Re-Cent: A Relation-Centric Framework for Joint Zero-Shot Relation Triplet Extraction

Zehan Li, Fu Zhang*, **Kailun Lyu, Jingwei Cheng, Tianyue Peng**
School of Computer Science and Engineering, Northeastern University, China
Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, China
`{lizehan1999,LyuKaiLun,pengtianyue0716}@163.com`,
`{zhangfu,chengjingwei}@mail.neu.edu.cn`

Abstract

Zero-shot Relation Triplet Extraction (ZSRTE) aims to extract triplets from the context where the relation patterns are unseen during training. Due to the inherent challenges of the ZSRTE task, existing extractive ZSRTE methods often decompose it into named entity recognition and relation classification, which overlooks the interdependence of two tasks and may introduce error propagation. Motivated by the intuition that crucial entity attributes might be implicit in the relation labels, we propose a **Re**lation-**Cent**ric joint ZSRTE method named **Re-Cent**. This approach uses minimal information, specifically unseen relation labels, to extract triplets in one go through a unified model. We develop two span-based extractors to identify the subjects and objects corresponding to relation labels, forming span-pairs. Additionally, we introduce a relation-based correction mechanism that further refines the triplets by calculating the relevance between span-pairs and relation labels. Experiments demonstrate that Re-Cent achieves state-of-the-art performance with fewer parameters and does not rely on synthetic data or manual labor.

1 Introduction

Relation Triplet Extraction (RTE) aims to extract entity pairs and their relations from context and organize them into triplet: `<subject entity, relation, object entity>` (Han et al., 2020). The existing research on RTE focuses on end-to-end joint extraction models (Wei et al., 2020; Ning et al., 2023) rather than pipeline methods (Zelenko et al., 2003) which run Named Entity Recognition (NER) and Relation Classification (RC) separately (Yuan et al., 2021). However, these approaches require a large amount of annotated data to retrain for novel relations (Yu et al., 2024), which limits their scalability in real-world scenarios.

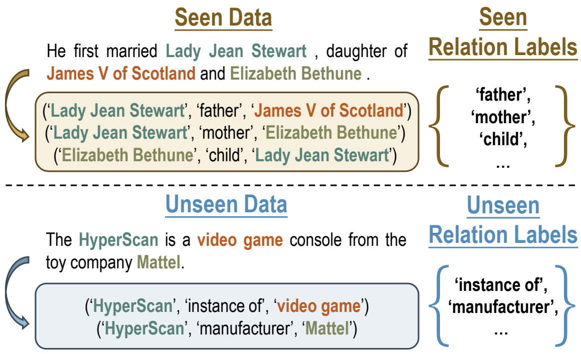

Figure 1: Examples for Zero-shot RTE. Each context may contain multiple triplets. The intersection of seen and unseen relation labels is empty.

Zero-shot learning (ZSL) transfers knowledge from seen classes to unseen ones, alleviating the reliance on annotated data (Wang et al., 2019). Existing efforts focus on Zero-shot Relation Classification (ZSRC)[1] (Chen and Li, 2021; Li et al., 2024b), which classifies the relations between *annotated* entities. However, annotated entities are not always provided. Chia et al. (2022) proposed Zero-shot RTE (ZSRTE), which can directly extract triplets from context *without annotated* data for training unseen relation patterns (Figure 1).

Existing ZSRTE methods can be roughly divided into two categories. *Generative methods* typically rely on additional synthetic data (Xu et al., 2024) or handcrafted templates (Kim et al., 2023) to directly generate triplets. Although Large Language Models (LLMs) can perform ZSRTE due to their ZSL capabilities (Xu et al., 2023), they often come with high inference (Wei et al., 2023) or training costs (Li et al., 2024a). *Extractive methods* (Lv et al., 2023; Gong and Eldardiry, 2024) decompose RTE into two tasks: NER and RC, due to the complexity of the task, but this increases deployment costs and also introduces the risk of error propagation.

* Corresponding author.

[1]In some works, it is referred to as Zero-shot Relation Extraction (ZSRE). In this paper, we refer to it as ZSRC.

Proceedings of the 31st International Conference on Computational Linguistics, pages 7344–7354
January 19–24, 2025.

We argue that a joint extractive ZSRTE system could be appealing. During inference, only relation labels are available, so we consider them as the pivot of ZSRTE. Despite their brevity, we believe these labels are rich in semantics and have the potential to offer valuable insights regarding the subject and object involved. For example, for the relation `<Director>`, the subject may be a movie or TV show, while the object is always a person. However, for the relation `<Instance of>`, we can only infer that the subject is a specific individual and the object represents an abstract class or type. We believe that such nuanced entity information, which is difficult to formalize, could be more effectively learned by models.

Based on this intuition, we propose **Re-Cent**, a **Re**lation-**Cent**ric joint zero-shot relation triplet extraction framework. We extract triplets in a unified model, relying solely on relation labels and avoiding the involvement of pseudo-data and manual labor. Specifically, inspired by the type and span matching approach in the NER task (Zaratiana et al., 2024), we design two span extractors to separately extract subjects and objects based on relation labels, resulting in a subject set containing n spans and an object set containing m spans. We pair them to form $n \times m$ span pairs. However, when multiple triplets are present in the context, there is a redundancy risk. To address this, we introduce a span-pair correction mechanism that re-scores the span-pairs with their relations to filter out irrelevant triplets. Re-Cent treats zero-shot NER and zero-shot RC as a unified task and optimizes them collaboratively, thus avoiding error propagation. Our main contributions are as follows:

- We introduce a novel perspective to revisit ZSRTE, which leverages the implicit information within relation labels to extract the subject and object, thereby achieving joint extraction.
- We propose a relation-based correction mechanism that improves the performance of multi-triplet extractions by introducing a relevance score between span-pairs and relations.
- Extensive experiments show that Re-Cent outperforms state-of-the-art methods, achieving **12.73** and **6.77** gains in F_1 scores on the two widely used datasets, respectively. Moreover, Re-Cent performs better than LLMs-based methods with significantly fewer model parameters.

2 Related Work

2.1 Joint Relation Triplet Extraction

Previous pipeline-based methods (Chan and Roth, 2011) decompose the RTE task into NER and RC, which overlook the interdependence between the two tasks (Shen et al., 2021) and introduce the risk of error propagation (Zhong and Chen, 2021). Recent research focuses on end-to-end extraction models, which alleviate the limitations of pipeline models through joint optimization. Common paradigms include span pair-classification (Wadden et al., 2019; Ji et al., 2020), table-filling (Shang et al., 2022; Ning et al., 2023), and seq2seq models (Zeng et al., 2018, 2020). However, a common limitation of these methods is their reliance on annotated data for novel relations. As a result, there has been growing interest in RTE tasks under few-shot and zero-shot settings (Deng et al., 2022), which require limited or no annotated data for unseen relations.

2.2 Zero-shot Relation Triplet Extraction

The existing research on Zero-shot relation classification (Chen and Li, 2021) uses methods such as prototype-matching (Zhao et al., 2023; Li et al., 2024b), text-entailment (Obamuyide and Vlachos, 2018; Sainz et al., 2021) or Large Language Models (LLMs) (Zhang et al., 2023; Li et al., 2023) to classify the relation between two *annotated* entities. However, in real-world scenarios, the annotated entities are not always given. To this end, Chia et al. (2022) proposed a challenging task called Zero-shot Relation Triplet Extraction (ZSRTE), which can be roughly divided into two types:

Generative methods aim to directly generate relation triplets. RelationPrompt (Chia et al., 2022) designs a Generator-Extractor framework, and subsequent work (Xu et al., 2024) incorporates reinforcement learning to generate higher-quality additional samples. Kim et al. (2023) manually construct templates for each relation and perform generation through template-filling. Zero-shot capable LLMs have also been applied to the ZSRTE task, utilizing multi-turn Q&A (Wei et al., 2023) or fine-tuning with tabular prompts (Li et al., 2024a).

Extractive methods identify the positions of entities and classify their relations. Due to the complexity of ZSRTE, existing work decomposes ZSRTE into two stages: NER and RC, executing them in a pipeline manner. Lv et al. (2023) introduced discriminative soft prompt, fine-tuning the

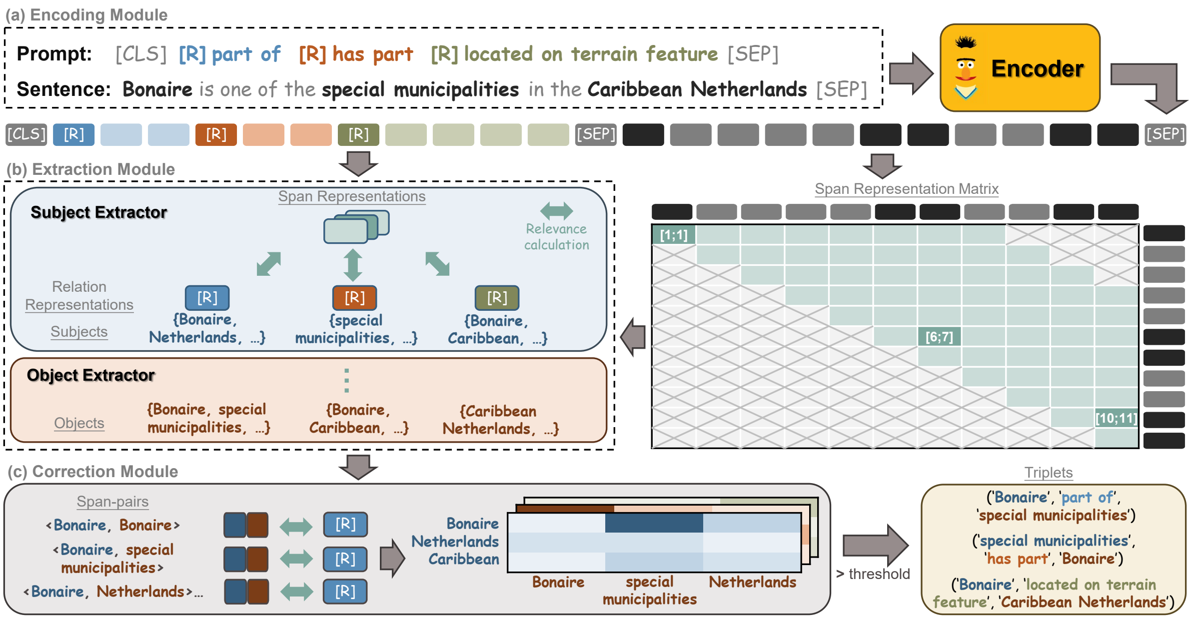

Figure 2: An illustration of the proposed **Re-Cent**. Take an example with unseen relation $m = 3$, a maximum span length $L = 8$. The input sentence is concatenated with the unseen relation labels and fed into the **(a) Encoding Module** (§3.3) to generate contextual representations. Then, the **(b) Extraction Module** (§3.4) identifies potential subjects and objects from the span representation matrix corresponding to the relation patterns. Finally, the **(c) Correction Module** (§3.5) filters out redundant triplets by calculating the relevance scores between span-pairs and unseen relation labels, resulting in the final triplets.

NER and RE models separately, while Gong and Eldardiry (2024) generate data for unseen relations based on a knowledge graph, and use an additional extractor to obtain entities. Although running in stages can simplify the complexity of ZSRTE (Lan et al., 2024), the lack of interaction between tasks may lead to potential performance degradation and is susceptible to error propagation.

Different from existing methods, we treat ZSRTE as a relation-centric subject-object span classification task and jointly optimize the model to minimize error propagation. Additionally, by further introducing the relevance scores between span-pairs and relation labels, we further improve extraction efficiency. Our framework uses minimal information, specifically unseen relation labels, without requiring additional generated training data.

3 Methodology

3.1 Task Formulation

Given a dataset $D = (\{(X_i, \Gamma_i)\}_{i=1}^{|D|}, R)$, where X_i represents the i-th input sentence and Γ_i represents the latent triplet set in X_i. The goal of Relation Triplet Extraction (RTE) is to extract $\Gamma_i = \{(e_j^{subj}, r_j, e_j^{obj})\}_{j=1}^{|\Gamma|}$ from X_i, where e_j^{subj}, $e_j^{obj} \in E$ represent the subject and object entity, respectively, with E being the set of all entities in X_i. $r_j \in R$ and R is a predefined set of relations.

The Zero-shot RTE (ZSRTE) requires learning from the seen dataset D_s and transferring the capabilities to predict the unseen dataset D_u. D_s and D_u are originally split from D. Similarly, given seen relations $R_s = \{r_1^s, \ldots, r_n^s\} \in D_s$ and unseen relations $R_u = \{r_1^u, \ldots, r_m^u\} \in D_u$, where n and m represent the number of relations. It is worth noting that there is no overlap between R_s and R_u. Based on ZSRTE, the goal of ZSRC is to identify the relation $r_u \in R_u$ between the annotated entity pairs (e_u^{subj}, e_u^{obj}) in the unseen dataset D_u.

3.2 Overview

An overview of our proposed model is shown in Figure 2. It includes three main components:

(1) **Prompt and Sentence Encoding** aims to utilize a pre-trained bidirectional encoder to generate representations for relation labels and sentences. (2) **Relation-Centered Extractors** are used to identify the sets of subjects and objects for specific relation patterns. (3) **Multiple Triplet Correction Module** is designed to assess the association between subject-object pairs and relations, filtering out irrelevant triplets.

3.3 Prompt and Sentence Encoding

To ensure full interaction between the sentence and the relation labels while enhancing the distinction between different relation labels, inspired by previous works (Lv et al., 2023; Zaratiana et al., 2024), we concatenate all the candidate relation labels into the prompt and obtain the representation of both prompt and sentence through a bidirectional transformer encoder . Specifically, for a given sentence $X = \{w_1, w_2, \ldots, w_l\}$, we prepend a prompt $X_p = \{[\mathcal{R}]r_1, [\mathcal{R}]r_2, \ldots, [\mathcal{R}]r_n\}$ to X, where l and n denote the number of tokens in the sentence and the number of candidate relations, respectively. The symbol $[\mathcal{R}]$ is a learnable marker that is concatenated before each relation r_i to represent the semantic of the i-th relation. The encoded hidden state embeddings can be expressed as:

$$\begin{aligned}&\{h_{[CLS]}, h_{[\mathcal{R}]}, h_{r_1}, \ldots, h_{[\mathcal{R}]}, h_{r_n}, h_{[SEP]},\\ &h_{w_1}, \ldots, h_{w_l}, h_{[SEP]}\} = \text{Encoder}(X_p X).\end{aligned} \quad (1)$$

3.4 Relation-Centered Extractors

We consider relations as a bridge between the subject and the object, which may encompass rich entity attributes such as role types, semantic directions, and hierarchical structures.

Taking the sentence in Figure 2 as an example, for the relation `<part of>`, the subject `<Bonaire>` as a geographic entity, is a part of the object `<special municipalities>`, which is an administrative division. The semantic direction between the two entities clarifies a hierarchical relationship from the specific to the general. However, these abstract attributes may be difficult to express explicitly in natural language. Therefore, we develop two *Relation-Centered* extractors, which are designed to explore the semantics within relations and to extract the sets of subjects and objects.

Span Representation Matrix. Initially, we represent entities using spans and construct a representation matrix $\mathcal{S}$:

$$\mathcal{S} = \{s^{ij} = [h_i; h_j] | i, j \in [1, l], i \le j, j - i \le L\}, \quad (2)$$

where s^{ij} represents the entity span from the i-th token to the j-th token, $[;]$ denotes the concatenation operation, and L indicates the maximum length of spans to limit computational complexity.

Subject / Object Extractor. We assume that each span in a sentence has the potential to serve as either the subject or object in a triplet, depending on its role in the relationship. To extract the subject set E_k^{subj} and the object set E_k^{obj} corresponding to the k-th relation label r_k, we compute the relevance scores $\mathcal{P}_{subj}$ and $\mathcal{P}_{obj}$ for each entity span when considered as a subject and as an object, respectively. These scores are represented as $\mathcal{P}_{subj}, \mathcal{P}_{obj} \in \mathbb{R}^{|\mathcal{S}| \times n}$:

$$\mathcal{P}_{subj} = \|_{k=1}^{n} \sigma(\sum_d \text{FFN}_{subj}(s_i) \times \text{FFN}_{rel}(h_{[\mathcal{R}],k})), \quad (3)$$

$$\mathcal{P}_{obj} = \|_{k=1}^{n} \sigma(\sum_d \text{FFN}_{obj}(s_i) \times \text{FFN}_{rel}(h_{[\mathcal{R}],k})), \quad (4)$$

where σ denotes the sigmoid activation function, d represents the token embedding dimension, and FFN is a two-layer feed-forward network, $s_i \in \mathcal{S}$. We add entity spans to E_k if their relevance score with r_k exceeds a threshold θ.

3.5 Multiple Triplet Correction Module

Directly pairing the spans in the set without considering the connection between the subject and the object may introduce redundant triplets. Therefore, we additionally introduce span-pair relevance scores. For the k-th relation, we first construct a span-pair representation matrix $\mathcal{T}$ from the spans in E_j^{subj} and E_j^{obj}:

$$\mathcal{T}_k = \{t_k^{ij} = [s_{subj}^{i}; s_{obj}^{j}] | s_{subj} \in E_k^{subj}, s_{obj} \in E_k^{obj}\}. \quad (5)$$

The relevance score $\mathcal{P}_{pair} \in \mathbb{R}^{|\mathcal{T} \times n|}$ is then calculated using the relation r_k associated with $\mathcal{T}_k$:

$$\mathcal{P}_{pair} = \|_{k=1}^{n} \sigma(\sum_d \text{FFN}_{pair}(t_k^{ij}) \times \text{FFN}_{rel}(h_{[\mathcal{R}],k})), \quad (6)$$

where $t_k^{ij} \in \mathcal{T}_k$. We take the triplets with a relevance score greater than the threshold τ as the final extraction results.

3.6 Training Objective

The training objectives include two main goals: for span extractors, the aim is to increase the relevance between spans and positive relations while decreasing their relevance with negative relations. Similarly, the span-pair scorer is designed to assign higher scores to positive span-pairs. Notably, to enhance robustness while reducing computational complexity, we select the top $\mathcal{K}$ spans with the highest relevance from the set E as hard negative samples for computation in the matrix $\mathcal{T}$. We use binary cross-entropy loss:

$$\mathcal{L}(\mathcal{P}) = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)], p_i \in \mathcal{P}, \quad (7)$$

where $y = 1$ when the relation is positive, and $y = 0$ otherwise. The final loss function needs to unify the two training objectives. We use a hyper-parameter α to control the weight of each task to obtain the final optimization objective:

$$Loss = \alpha(\mathcal{L}(\mathcal{P}_{subj}) + \mathcal{L}(\mathcal{P}_{obj})) + (1-\alpha)\mathcal{L}(\mathcal{P}_{pair}). \quad (8)$$

4 Experiments Setup

4.1 Datasets

We evaluate our framework on two wildly used datasets, FewRel (Han et al., 2018) and Wiki-ZSL (Chen and Li, 2021). The key difference is that FewRel is manually annotated, while Wiki-ZSL is constructed through distant supervision, which may introduce more noise (Li et al., 2024b). Data statistics can be found in Table 1. To ensure a fair and comprehensive comparison, we follow the setup of previous works (Chia et al., 2022) by splitting the data into training, validation, and test sets. To maintain a zero-shot setting, there is no overlap in relation labels across these data folds. Specifically, for each dataset, the test set contains m unseen relation labels, and we evaluate our method under the settings of $m = \{5, 10, 15\}$. The validation set contains 5 unseen labels for early stopping and hyperparameter tuning, while the remaining labels are included in the training set.

4.2 Implementation Details

The hyperparameters are determined manually on the valid set using grid search. We use AdamW as the optimizer, applying different learning rates of $1e-5$ and $3e-5$ for the encoder and other layers, respectively. We set the batch size to 32 and 10 epochs and apply an early stopping strategy. The maximum span length L is set to 12. The ratio of positive-to-negative relation type is $1:3$, the number of hard negative spans $\mathcal{K} = 5$, and the loss weight $\alpha = 0.5$. For the selection of thresholds θ and τ, we conduct a detailed analysis of two benchmarks in §5.6 and set the same optimal thresholds for all experiments. We choose DeBERTa-v3-base (He et al., 2023), a bidirectional transformer encoder, as the backbone to evaluate performance.

For the metrics of ZSRTE, we align with the settings of Chia et al. (2022), employing Accuracy ($Acc.$) as the metric for **Single Triplet** evaluation, where each sentence contains only one triplet. For **Multi Triplet** evaluation, where each sentence contains two or more triplets, we report the standard micro precision ($P.$), recall ($R.$), and F_1 score. To reduce experimental randomness, we report the results as the average performance across five data folds, with the data splits sourced from previous works (Chia et al., 2022; Kim et al., 2023). Our code is available[2].

Dataset	Instances	Entities	Relations	Avg. Len.
Wiki-ZSL	94,383	77,623	113	24.85
FewRel	56,000	72,954	80	24.95

Table 1: The statistics for two public datasets. Avg. Len. represents the average length of each sentence.

4.3 Compared Baselines

We compare our proposed model with the following seven strong baselines in ZSRTE. For *generative* methods, **TableSequence** (Wang and Lu, 2020) and **RelationPrompt** (Chia et al., 2022) are trained using synthetic data of unseen relations. **TAG** (Xu et al., 2024) enhances data quality based on these methods through reinforcement learning. **ZETT** (Kim et al., 2023) utilizes handcrafted relation templates to directly generate the corresponding entities. For *extractive* methods, **DSP** (Lv et al., 2023) constructs different discriminative models for NER and RE tasks. **RSED** (Lan et al., 2024) filters potential relations and detects entity boundaries based on the relations. **ZS-SKA** (Gong and Eldardiry, 2024) introduces different levels of semantic information for knowledge augmentation.

Our span-based extractive method eliminates the reliance on pseudo-data compared to generative methods, and the most notable difference from extractive methods like DSP and ZS-SKA is that we jointly optimize NER and RE with a unified model.

5 Analysis and Discussion

5.1 Main Results

In Table 2, we compare our main experimental results with previous methods and observe that Re-Cent *achieves significant performance improvements* on two public datasets:

In the Single Triplet evaluation, we surpass the SOTA method (ZS-SKA) by **3.34** and **7.74** on Wiki-ZSL and FewRel, respectively. In the Multi Triplet evaluation, we also achieve notable advantages, with an absolute F1 improvement of **12.73** and **6.77** over ZS-SKA on the two benchmarks, facilitated by a more balanced precision-recall ratio.

[2]https://github.com/lizehan1999/Re-Cent

Unseen Labels	Methods	Single Triplet Wiki-ZSL	Single Triplet FewRel	Multi Triplet Wiki-ZSL			Multi Triplet FewRel		
		Acc.	*Acc.*	*P.*	*R.*	F_1	*P.*	*R.*	F_1
m=5	TableSequence (Wang and Lu, 2020)	14.47	11.82	43.68	3.51	6.29	15.23	1.91	3.40
	NoGen (Chia et al., 2022)	9.05	11.49	15.58	43.23	22.26	9.45	36.74	14.57
	RelationPrompt (Chia et al., 2022)	16.64	22.27	29.11	31.00	30.01	20.80	24.32	22.34
	ZETT (Kim et al., 2023)	21.49	30.71	35.89	28.38	31.74	38.14	30.58	33.71
	DSP (Lv et al., 2023)	-	-	42.70	43.40	43.00	40.10	27.00	32.30
	RSED (Lan et al., 2024)	18.40	22.67	38.14	36.84	37.48	43.91	34.97	38.93
	ZS-SKA (Gong and Eldardiry, 2024)	**44.00**	32.86	66.70	27.24	38.68	57.50	26.24	36.04
	TAG (Xu et al., 2024)	23.12	28.94	39.36	37.51	38.24	37.56	40.24	38.81
	Re-Cent (ours)	43.32	**46.18**	53.90	58.55	**55.66**	46.88	44.56	**44.80**
m=10	TableSequence (Wang and Lu, 2020)	9.61	12.54	45.31	3.57	6.40	28.93	3.60	6.37
	NoGen (Chia et al., 2022)	7.10	12.40	9.63	45.01	15.70	6.40	41.70	11.02
	RelationPrompt (Chia et al., 2022)	16.48	23.18	30.20	32.31	31.19	21.59	28.68	24.61
	ZETT (Kim et al., 2023)	17.16	27.79	24.49	26.99	24.87	30.65	32.44	31.28
	DSP (Lv et al., 2023)	-	-	26.30	48.00	34.00	35.90	27.10	30.90
	RSED (Lan et al., 2024)	22.30	24.91	27.09	39.09	32.00	30.89	29.90	30.39
	ZS-SKA (Gong and Eldardiry, 2024)	26.40	34.03	45.38	29.27	35.30	60.48	23.22	33.28
	TAG (Xu et al., 2024)	17.24	28.16	31.37	32.53	31.88	31.04	33.49	32.18
	Re-Cent (ours)	**30.30**	**36.53**	42.22	50.56	**45.95**	39.87	39.10	**39.05**
m=15	TableSequence (Wang and Lu, 2020)	9.20	11.65	44.43	3.53	6.39	19.03	1.99	3.48
	NoGen (Chia et al., 2022)	6.61	10.93	7.25	44.68	12.34	4.61	36.39	8.15
	RelationPrompt (Chia et al., 2022)	16.16	18.97	26.19	32.12	28.85	17.73	23.20	20.08
	ZETT (Kim et al., 2023)	12.78	26.17	19.45	23.31	21.21	22.50	27.09	24.39
	DSP (Lv et al., 2023)	-	-	27.70	32.40	29.90	27.90	25.40	26.60
	RSED (Lan et al., 2024)	21.64	25.14	25.37	33.80	28.98	27.00	23.55	25.16
	ZS-SKA (Gong and Eldardiry, 2024)	20.26	23.86	31.23	27.20	29.19	37.29	19.13	25.29
	TAG (Xu et al., 2024)	16.41	22.53	26.52	31.34	29.18	25.35	25.88	25.59
	Re-Cent (ours)	**27.06**	**31.27**	35.79	45.19	**39.75**	30.53	32.53	**31.07**

Table 2: Main comparison results of the proposed Re-Cent with the prior works. **Bold** marks the highest score, underline marks the second-best score. All baseline results are sourced from the original papers.

This indicates that our framework adapts better to the challenging ZSRTE task compared to previous generative and extractive baselines. This may be attributed to our joint optimization of NER and RC as a unified objective, which enables information sharing while avoiding error accumulation.

Additionally, compared to the baseline methods, Re-Cent shows a larger performance improvement on Wiki-ZSL, suggesting that it is more robust in handling noisy data, which is beneficial for real-world scenarios. It is worth noting that our method completes the ZSRTE task based on a unified model without requiring additional manual labor or the intervention of pseudo-data, further reducing the deployment cost.

5.2 Comparison with Large Language Models

Generative Large Language Models (LLMs) possess strong zero-shot learning capabilities. It may raise concerns about whether Re-Cent can outperform LLMs-based methods. In Table 3, we report the performance of four generative baselines alongside the parameter count of the backbone models, particularly MICRE, which fine-tuned LLMs like LLaMA-7B (Touvron et al., 2023) on 12 public datasets through tabular-prompting. Additionally, we compare the performance of Re-Cent at different parameter scales by replacing the backbone model with three different sizes of DeBERTa-v3.

The results show that *even with fewer parameters, our model consistently outperforms most generative methods*. Moreover, as the backbone model size increases, there is a significant performance boost, emphasizing the scalability of Re-Cent. We also observe that the large-sized backbone makes Re-Cent's advantages more pronounced, with performance exceeding LLaMA by 39% with only 4% of LLaMA's parameters.

5.3 Scalability Analysis of Re-Cent on ZSRC

We make a simple modification to the structure of Re-Cent to extend it to the zero-shot relation

Methods	Backbone Params	Wiki-ZSL m=5	Wiki-ZSL m=10	Wiki-ZSL m=15	FewRel m=5	FewRel m=10	FewRel m=15	*Avg.*
RelationPrompt$_{BART\&GPT-2}$ †	264M	16.64	16.48	16.16	22.27	23.18	18.97	18.95
ZETT$_{T5-base}$ ‡	220M	21.49	17.27	12.78	30.71	27.90	26.17	22.72
TAG$_{BART\&GPT-2}$ †	264M	23.12	17.24	16.41	28.94	28.16	22.53	22.73
MICRE $_{T5-3B}$ ‡	3,000M	25.20	23.65	21.80	36.75	33.18	30.44	28.50
MICRE $_{LLaMA}$ ‡	7,000M	27.74	24.64	22.23	37.53	34.77	<u>32.42</u>	29.89
Re-Cent (ours) $_{DeBERTa_{-v3-small}}$	44M	37.53	27.16	25.35	39.44	32.16	25.19	31.14
Re-Cent (ours) $_{DeBERTa_{-v3-base}}$	86M	<u>43.32</u>	<u>30.30</u>	<u>27.06</u>	<u>46.18</u>	<u>36.53</u>	31.27	<u>35.78</u>
Re-Cent (ours) $_{DeBERTa_{-v3-large}}$	304M	**48.74**	**35.47**	**33.24**	**48.44**	**43.30**	**40.76**	**41.66**

Table 3: Accuracy comparison results with LLMs-based methods under the single triplet setting. We report the backbone models used in their baselines along with the corresponding number of parameters. † and ‡ respectively mark the results reported by Xu et al. (2024) and Li et al. (2024a). *Avg.* denotes the average of accuracy scores.

Methods	Wiki-ZSL *P.*	Wiki-ZSL *R.*	Wiki-ZSL F_1	FewRel *P.*	FewRel *R.*	FewRel F_1
ZS-BERT	34.12	34.38	34.25	35.54	38.19	36.82
REPrompt	63.69	**67.93**	65.74	74.33	72.51	73.40
RE-Matching	67.31	67.33	67.32	73.80	73.52	73.66
SUMASK	43.55	40.27	41.85	44.76	41.13	42.87
ZS-SKA	41.78	40.50	39.30	45.03	51.86	46.99
MICRE	67.14	68.87	67.99	73.47	75.83	74.77
AlignRE	<u>69.01</u>	<u>67.52</u>	<u>68.26</u>	**77.63**	**77.00**	**77.31**
Re-Cent (RC)	**69.83**	67.49	**68.61**	<u>77.24</u>	<u>74.90</u>	<u>76.04</u>

Table 4: ZSRC experiment of Re-Cent. We report the performance on FewRel and Wiki-ZSL under m=15. All baseline results are sourced from the original papers.

classification (ZSRC) task for more insights. Since entities are pre-annotated in ZSRC, we remove the span extraction objective and directly concatenate the span representations of the subject and object as the sentence representation. The triplet correction module is repurposed as a relation predictor to provide a relevance score for each relation, with the highest score selected as the final prediction. The positive-to-negative relation type ratio is set to 1:10, and the other hyperparameters remain the same as in the ZSRTE task. We compare seven strong baselines: ZS-BERT (Chen and Li, 2021), REPrompt (Chia et al., 2022), RE-Matching (Zhao et al., 2023), SUMASK (Li et al., 2023), ZS-SKA (Gong and Eldardiry, 2024), MICRE (Li et al., 2024a) and AlignRE (Li et al., 2024b). Following the settings of previous works, we use the Macro F_1 metric for evaluation.

The comparison results in Table 4 indicate that Re-Cent *demonstrates strong competitiveness and generalization in ZSRC*, particularly on the noisier Wiki-ZSL dataset. However, compared to the method specifically optimized for ZSRC

Methods	m=5	m=10	m=15
Re-Cent	**44.80**	**39.10**	**31.07**
w/o. Two Extractors	40.08	31.06	23.61
w/o. Triplet Correction	41.91	34.08	27.49
w/o. both	5.69	5.55	3.70

Table 5: Ablation study of Re-Cent. We report the multi triplet F_1 performance on FewRel under m={5,10,15}.

(AlignRE), there remains a slight performance gap on FewRel, which may be due to the absence of side information such as descriptions and aliases. Therefore, leveraging additional information sources beyond relation labels to improve performance could be a promising direction.

5.4 Ablation Experiment

We report the impact of each component under multi-triplet setting on FewRel in Table 5.

We use a unified extractor to extract a single span set for each relation label, replacing the subject extractor and the object extractor (**w/o.** Two Extractors). The performance drop reveals the importance of separately identifying the subject and object. The reason may be that when matching all spans without distinguishing directions, redundant triplets burden the correction module. When we omit the span-pair matching score to filter triplets (**w/o.** Triplet Correction), the performance declines, indicating that the module effectively identifies irrelevant triplets by computing the relevance between entity pairs and relations. When we remove both components (**w/o.** both), there is a dramatic performance drop. This is because the triplets lose directional information and lack additional correction, leading to a deviation in the optimization objective and thus making the extraction ineffective.

Dataset	Methods	Subject	Object	Triplets
Wiki-ZSL	*Original*	25.78	29.90	27.06
	Reversed	8.70 ↓	7.16 ↓	1.46 ↓
FewRel	*Original*	31.25	29.76	31.27
	Reversed	7.43 ↓	8.97 ↓	3.28 ↓

Table 6: Experimental analysis of the role of Relation-Centric for subject extraction, object extraction, and single triplet extraction when m=15 (*Acc.*). *Reversed* indicates that the two extractors are swapped during the inference phase based on the *original* model.

Sentence:	**ATLAS Hydrographic** is an oceanographic systems company, part of the **ATLAS Elektronik** group that is owned by **thyssenkrupp** and **Airbus**.	
Subject Set:	ATLAS Hydrographic	ATLAS Elektronik
Object Set:	thyssenkrupp	Airbus
Triplets:	<ATLAS Elektronik, owned by, thyssenkrupp>	
	<ATLAS Elektronik, owned by, Airbus>	

Sentence:	**Bernard Hopkins** vs. **Chad Dawson** was a boxing match contested for both the **WBC** and "The Ring" **light heavyweight** championships.	
Subject Set:	Bernard Hopkins	Chad Dawson
Object Set:	WBC	light heavyweight
Triplets:	<Bernard Hopkins, competition class, light heavyweight>	
	<Chad Dawson, competition class, light heavyweight>	

Figure 3: Two case studies of extraction results on FewRel, where we provide the extracted subject set, object set, and all triplets.

5.5 Relation-Centric Role Analysis

To examine whether Re-Cent effectively captures the entity attribute information in relation labels and accurately identifies the corresponding subject and object, we conduct an experimental analysis. During the inference phase, we reverse the two extractors (Table 6), such as using the subject extractor to identify the object. We observe a significant decrease in performance, indicating that our framework effectively captures attribute information from the relation labels and focuses on extracting the corresponding subject and object entity sets. Furthermore, in Figure 3, we provide two examples of Re-Cent extraction results. We observe that there is no overlap between the subject set and the object set, and Re-Cent achieves more accurate extraction by filtering redundant triplets.

5.6 Hyperparameter Study

Thresholds can affect the performance of Re-Cent, prompting us to conduct a two-factor experiment to explore the entity span threshold (θ) and the triplet threshold (τ), which are described in §3.4 and §3.5, respectively. We conduct experiments

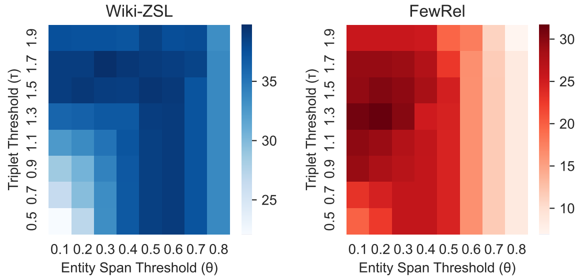

Figure 4: Threshold study on two datasets with m=15. Darker blocks represent higher multi-triplet F_1 scores under the corresponding group of thresholds.

with $m = 15$ on two datasets and report the performance under different thresholds in Figure 4. We observe that when θ is smaller, adjusting τ makes it easier to achieve optimal performance. This is because a smaller θ recalls more positive spans, reducing omissions, which allows the correction module to more precisely prune irrelevant triplets through τ. However, as θ decreases further, this advantage diminishes. Additionally, the Wiki-ZSL dataset seems to require higher thresholds to achieve good performance, which may be due to its noisier nature, necessitating higher confidence to ensure robust predictions. We select the best thresholds for each benchmark, $\theta = 0.3, \tau = 1.7$ for Wiki-ZSL and $\theta = 0.2, \tau = 1.3$ for FewRel, and apply them in all experiments.

6 Conclusion

In this paper, we present a novel approach named Re-Cent to zero-shot relation triplet extraction (ZSRTE) which adopts a relation-centric joint extraction framework. By leveraging the implicit information within relation labels, Re-Cent simultaneously extracts subjects and objects, forming triplets without relying on pseudo-data or manual labor. The introduction of a relation-based span-pair correction mechanism further enhances its ability to handle multi-triplet extractions by filtering irrelevant triplets. Extensive experiments demonstrate that Re-Cent outperforms existing state-of-the-art methods, providing a more efficient and scalable solution for ZSRTE while significantly reducing model complexity compared to LLMs-based methods. Our work highlights the potential of using an end-to-end framework in the complex ZSRTE task. In the future, we will further investigate how to migrate this approach to other zero-shot tasks.

Limitations

Despite Re-Cent demonstrating advantages in the ZSRTE task, there are still some limitations.

Firstly, Re-Cent relies on a strong encoder backbone and does not exhibit performance advantages on models like bert-base. Additionally, the span-based Re-Cent fails to handle discontinuous entities, which distinguishes it from generative methods. Furthermore, we found from experiments that the value of the threshold has a significant impact on Re-Cent, which also prompts us to explore the possibility of dynamic threshold methods.

Acknowledgments. The authors sincerely thank the reviewers for their valuable comments, which improved the paper. The work is supported by the National Natural Science Foundation of China (62276057), and Sponsored by CAAI-MindSpore Open Fund, developed on OpenI Community.

References

Yee Seng Chan and Dan Roth. 2011. Exploiting syntactico-semantic structures for relation extraction. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 551–560.

Chih-Yao Chen and Cheng-Te Li. 2021. Zs-bert: Towards zero-shot relation extraction with attribute representation learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3470–3479.

Yew Ken Chia, Lidong Bing, Soujanya Poria, and Luo Si. 2022. Relationprompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 45–57.

Shumin Deng, Yubo Ma, Ningyu Zhang, Yixin Cao, and Bryan Hooi. 2022. Information extraction in low-resource scenarios: Survey and perspective. *arXiv preprint arXiv:2202.08063*.

Jiaying Gong and Hoda Eldardiry. 2024. Prompt-based zero-shot relation extraction with semantic knowledge augmentation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13143–13156.

Xu Han, Tianyu Gao, Yankai Lin, Hao Peng, Yaoliang Yang, Chaojun Xiao, Zhiyuan Liu, Peng Li, Jie Zhou, and Maosong Sun. 2020. More data, more relations, more context and more openness: A review and outlook for relation extraction. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 745–758.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2023. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. In *The Eleventh International Conference on Learning Representations (ICLR 2023)*.

Bin Ji, Jie Yu, Shasha Li, Jun Ma, Qingbo Wu, Yusong Tan, and Huijun Liu. 2020. Span-based joint entity and relation extraction with attention-based span-specific and contextual semantic representations. In *Proceedings of the 28th international conference on computational linguistics*, pages 88–99.

Bosung Kim, Hayate Iso, Nikita Bhutani, Estevam Hruschka, Ndapandula Nakashole, and Tom Mitchell. 2023. Zero-shot triplet extraction by template infilling. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 272–284.

Yuquan Lan, Dongxu Li, Yunqi Zhang, Hui Zhao, and Gang Zhao. 2024. Rsed: Zero-shot relation triplet extraction via relation selection and entity boundary detection. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11256–11260. IEEE.

Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6877–6892.

Guozheng Li, Peng Wang, Jiajun Liu, Yikai Guo, Ke Ji, Ziyu Shang, and Zijie Xu. 2024a. Meta in-context learning makes large language models better zero and few-shot relation extractors. *arXiv preprint arXiv:2404.17807*.

Zehan Li, Fu Zhang, and Jingwei Cheng. 2024b. Alignre: An encoding and semantic alignment approach for zero-shot relation extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 2957–2966.

Bo Lv, Xin Liu, Shaojie Dai, Nayu Liu, Fan Yang, Ping Luo, and Yue Yu. 2023. Dsp: Discriminative soft prompts for zero-shot entity and relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5491–5505.

Jinzhong Ning, Zhihao Yang, Yuanyuan Sun, Zhizheng Wang, and Hongfei Lin. 2023. Od-rte: A one-stage object detection framework for relational triple extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11120–11135.

Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot relation classification as textual entailment. In *Proceedings of the first workshop on fact extraction and VERification (FEVER)*, pages 72–78.

Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero and few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1199–1212.

Yu-Ming Shang, Heyan Huang, and Xianling Mao. 2022. Onerel: Joint entity and relation extraction with one module in one step. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 11285–11293.

Yongliang Shen, Xinyin Ma, Yechun Tang, and Weiming Lu. 2021. A trigger-sense memory flow framework for joint entity and relation extraction. In *Proceedings of the web conference 2021*, pages 1704–1715.

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*.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. 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 5784–5789.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1706–1721.

Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. 2019. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–37.

Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, et al. 2023. Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1476–1488.

Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, and Enhong Chen. 2023. Large language models for generative information extraction: A survey. *arXiv preprint arXiv:2312.17617*.

Ting Xu, Haiqin Yang, Fei Zhao, Zhen Wu, and Xinyu Dai. 2024. A two-agent game for zero-shot relation triplet extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 7510–7527.

Junjie Yu, Xing Wang, and Wenliang Chen. 2024. Reliable data generation and selection for low-resource relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19440–19448.

Yue Yuan, Xiaofei Zhou, Shirui Pan, Qiannan Zhu, Zeliang Song, and Li Guo. 2021. A relation-specific attention network for joint entity and relation extraction. In *International joint conference on artificial intelligence*. International Joint Conference on Artificial Intelligence.

Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. 2024. Gliner: Generalist model for named entity recognition using bidirectional transformer. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5364–5376.

Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2003. Kernel methods for relation extraction. *Journal of machine learning research*, 3(Feb):1083–1106.

Daojian Zeng, Haoran Zhang, and Qianying Liu. 2020. Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 9507–9514.

Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 506–514.

Kai Zhang, Bernal Jiménez Gutiérrez, and Yu Su. 2023. Aligning instruction tasks unlocks large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 794–812.

Jun Zhao, Wenyu Zhan, Wayne Xin Zhao, Qi Zhang, Tao Gui, Zhongyu Wei, Junzhe Wang, Minlong Peng, and Mingming Sun. 2023. Re-matching: A fine-grained semantic matching method for zero-shot relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6680–6691.

Zexuan Zhong and Danqi Chen. 2021. A frustratingly easy approach for entity and relation extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 50–61.