wang, Z. Wu, K. Zeng, B. Xu, L. Hou, J. Li, When does in-context learning fall short and why? a study on specification-heavy tasks, 2023. arXiv:2311.08993.

[11] C. Si, D. Friedman, N. Joshi, S. Feng, D. Chen, H. He, Measuring inductive biases of in-context learning with underspecified demonstrations, in: Proceedings of ACL, 2023, pp. 11289–11310. URL: https://aclanthology.org/2023.acl-long.632. doi:10.18653/v1/2023.acl-long.632.

[12] P. Cao, X. Zuo, Y. Chen, K. Liu, J. Zhao, Y. Chen, W. Peng, Knowledge-enriched event causality identification via latent structure induction networks, in: Proceedings of ACL-IJCNLP, 2021, pp. 4862–4872. URL: https://aclanthology.org/2021.acl-long.376. doi:10.18653/v1/2021.acl-long.376.

[13] T. Lai, H. Ji, C. Zhai, Q. H. Tran, Joint biomedical entity and relation extraction with knowledge-enhanced collective inference, in: Proceedings of ACL-IJCNLP, 2021, pp. 6248–6260. URL: https://aclanthology.org/2021.acl-long.488. doi:10.18653/v1/2021.acl-long.488.

[14] A. Mallen, A. Asai, V. Zhong, R. Das, D. Khashabi, H. Hajishirzi, When not to trust language

models: Investigating effectiveness of parametric and non-parametric memories, in: Proceedings of ACL, 2023, pp. 9802–9822. URL: https://aclanthology.org/2023.acl-long.546. doi:10.18653/v1/2023.acl-long.546.

[15] M. Mintz, S. Bills, R. Snow, D. Jurafsky, Distant supervision for relation extraction without labeled data, in: Proceedings of ACL-AFNLP, 2009, pp. 1003–1011. URL: https://aclanthology.org/P09-1113.

[16] R. Hoffmann, C. Zhang, D. S. Weld, Learning 5000 relational extractors, in: Proceedings of ACL, 2010, pp. 286–295. URL: https://aclanthology.org/P10-1030.

[17] M. Chen, L. Huang, M. Li, B. Zhou, H. Ji, D. Roth, New frontiers of information extraction, in: Proceedings of NAACL-HLT (Tutorials), 2022, pp. 14–25. URL: https://aclanthology.org/2022.naacl-tutorials.3. doi:10.18653/v1/2022.naacl-tutorials.3.

[18] P.-L. Huguet Cabot, R. Navigli, REBEL: Relation extraction by end-to-end language generation, in: Findings of EMNLP, 2021, pp. 2370–2381. URL: https://aclanthology.org/2021.findings-emnlp.204. doi:10.18653/v1/2021.findings-emnlp.204.

[19] X. Zhao, M. Zhang, M. Ma, C. Su, Y. Liu, M. Wang, X. Qiao, J. Guo, Y. Li, W. Ma, HW-TSC at SemEval-2023 task 7: Exploring the natural language inference capabilities of ChatGPT and pre-trained language model for clinical trial, in: Proceedings of SemEval-2023, 2023, pp. 1603–1608. URL: https://aclanthology.org/2023.semeval-1.221. doi:10.18653/v1/2023.semeval-1.221.

[20] L. Pan, A. Albalak, X. Wang, W. Wang, Logic-LM: Empowering large language models with symbolic solvers for faithful logical reasoning, in: Findings of EMNLP, 2023, pp. 3806–3824. URL: https://aclanthology.org/2023.findings-emnlp.248. doi:10.18653/v1/2023.findings-emnlp.248.

[21] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Commun. ACM 57 (2014) 78–85. URL: https://doi.org/10.1145/2629489. doi:10.1145/2629489.

[22] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, M. Sun, DocRED: A large-scale document-level relation extraction dataset, in: Proceedings of ACL, 2019, pp. 764–777. URL: https://aclanthology.org/P19-1074. doi:10.18653/v1/P19-1074.

[23] L. Huguet Cabot, S. Tedeschi, A.-C. Ngonga Ngomo, R. Navigli, RED$^{\text{fm}}$: a filtered and multilingual relation extraction dataset, in: Proceedings of ACL, 2023, pp. 4326–4343. URL: https://aclanthology.org/2023.acl-long.237. doi:10.18653/v1/2023.acl-long.237.

[24] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, T. Scialom, Llama 2: Open foundation and fine-tuned chat models, 2023. arXiv:2307.09288.

[25] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, Y. Li, X. Wang, M. Dehghani, S. Brahma, A. Webson, S. S. Gu, Z. Dai, M. Suzgun, X. Chen, A. Chowdhery, A. Castro-Ros, M. Pellat, K. Robinson, D. Valter, S. Narang, G. Mishra, A. Yu, V. Zhao, Y. Huang, A. Dai, H. Yu, S. Petrov, E. H. Chi, J. Dean, J. Devlin, A. Roberts, D. Zhou, Q. V. Le, J. Wei, Scaling instruction-finetuned language models, 2022. arXiv:2210.11416.

[26] P. Verga, E. Strubell, A. McCallum, Simultaneously self-attending to all mentions for full-abstract biological relation extraction, in: Proceedings of NAACL-HLT, 2018, pp. 872–884. URL: https://aclanthology.org/N18-1080. doi:10.18653/v1/N18-1080.

[27] L. Baldini Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Matching the blanks: Distributional similarity for relation learning, in: Proceedings of ACL, 2019, pp. 2895–2905. URL: https://aclanthology.org/P19-1279. doi:10.18653/v1/P19-1279.

[28] S. Zeng, R. Xu, B. Chang, L. Li, Double graph based reasoning for document-level relation extraction, in: Proceedings of EMNLP, 2020, pp. 1630–1640. URL: https://aclanthology.org/2020.emnlp-main.127. doi:10.18653/v1/2020.emnlp-main.127.

[29] S. Zeng, Y. Wu, B. Chang, SIRE: Separate intra- and inter-sentential reasoning for document-level

relation extraction, in: Findings of ACL-IJCNLP, 2021, pp. 524–534. URL: https://aclanthology.org/2021.findings-acl.47. doi:10.18653/v1/2021.findings-acl.47.

[30] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, P. Wang, K-bert: Enabling language representation with knowledge graph, Proceedings of AAAI 34 (2020) 2901–2908. URL: https://ojs.aaai.org/index.php/AAAI/article/view/5681. doi:10.1609/aaai.v34i03.5681.

[31] X. Chen, N. Zhang, X. Xie, S. Deng, Y. Yao, C. Tan, F. Huang, L. Si, H. Chen, Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction, in: Proceedings of WWW, 2022, p. 2778–2788. URL: https://doi.org/10.1145/3485447.3511998. doi:10.1145/3485447.3511998.

[32] A. Roy, S. Pan, Incorporating medical knowledge in BERT for clinical relation extraction, in: Proceedings of EMNLP, 2021, pp. 5357–5366. URL: https://aclanthology.org/2021.emnlp-main.435. doi:10.18653/v1/2021.emnlp-main.435.

[33] H. Peng, X. Wang, F. Yao, Z. Wang, C. Zhu, K. Zeng, L. Hou, J. Li, OmniEvent: A comprehensive, fair, and easy-to-use toolkit for event understanding, in: Proceedings of EMNLP (Demo), 2023, pp. 508–517. URL: https://aclanthology.org/2023.emnlp-demo.46. doi:10.18653/v1/2023.emnlp-demo.46.

[34] R. Han, T. Peng, C. Yang, B. Wang, L. Liu, X. Wan, Is information extraction solved by chatgpt? an analysis of performance, evaluation criteria, robustness and errors, 2023. arXiv:2305.14450.

[35] B. Li, G. Fang, Y. Yang, Q. Wang, W. Ye, W. Zhao, S. Zhang, Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness, 2023. arXiv:2304.11633.

[36] K. Zhang, B. Jimenez Gutierrez, Y. Su, Aligning instruction tasks unlocks large language models as zero-shot relation extractors, in: Findings of ACL, 2023, pp. 794–812. URL: https://aclanthology.org/2023.findings-acl.50. doi:10.18653/v1/2023.findings-acl.50.

[37] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, S. Kurohashi, GPT-RE: In-context learning for relation extraction using large language models, in: Proceedings of EMNLP, 2023, pp. 3534–3547. URL: https://aclanthology.org/2023.emnlp-main.214. doi:10.18653/v1/2023.emnlp-main.214.

[38] J. Li, Z. Jia, Z. Zheng, Semi-automatic data enhancement for document-level relation extraction with distant supervision from large language models, in: Proceedings of EMNLP, 2023, pp. 5495–5505. URL: https://aclanthology.org/2023.emnlp-main.334. doi:10.18653/v1/2023.emnlp-main.334.

[39] W. Shen, J. Wang, J. Han, Entity linking with a knowledge base: Issues, techniques, and solutions, IEEE Transactions on Knowledge and Data Engineering 27 (2015) 443–460. doi:10.1109/TKDE.2014.2327028.