AlignRE: An Encoding and Semantic Alignment Approach for Zero-Shot Relation Extraction

Zehan Li[1], **Fu Zhang**[1*], **Jingwei Cheng**[1]
[1]School of Computer Science and Engineering, Northeastern University, China
lizehan1999@163.com, {zhangfu,chengjingwei}@mail.neu.edu.cn

Abstract

Zero-shot Relation Extraction (ZSRE) aims to predict unseen relations between entity pairs from input sentences. Existing prototype-based ZSRE methods encode relation descriptions into prototype embeddings and predict by measuring the similarity between sentence embeddings and prototype embeddings. However, these methods often overlook abundant side information of relations and suffer from a significant encoding gap between prototypes and sentences, limiting performance. To this end, we propose a framework named **AlignRE**, based on two **Align**ment methods for ZS**RE**. Specifically, we present a novel perspective centered on *encoding schema alignment* to enhance prototype-based ZSRE methods. We utilize well-designed prompt-tuning to bridge the encoding gap. To improve prototype quality, we explore and leverage multiple side information and propose a *prototype* aggregation method based on *semantic alignment* to create comprehensive relation prototype representations. We conduct experiments on FewRel and Wiki-ZSL datasets and consistently outperform state-of-the-art methods. Moreover, our method exhibits substantially faster performance and reduces the need for extensive manual labor in prototype construction. Code is available at https://github.com/lizehan1999/AlignRE.

1 Introduction

Relation Extraction (RE) extracts structured facts from unstructured text, aiming to predict a relation between two entities in a sentence. Despite the impressive performance achieved by supervised RE (Zheng et al., 2021; Shang et al., 2022; Zhang et al., 2023b), in real scenarios, these methods have difficulty adapting to relations that are unseen during training (Han et al., 2021). Zero-shot Relation Extraction (ZSRE) aims to predict unseen relations by leveraging prior knowledge from seen relations.

*Corresponding author.

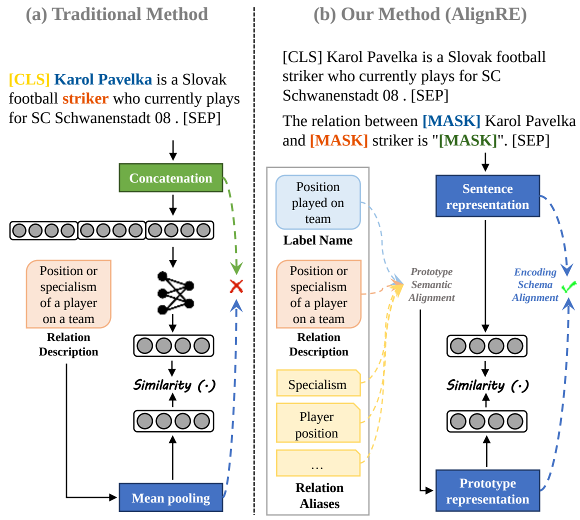

Figure 1: Comparison between traditional method (a) and our method (b). The colored words indicate the features used in constructing sentence embeddings.

In contrast to supervised RE, ZSRE does not necessitate an abundance of labeled instances, which possesses the capacity to classify unseen relations not present in the training data (Wang et al., 2019).

Presently, the main strategies for addressing ZSRE can be roughly classified into three types: prototype-based, classification-based, and generation-based methods. As shown in Figure 1(a), traditional prototype-based methods (e.g., ZS-BERT (Chen and Li, 2021)) obtain representations of different features of a sentence, concatenate them, and reduce dimensionality to encode sentence embeddings. Then, they match it with prototype embeddings constructed from descriptions of relations. However, *on one hand*, prototype embeddings are encoded by computing the mean pooling of all tokens (Reimers and Gurevych, 2019). When calculating the matching similarity between **sentence embeddings based on concatenation** and **prototype embeddings based on mean pooling**, differences in feature representation methods may lead to semantic *matching gaps* in feature

Findings of the Association for Computational Linguistics ACL 2024, pages 2957–2966
August 11-16, 2024 ©2024 Association for Computational Linguistics

space, possibly causing challenges in aligning embeddings and complicating model convergence. *On the other hand*, the previous prototype-based methods **mainly rely on the quality of relation description** information, and any noise in the description information has the potential to compromise the *faithfulness* of the prototype embeddings. *Other side information* related to the relations, such as label names and aliases of relations, may contribute to making prototype embeddings more robust.

In this work, we propose AlignRE, a simple yet effective ZSRE framework based on encoding schema alignment and prototype semantic alignment. *Firstly*, we propose a new perspective based on *encoding alignment* to eliminate the gap between sentence and prototype embeddings. Inspired by prompt-tuning (Han et al., 2022a; Liu et al., 2023), we devise a unified prompt containing [MASK] to guide the model to capture more accurate semantic information at the positions of [MASK] (Figure 1 (b)). On this basis, we obtain sentence representation using the same encoding schema as the prototype. *Secondly*, to enhance the quality of prototypes, we propose a prototype aggregation method based on *semantic alignment* to aggregate multiple types of side information, such as label names, descriptions, and aliases of relations. We assign higher attention to more representative prototypes and then aggregate them to obtain a comprehensive relation prototype representation. We finally match the sentence and prototype embeddings after aligning the encoding and semantics to complete prediction.

In summary, our contributions are as follows:

- We for the first time propose a new perspective based on encoding schema alignment to improve prototype-based ZSRE. We innovatively unified encoding schemas of relation prototypes and sentences through well-designed prompt-tuning to bridge the encoding gap.
- We explore and leverage side information besides relation descriptions to construct multiple prototypes for each relation and propose a prototype aggregation method based on semantic alignment, further refining features of relation prototypes.
- We conduct multiple experiments on FewRel and Wiki-ZSL datasets and report average results. AlignRE consistently outperforms state-of-the-art methods in terms of F1 (**+2.33** on FewRel and **+1.47** on Wiki-ZSL). Specifically, by utilizing alignment-based sentence embedding instead of the traditional methods, the F1 score of ZSRE can be significantly improved by approximately **33**. Moreover, AlignRE exhibits faster performance compared to classification-based and generation-based methods, reduces the need for manual labor in prompt and prototype construction.

2 Related Work

Zero-Shot Relation Extraction (ZSRE) was first introduced by Levy et al. (2017), and existing methods can be roughly categorized into three types.

Prototype-based Methods transform relation descriptions into prototype embeddings. They minimize the distance between sentence embeddings and prototype embeddings, employing nearest neighbor search techniques to predict unseen relations (Chen and Li, 2021). Zhao et al. (2023) manually designs templates for each relation, facilitating more fine-grained semantic matching between sentences and prototypes. **Classification-based Methods** (Obamuyide and Vlachos, 2018; Sainz et al., 2021; Liu et al., 2022) integrate classification networks into their frameworks. However, as the unseen relations increases, the computational cost of these methods also rises (Zhao et al., 2023). **Generation-based Methods** generate relation label directly for the input sentence (Lu et al., 2022; Chia et al., 2022). Recently, large generative language models (Huang et al., 2023; Ma et al., 2023) have injected new vitality into ZSRE tasks (Wei et al., 2023; Zhang et al., 2023a; Li et al., 2023), but they also come with elevated inference costs.

Moreover, the recent prompt-tuning techniques (Zhang et al., 2022) have been applied into the supervised RE task (Han et al., 2022b; Chen et al., 2022), achieving promising results. Recent work (Sainz et al., 2021; Chia et al., 2022) also effectively utilizes prompts in ZSRE. Different from these methods, our approach is based on the prototype framework and for the first time considers the alignment of encoding schemas of sentences and prototypes. To this end, we construct only a unified prompt template with [MASK] to accommodate all relations, avoiding the complexity of prompt design. Further, we innovatively propose a method utilizing a weighted pooling mechanism of [MASK] embeddings for generating sentence embeddings. We also propose a prototype aggre-

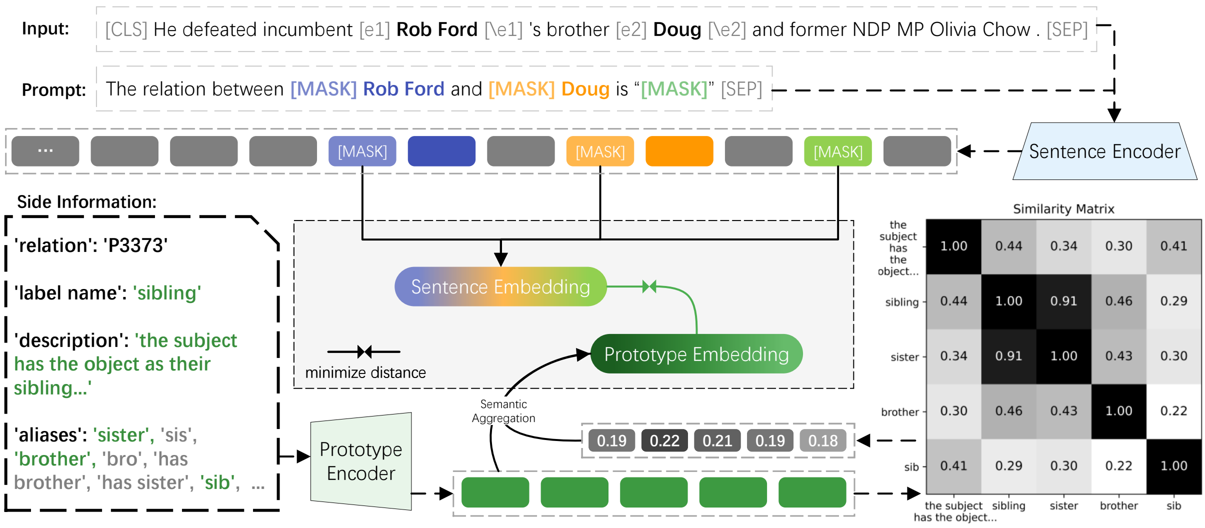

Figure 2: Framework of the proposed AlignRE.

gation method based on semantic alignment to enhance the quality of prototypes. Our method significantly reduces training and inference costs while also achieving SOTA results in multiple settings.

3 Methodology

3.1 Task Definition

Given an instance $I = (x, e^h, e^t, r)$, consisting of the input sentence x, the head entity e^h, the tail entity e^t, and their relation r. The training dataset S contains n seen relations $r_s \in R_s$ along with side information, which includes the label name, description, and alias of r_s. The objective of ZSRE is to learn from S and transfer knowledge to predict unseen relations $r_u \in R_u$ in the test dataset U. Prediction is solely based on the side information of r_u. There is no overlap between R_s and R_u.

3.2 Framework of AlignRE

As shown in Figure 2, AlignRE consists of two modules, **Encoding alignment-based Sentence Representation (EaSR)** and **Semantic alignment-based Relation Prototype Aggregation (SaPA)**.

In the **EaSR** module, we first feed the input and well-designed unified prompt with [MASK] for all relations into a sentence encoder. Then, we extract embeddings of [MASK] tokens corresponding to the head entity, tail entity, and relation, respectively. Unlike previous methods that directly concatenate all tokens to obtain sentence embeddings, we for the first time propose the idea of aligning sentence encoding with prototype encoding. To achieve this, we employ only the above three [MASK] tokens with a weighted pooling mechanism as the sentence representation. Such a seemingly simple method can eliminate the gap between encoding schemas of sentences and prototypes, and thus achieves significant performance improvements.

In the **SaPA** module, the label names, descriptions, and randomly selected aliases of relations in the side information are separately fed into a prototype encoder, resulting in several initial prototype embeddings. We propose a prototype aggregation method based on semantic alignment to aggregate them to obtain final prototype embeddings.

It's worth noting that during the training phase, we optimize only the sentence encoder, which means that prototype embeddings are fixed and can be computed in advance.

3.3 Encoding alignment-based Sentence Representation

Given an input sentence $X = \{w_1, w_2, \ldots, w_n\}$, following the conventional setup (Wu and He, 2019), we first introduce entity markers $[e]$ and $[\backslash e]$ into X, placing them on both sides of the head entity e^h and tail entity e^t to indicate their positions. The obtained X' can be represented as:

$$X' = \{w_1, w_2, \ldots, [e1], e^h, [\backslash e1], \ldots, [e2], e^t, [\backslash e2], \ldots, w_n\}. \quad (1)$$

As illustrated in Figure 2, we design a unified prompt template and transform the input sentence X into X_{prompt}. Considering that the type information of entities is crucial for relation extraction (Yao et al., 2019; Lyu and Chen, 2021), we introduce three [MASK] tokens into the prompt, where $[\text{MASK}]_r$ denotes the relation slot in the prompt template, and $[\text{MASK}]_h$ and $[\text{MASK}]_t$ represent the type information of the head entity and the tail entity, respectively. The representation of X_{prompt} can be expressed as:

$$\begin{aligned}X_{\text{prompt}} =& \{w_{p1}, w_{p2}, \ldots, [\text{MASK}]_h e^h, \ldots, \\ & [\text{MASK}]_t e^t, \ldots, [\text{MASK}]_r, \ldots, w_{pm}\},\end{aligned} \tag{2}$$

where $w_{p1}, w_{p2}, \ldots, w_{pm}$ represent tokens in our constructed prompt, e.g., the *prompt* in Figure 2. We will analyze the impact of different prompt templates on performance in Section 5.2 in detail.

We concatenate X' and X_{prompt}, using [SEP] as the separator token. As a result, the final input structure for prompt-tuning is as follows:

$$X'_{\text{prompt}} = [\text{CLS}]X'[\text{SEP}]X_{\text{prompt}}[\text{SEP}]. \tag{3}$$

Then we feed X'_{prompt} into the sentence encoder. Traditional prototype-based methods (Chen and Li, 2021; Zhao et al., 2023) obtain sentence embeddings by concatenating tokens and then reducing dimension. However, they calculate prototype embeddings of relations by pooling all tokens, there may be a semantic matching gap as mentioned in Section 1. Hence, we for the first time propose the idea of aligning sentence encoding schema with prototype encoding schema. We first directly extract $H_{[\text{MASK}]_h}$, $H_{[\text{MASK}]_t}$, $H_{[\text{MASK}]_r}$, denoting the embeddings associated with the head entity, the tail entity, and the relation, respectively. Then, we obtain the final sentence embedding through a weighted pooling mechanism, with hyperparameters α, β, γ governing their respective contributions. The *main motivation behind this* is that using the same encoding schemas can make prototype embeddings and sentence embeddings more easily matchable in the feature space. Experimental results indicate that this alignment approach significantly improves model's performance. The final embedding of input sentence X is formulated as:

$$H_X = \alpha * H_{[\text{MASK}]_h} + \beta * H_{[\text{MASK}]_t} + \gamma * H_{[\text{MASK}]_r}. \tag{4}$$

3.4 Semantic alignment-based Relation Prototype Aggregation

The traditional prototype-based methods mainly depend on the quality of relation description information. Noise in this information may undermine the accuracy of prototype embeddings. In order to enhance the quality of prototypes, we propose to leverage side information besides relation descriptions, and propose a prototype aggregation method based on semantic alignment. As shown in Figure 2, each relation has a unique label name, a description, and several aliases. Different types of side information have distinct characteristics. They delineate distinct embedding spaces by virtue of disparities in data distribution, and we furnish an exhaustive analysis of them in Section 5.4.

To benefit collectively from side information, we design a prototype aggregation method based on semantic alignment. We construct a weight matrix to amalgamate information of diverse types pertaining to the same relations.

Specifically, we employ a prototype encoder to embed the side information $side_i$ for each relation r_i, yielding a set of initial prototype embeddings $\{p_i^1, p_i^2, \ldots, p_i^n\}$. Then, we assign weights to each side information to characterize the importance of different side information.

As shown in Figure 2, the weight for the j-th side information can be expressed as follows:

$$W_j = \frac{\exp\left(\sum_{k=1}^{n} \text{sim}(p_i^k, p_i^j)\right)}{\sum_{l=1}^{n} \exp\left(\sum_{k=1}^{n} \text{sim}(p_i^k, p_i^l)\right)}. \tag{5}$$

We adopt cosine similarity as the measurement function $\text{sim}(\cdot)$, with n denoting the number of side information. The final prototype embedding of the i-th relation can be computed as:

$$P_i = \sum_{j=1}^{n} W_j * p_i^j. \tag{6}$$

3.5 Train and Test

Our training objective is to minimize the distance between sentence embeddings and prototype embeddings of positive relations, while maximizing the distance between sentence and prototype embeddings of negative relations. In order to enhance model robustness, we employ the margin loss as the objective function, which is formulated as follows:

$$\begin{aligned}L = \sum_{i=1}^{N} & \max(0, \max(\text{sim}(H_{X,i}, P_j)) \\ & - \text{sim}(H_{X,i}, P_i) + \Delta), i \neq j,\end{aligned} \tag{7}$$

where N represents the batch size.

The objective is to ensure that the distance between sentence embeddings and positive prototype embeddings is greater than the distance between

Unseen Labels	Model	Wiki-ZSL			FewRel		
		P.	R.	F1	P.	R.	F1
$m = 5$	R-BERT (Wu and He, 2019)[†]	39.22	43.27	41.15	42.19	48.61	45.17
	ZS-BERT (Chen and Li, 2021)[†]	71.54	72.39	71.96	76.96	78.86	77.90
	LaVeEntail (Sainz et al., 2021)*	77.39	75.90	76.63	91.14	90.86	91.00
	RelationPrompt (Chia et al., 2022)*	70.66	**83.75**	76.63	90.15	88.50	89.30
	RE-Matching (Zhao et al., 2023)*	78.19	78.41	78.30	92.82	92.34	92.58
	SUMASK (Li et al., 2023)[†]	75.64	70.96	73.23	78.27	72.55	75.30
	AlignRE(ours)	**83.11**	80.30	**81.64**	**93.30**	**92.90**	**93.09**
$m = 10$	R-BERT (Wu and He, 2019)[†]	26.18	29.69	27.82	25.52	33.02	25.52
	ZS-BERT (Chen and Li, 2021)[†]	60.51	60.98	60.74	56.92	57.59	57.25
	LaVeEntail (Sainz et al., 2021)*	71.86	71.14	71.50	83.05	82.55	82.80
	RelationPrompt (Chia et al., 2022)*	68.51	**74.76**	71.50	80.33	79.62	79.96
	RE-Matching (Zhao et al., 2023)*	74.39	73.54	73.96	83.21	82.64	82.93
	SUMASK (Li et al., 2023)[†]	62.31	61.08	61.69	64.77	60.94	62.80
	AlignRE(ours)	**75.00**	73.26	**74.10**	**86.41**	**85.14**	**85.75**
$m = 15$	R-BERT (Wu and He, 2019)[†]	17.31	18.82	18.03	16.95	19.37	18.08
	ZS-BERT (Chen and Li, 2021)[†]	34.12	34.38	34.25	35.54	38.19	36.82
	LaVeEntail (Sainz et al., 2021)*	62.13	61.76	61.95	72.83	72.10	72.46
	RelationPrompt (Chia et al., 2022)*	63.69	**67.93**	65.74	74.33	72.51	73.40
	RE-Matching (Zhao et al., 2023)*	67.31	67.33	67.32	73.80	73.52	73.66
	SUMASK (Li et al., 2023)[†]	43.55	40.27	41.85	44.76	41.13	42.87
	AlignRE(ours)	**69.01**	67.52	**68.26**	**77.63**	**77.00**	**77.31**

Table 1: Main results on Wiki-ZSL and FewRel datasets. We report the average results across five random seeds. The results indicated by † and $*$ are reported in SUMASK (Li et al., 2023) and RE-Matching (Zhao et al., 2023), respectively. Here, P., R. and F1 denote Precision, Recall and F1 score, respectively. Best results are in bold.

sentence embeddings and negative prototype embeddings, and this difference must surpass a predefined threshold denoted as Δ. We randomly selected 7 different P_j within the same batch as negative prototype embeddings.

4 Experiments

4.1 Setup

Dataset: We evaluate our model on two benchmark datasets: FewRel and Wiki-ZSL. **FewRel** (Gao et al., 2019) comprises a total of 56,000 instances, representing 64 relations from the training set and 16 relations from the validation set. **Wiki-ZSL** (Chen and Li, 2021) is generated through a distant supervision approach using the Wiki-KB (Sorokin and Gurevych, 2017), which comprises 113 relations and 94,383 instances. The description of relations can be obtained from (Chen and Li, 2021) or directly accessed from Wikidata. Similarly, the other side information, including label names and aliases of relations, can also be directly accessed from Wikidata.

Training and Evaluation: Following the standard experimental setup (Chen and Li, 2021), we utilize the union of their respective training and validation sets from FewRel and Wiki-ZSL, and randomly select m relations as unseen labels and partition the dataset into training and validation subsets. It is ensured that there exists no overlap between the sets of seen and unseen relations. Our experiments encompass the most widely used variations of m values, specifically, $m = 5, 10, 15$. We selected the label name, the description, and three random aliases of relations from the side information as the initial prototypes.

Implementation Details: We establish consistent hyperparameters for all experiments. To ensure the comparability of results, we conduct all our experiments across five different random data partition seeds, which remain consistent with those used by Zhao et al. (2023). We use *bert-base-uncased* (Kenton and Toutanova, 2019) as the sentence encoder and *stsb-bert-base* (Reimers and Gurevych, 2019) as the prototype encoder, both of which have 110 million parameters. We utilize the AdamW optimizer with a learning rate of $2e^{-6}$, Δ at 0.1, 5 epochs, and a batch size of 64. We test with different values of α, β, and γ, choosing α and

		Label Name	Relation Description	Relation Aliases	Semantic Aggregation
A1	concat($[e1]$, $[e2]$)	34.67	35.61	32.68	43.71
A2	concat([CLS],e^h, e^t)	36.23	36.89	33.14	52.80
A3	concat([MASK]$_{prompt}$,$[e1]$, $[e2]$)	39.30	38.52	28.87	50.38
A4	concat([MASK]$_{prompt}$,e^h, e^t)	44.19	32.86	32.66	42.74
A5	[MASK]$_{prompt}$	70.58	68.65	68.41	**77.31**

Table 2: Ablation experiments (F1) on the FewRel dataset with $m = 15$. **Semantic Aggregation** denotes the results of employing our semantic alignment-based prototype aggregation method on the label name, relation description, and three random relation aliases. The results of AlignRE are denoted in bold.

β from the set $\{0, 0.1, 0.2, 0.33, 0.4, 0.5\}$, and γ from the set $\{0, 0.2, 0.33, 0.6, 0.8, 1\}$. We conduct experiments on the FewRel dataset with $m = 15$ and select the hyperparameters $\alpha = \beta = \gamma = 0.33$ when the model performs the best. All experiments are conducted on a 24GB NVIDIA 3090Ti GPU.

Baseline: We compare three types of baseline models in this study. These types include *prototype-based methods*, such as ZS-BERT (Chen and Li, 2021) and the recent competitive RE-Matching (Zhao et al., 2023); *classification-based methods*, including R-BERT (Wu and He, 2019) and LaVeEntail (Sainz et al., 2021); and *generation-based methods* RelationPrompt (Chia et al., 2022) and SUMASK (Li et al., 2023). Our method is based on prototypes. Compared to Classification-based methods and Generation-based methods, our approach significantly reduces training and inference costs by directly matching embeddings. In contrast to other Prototype-based methods, we align the encoding method of sentence encoder with prototype encoder and reduce additional manual labor through prototype aggregation.

4.2 Main Results

All experimental results are presented in Table 1. Our model consistently outperforms all baselines in F1 on two datasets. Moreover, we draw several interesting conclusions:

(1) Compared to the classification-based method (LaVeEntail) and the generation-based method (RelationPrompt), AlignRE demonstrates significant performance advantages with fewer parameters. Even the method based on powerful large language models (SUMASK), it still cannot handle the ZSRE task well, showcasing the potential of the prototype-based framework.

(2) Compared to the other prototype-based method (ZS-BERT), with the expansion of the set of unseen relations, AlignRE maintains stable performance. Compared to the SOTA prototype-based method RE-Matching, AlignRE performs better on both datasets and achieves an average F1 improvement of **2.33** on the FewRel and **1.47** on Wiki-ZSL.

(3) Moreover, RE-Matching manually extracts entity type information from relation descriptions. LaVeEntail manually designs different prompt templates for each relation. Our results demonstrate that even in scenarios where no additional manual labor is required, our model still shows the superiority of the method.

	P.	R.	F1
[CLS]	73.83	72.73	73.27
MEAN(X')	74.79	72.95	73.85
MEAN(X'_{prompt})	74.78	73.13	73.94
MEAN([MASK])	77.63	77.00	**77.31**

Table 3: Comparison of selecting different tokens for generating sentence embeddings on FewRel ($m = 15$).

4.3 Ablation Experiments

We conduct ablation experiments under the setting of $m = 15$ on the FewRel dataset. The results are shown in Table 2. A1 and A2 are methods without prompt-tuning, also utilized in ZS-BERT (Chen and Li, 2021) and RE-Matching (Zhao et al., 2023) for building sentence representations. A3 and A4 are methods using [MASK]-based prompt-tuning while concatenating specific tokens. The last one [MASK]$_{prompt}$ is our method proposed in Section 3.3.

To investigate **the effectiveness of encoding schema alignment** for the ZSRE task, we extract feature embeddings such as [CLS], the head entity e^h, the tail entity e^t, entity markers' starting tokens $[e1]$, $[e2]$, and concatenate them to form different semantic representations. The results indicate that, when utilizing any of the same prototype information, the results obtained by our prompt-based sentence encoding method (A5) significantly outperform others (**an average F1 improvement of close to 33**).

	Prompt Template	P.	R.	F1
P1	[MASK] e^h [MASK] [MASK] e^t	76.78	75.38	76.07
P2	[MASK] e^h is [MASK] of [MASK] e^t	76.71	75.53	76.11
P3	The relation between e^h and e^t is " [MASK] "	77.42	76.35	76.87
P4	The relation between [MASK] and [MASK] is " [MASK] "	75.20	75.31	75.25
P5	The relation between [MASK] e^h and [MASK] e^t is " [MASK] " (**without** entity marker)	77.42	77.15	77.28
P6	The relation between [MASK] e^h and [MASK] e^t is " [MASK] "	77.63	77.00	**77.31**

Table 4: Comparison of different prompt templates on the FewRel dataset with $m = 15$.

We conduct an analysis to elucidate potential reasons for the observed results. **Prototype embeddings** are generated by computing the mean pooling of embeddings for each token in the prototype without fine-tuning. **Sentence embeddings** in the tradition models (ZS-BERT, RE-Matching) use concatenation by joining tokens (such as [CLS] or markers for the entities). Although the concatenation method is intuitive in traditional supervised RE (Wu and He, 2019; Soares et al., 2019), we argue that it is not suitable for ZSRE. The different ways of generating embeddings may prevent ZSRE models from fully leveraging the pre-training knowledge. This could make it more challenging for sentence embeddings to approach prototype embeddings, leading to a rapid decline in performance as the number of unseen relations increases. Our approach addresses this by *aligning the encoding schemas* of prototypes and sentences, making prototype embeddings and sentence embeddings more closely aligned and eliminating this gap. *This strategy thus yields superior results*, as shown in the comparison of A5 with A1-A4 in Table 2.

To investigate **the influence of semantically aggregated** prototype embeddings on the ZSRE task, we conduct experiments employing distinct prototypes. The results in Table 2 reveal a notable enhancement in model performance with **Semantic Aggregation**, which use the label name, description, and three random aliases of the relation.

It is noteworthy that, in contrast to description of relations, the encoding schema alignment approach appears to derive greater benefits from simplistic label names of relations. We posit that one explanation for this phenomenon lies in the fact that relation descriptions tend to contain a higher degree of noise, which may compromise the quality of prototypes and consequently lead to diminished discriminative power. Conversely, MLM (Masked Language Modeling) operates at the token-level rather than the sentence-level, utilizing relation names that share common words with relation prototypes, which inherently facilitates the matching process with the [MASK] token in the prompt.

5 Analysis

5.1 Effectiveness of Selected [MASK] Tokens

We change the schema of encoding sentence embeddings in ZSRE, advocating for aligning it with the encoding method of prototype embeddings (mean pooling). The ablation experiments in Section 4.3 have demonstrated the effectiveness of encoding schema alignment. When we encode sentence representation, we design a [MASK]-based prompt, and propose to employ only three [MASK] tokens with a weighted pooling mechanism as the sentence embedding. Therefore, to verify the effectiveness of our carefully selected tokens method, we compare the effects of using different tokens for mean pooling.

As shown in Table 3, [CLS] represents using only this token to denote sentence embedding. MEAN(X') represents the mean pooling of all tokens in the sentence. MEAN(X'_{prompt}) denotes the mean pooling of all tokens in the sentence and prompt. MEAN([MASK]) represent the method we used, using only the mean pooling of three [MASK] tokens. The results show that using only the [MASK] tokens in the prompt leads to a performance improvement of **3.37-4.04** compared to using the other tokens, demonstrating that *our* [MASK]*-based prompt-tuning* and *our carefully selected tokens* can capture more accurate semantic information about relations, which *aids in aligning and matching with relation prototypes*.

5.2 Impact of Prompt Templates

To analyze the impact of different prompt formats on the model's performance, we conduct experiments using various types of prompts. As indicated in Table 4, within the realm of semantically coherent expressions, different prompt formats exhibit no significant differences in effectiveness, affirming the *stability* of our method.

	Strategy	F1
S1	***name description aliases***	72.61
S2	***name*** (also known as ***aliases***) is ***description***	71.52
S3	avg.($p_{name}, p_{description}, p_{aliases}$)	76.88
S4	align.($p_{name}, p_{description}, p_{aliases}$)	**77.31**

Table 5: Comparison across different prototype aggregation methods on the FewRel dataset under the setting of $m = 15$. And p represents the initial prototype embedding.

Our findings underscore the necessity of including entity names within the prompt, as illustrated by P4 and P6. The results of P3 and P6 demonstrate that introducing [MASK] tokens of two entities can further enhance performance. Additionally, as shown by P5 and P6, the inclusion of entity markers in the input sentences yields marginal improvements in results.

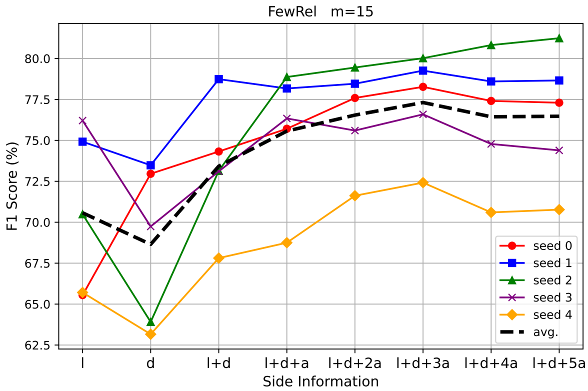

Figure 3: F1 scores for five random seeds under the FewRel dataset with m=15 setting. The dashed line represents the average of the five results. *l+d+3a* represents the result of semantic aggregation of the label name, description, and three aliases of relation (similar for others).

5.3 Analysis on Aggregation Strategy

To investigate the impact of different information aggregation methods on prototype quality, we conduct experiments involving four distinct aggregation strategies as shown in Table 5. Strategies S1 and S2 concatenate different side information with template to generate prototype embeddings through prototype encoder.

Strategy S3 aggregate initial prototype embeddings of different side information via mean pooling, while Strategy S4 represents our semantic alignment-based aggregation approach. The results indicate that embedding different side information separately and then aggregating them is better than directly concatenating them and then embedding. Moreover, the use of semantic alignment-based aggregation demonstrates a pronounced advantage.

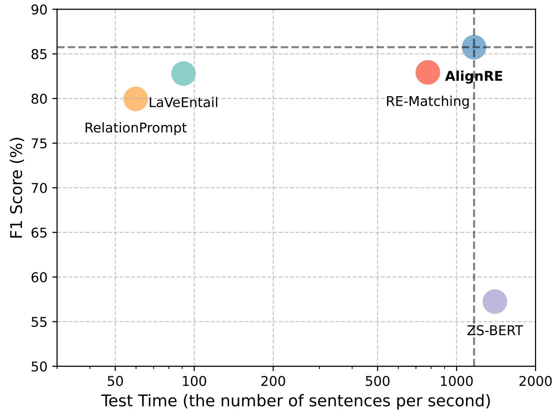

Figure 4: Analysis of model performance (F1) and inference time under the FewRel dataset with $m = 10$.

5.4 Analyzing the Contribution of Different Side Information to Prototype

In order to investigate the contributions of various types of information to the prototypes, we conduct separate experimental analyses for each.

Figure 3 illustrates the results of our experiments across five different seeds. In most of the seed configurations, relation description information does not confer an advantage in prototype generation. Conversely, more concise relation names are better suited to represent the essence of relations. Moreover, the results indicate that a greater variety and quantity of side information can lead to more stable and reliable prediction results. However, the degree of improvement tends to plateau as information accumulates. Therefore, we advocate for the creation of highly discriminative prototypes as a more cost-effective approach in ZSRE task.

5.5 Efficiency Comparison

In Figure 4, we compare the performance of different baselines in terms of test time with $m = 10$ on the FewRel dataset.

The results indicate that prototype-based methods (ZS-BERT, RE-Matching, and AlignRE), compared to the other classification-based (LaVeEntail) and generation-based (RelationPrompt) methods, use fewer parameters and simpler encoding processes, thus significantly improving inference speed. RE-Matching gain an advantage in both time and performance due to its fine-grained modeling, though it is slightly slower than ZS-BERT due to the increased number of matches. Although our method AlignRE incurs additional encoding costs for prompts, AlignRE does not add additional linear layers for feature dimension reduction, thus maintains high inference efficiency and performance.

6 Conclusion

In this paper, we propose a simple and effective ZSRE framework named AlignRE, based on encoding schema alignment and prototype semantic alignment.

We for the first time propose the idea of aligning encoding schemas of sentences and prototypes, achieving this through well-designed prompt-tuning and careful token selection, bridging the encoding gap and significantly enhancing the performance of ZSRE. We also leverage multiple side information to construct prototypes for relations, and propose a aggregation method based on semantic alignment to obtain a comprehensive relation prototype. Moreover, our method exhibits substantially faster performance and reduces the need for extensive manual labor in prototype construction.

In future work, our goal is to further optimize prototype representations and extend the task to settings involving a wider range of unseen relations.

Limitations

Due to noise in low-resource scenarios, different dataset partitions have an impact on the performance of AlignRE. We currently adopt the most widely used segmentation settings in existing work, namely $m = 5, 10, 15$. The performance of the model under a wider range of segmentation settings requires further exploration. Additionally, our used prototype information in the datasets is sourced from Wikidata, where varying degrees of noise may compromise prototype quality. How to effectively enhance the model robustness and the prototype quality are worthy of attention.

Acknowledgments

We thank all the reviewers for their efforts to make the paper comprehensive and solid. This work is supported by the National Natural Science Foundation of China (62276057) and in part sponsored by CAAI-MindSpore Open Fund, developed on OpenI Community.

References

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 (NAACL-HLT)*, pages 3470–3479.

Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *Proceedings of the ACM Web conference (WWW)*, pages 2778–2788.

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 ACL 2022*, pages 45–57.

Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019. Fewrel 2.0: Towards more challenging few-shot relation classification. 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 6250–6255.

Jiale Han, Bo Cheng, and Wei Lu. 2021. Exploring task difficulty for few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2605–2616.

Jiale Han, Shuai Zhao, Bo Cheng, Shengkun Ma, and Wei Lu. 2022a. Generative prompt tuning for relation classification. In *Findings of EMNLP 2022*, pages 3170–3185.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2022b. Ptr: Prompt tuning with rules for text classification. *AI Open*, 3:182–192.

Jiaxin Huang, Shixiang Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2023. Large language models can self-improve. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1051–1068.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. *CoNLL 2017*, page 333.

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

Fangchao Liu, Hongyu Lin, Xianpei Han, Boxi Cao, and Le Sun. 2022. Pre-training to match for unified low-shot relation extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 5785–5795.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.

Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma, and Muhao Chen. 2022. Summarization as indirect supervision for relation extraction. In *EMNLP 2022*, pages 6575–6594.

Shengfei Lyu and Huanhuan Chen. 2021. Relation classification with entity type restriction. In *Findings of ACL-IJCNLP 2021*, pages 390–395.

Yubo Ma, Yixin Cao, Yong Hong, and Aixin Sun. 2023. Large language model is not a good few-shot information extractor, but a good reranker for hard samples! In *Findings of EMNLP 2023*, pages 10572–10601.

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.

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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

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 (EMNLP)*, 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 (AAAI)*, pages 11285–11293.

Livio Baldini Soares, Nicholas Fitzgerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 2895–2905.

Daniil Sorokin and Iryna Gurevych. 2017. Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1784–1789.

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*.

Shanchan Wu and Yifan He. 2019. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 2361–2364.

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 (ACL)*, pages 764–777.

Haoxing Zhang, Xiaofeng Zhang, Haibo Huang, and Lei Yu. 2022. Prompt-based meta-learning for few-shot text classification. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1342–1357.

Kai Zhang, Bernal Jimenez Gutierrez, and Yu Su. 2023a. Aligning instruction tasks unlocks large language models as zero-shot relation extractors. In *Findings of ACL 2023*, pages 794–812.

Mi Zhang, Tieyun Qian, Ting Zhang, and Xin Miao. 2023b. Towards model robustness: Generating contextual counterfactuals for entities in relation extraction. In *Proceedings of the ACM Web Conference (WWW)*, pages 1832–1842.

Jun Zhao, WenYu Zhan, 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 (ACL)*, pages 6680–6691.

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. 2021. Prgc: Potential relation and global correspondence based joint relational triple extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 6225–6235.