CE-DA: Custom Embedding and Dynamic Aggregation for Zero-Shot Relation Extraction

Fu Zhang[†*], **He Liu**[†], **Zehan Li**, **Jingwei Cheng**
School of Computer Science and Engineering, Northeastern University, China
Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, China
{zhangfu,chengjingwei}@mail.neu.edu.cn,
liuhe2000@stumail.neu.edu.cn, lizehan1999@163.com

Abstract

Zero-shot Relation Extraction (ZSRE) aims to predict novel relations from sentences with given entity pairs, where the relations have not been encountered during training. Prototype-based methods, which achieve ZSRE by aligning the sentence representation and the relation prototype representation, have shown great potential. However, most existing works focus solely on improving the quality of prototype representations, neglecting sentence representations, and lacking interaction between different types of relation side information. In this paper, we propose a novel ZSRE framework named **CE-DA**, which includes two modules: **C**ustom **E**mbedding and **D**ynamic **A**ggregation. We employ a two-stage approach to obtain customized embeddings of sentences. In the first stage, we train a sentence encoder through unsupervised contrastive learning, and in the second stage, we highlight the potential relations between entities in sentences using carefully designed entity emphasis prompts to further enhance sentence representations. Additionally, our dynamic aggregation method assigns different weights to different types of relation side information through a learnable network to enhance the quality of relation prototype representations. In contrast to traditional methods that treat the importance of all side information equally, our dynamic aggregation method further strengthens the interaction between different types of relation side information. Our method demonstrates competitive performance across various metrics on two ZSRE datasets.[1]

1 Introduction

Relation Extraction (RE) is an important task of natural language processing (NLP), which aims to identify the relation between two target entities within a given sentence (Pawar et al., 2017). Most of the existing supervised RE methods (Wu and He, 2019; Yamada et al., 2020; Sui et al., 2023) require large amounts of human-labeled data. However, it is costly and challenging to obtain large-scale labeled data for new relations in real-world applications (Wang et al., 2019). Hence, to address this issue, Levy et al. (2017) introduce the Zero-Shot Relation Extraction (ZSRE), and it has recently garnered increased attention (Chen and Li, 2021).

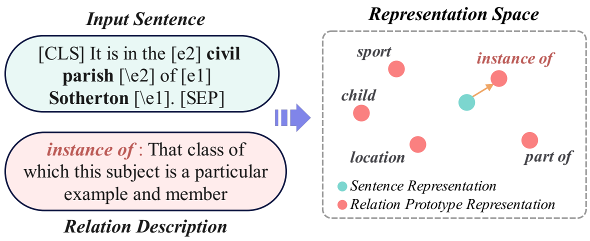

Figure 1: Example of a prototype-based ZSRE method. In the sentence, the head entity *Sotherton* and the tail entity *civil parish* are identified as having the most similar relation in the representation space as *instance of*.

As shown in Figure 1, the prototype-based method encodes the input sentence and the relation description separately, and *aligns* the *sentence representation* and the *relation prototype representation* corresponding to the relation description to predict unseen relations (Chen and Li, 2021), which has already demonstrated outstanding performance in ZSRE. Zhao et al. (2023) built upon this by aligning entities with hypernyms in relation descriptions, thereby achieving fine-grained semantic matching. However, this approach relies on manual annotation and still lacks interaction between sentences and relation descriptions (Li et al., 2024b). AlignRE (Li et al., 2024c) introduces *relation side information*, including label names, descriptions, and aliases, to enhance prototype representation by assigning them fixed weights. However, this static approach lacks interaction with the task objectives, which may lead to limitations in the model's generalization ability and accuracy. Recently, Large

[1]Code: https://github.com/ReveriePoem/CE-DA
[†]Equal contribution. [*]Corresponding author.

Proceedings of the 31st International Conference on Computational Linguistics, pages 9814–9823
January 19–24, 2025. ©2025 Association for Computational Linguistics

Language Models (LLMs) have also made significant progress in ZSRE (Zhang et al., 2023; Li et al., 2023). Nevertheless, the design of prompts in such methods requires extensive manual work.

Most previous prototype-based methods primarily focus on improving the quality of relation prototypes and do not emphasize enhancing sentence representations. They often directly use models like BERT (Devlin et al., 2019) to encode sentences. Considering that more accurate sentence representations can further enhance the model's ZSRE capability, inspired by Gao et al. (2021), which enhances sentence embeddings through contrastive learning, we introduce the idea of unsupervised contrastive learning and propose a two-stage training method to obtain better sentence representations. Additionally, traditional methods enhance relation prototype representations by incorporating relation side information, but when obtaining the relation prototype, they treat the importance of all side information equally. However, different side information often contributes unevenly to the prototype representation. Dynamically adjusting the contribution of side information may therefore help enhance the expressiveness of the prototype.

To address these issues, we propose a novel zero-shot relation extraction framework named **CE-DA**, which enhances model performance through two methods: **c**ustom sentence **e**mbedding and **d**ynamic relation **a**ggregation. Specifically, we introduce a two-stage training process to obtain Customized Embeddings of input sentences. For the first stage, we introduce noise into an input sentence by randomly *sampling dropout masks* based on contrastive learning to adapt to unseen noise. In the second stage, we introduce an *entity emphasis prompt* in the sentence to further enhance the interaction between the head and tail entities and their context. This staged optimization approach can adapt to the characteristics of ZSRE, enhancing the overall generalization capability of the model. In Dynamic Aggregation, we apply a new dynamic weighted aggregation method to integrate side information, aiming to learn the associations and importance among different types of relation information to obtain better prototype embeddings. Experimental results demonstrate that this method achieves competitive performance compared to state-of-the-art (SOTA) methods for ZSRE while eliminating the need for extensive manual work.

Our contributions are summarized as follows:

- We propose an innovative two-stage training framework to generate customized embeddings for sentences, thereby enhancing sentence representation. This framework improves the model's ability to handle noise in input sentences through unsupervised contrastive learning. Additionally, we introduce a simple entity emphasis prompting method, which further highlights the role of entities within sentences at the semantic level, to improve the quality of sentence representation.
- We propose a dynamic weighting method to aggregate relation side information to achieve comprehensive prototype representation.
- Experiments demonstrate that our method achieves new SOTA performance on the FewRel dataset and surpasses most baseline models on Wiki-ZSL, validating the effectiveness of our approach.

2 Related Work

Contrastive Learning. Contrastive learning is a dominant paradigm for representation learning, it has been widely applied in fields such as computer vision (Chen et al., 2020; He et al., 2020) and natural language processing (Giorgi et al., 2021; Rethmeier and Augenstein, 2023). Gao et al. (2021) proposed a simple contrastive learning framework for sentence embeddings. Their analysis shows that this method improves distribution uniformity and avoids degenerate alignment through dropout noise, enhancing expressiveness.

Zero-shot Relation Extraction. We follow Li et al. (2024c) and categorize existing ZSRE methods into three types. ***Prototype-based Methods*** minimize the distance between sentence embeddings and relation description embeddings, and predict unseen relations via nearest neighbor search (Chen and Li, 2021). Zhao et al. (2023) propose a fine-grained semantic matching method that decomposes the sentence-level similarity score into entity and context-matching scores. EMMA (Li et al., 2024b) is an efficient multi-grained matching method that combines coarse-grained recall with fine-grained classification. They generate virtual entity representations of descriptions in semantic matching instead of annotating descriptions to avoid manual costs. The coarse-grained filter selects candidate relations, while the fine-grained classifier refines selection for improved prediction

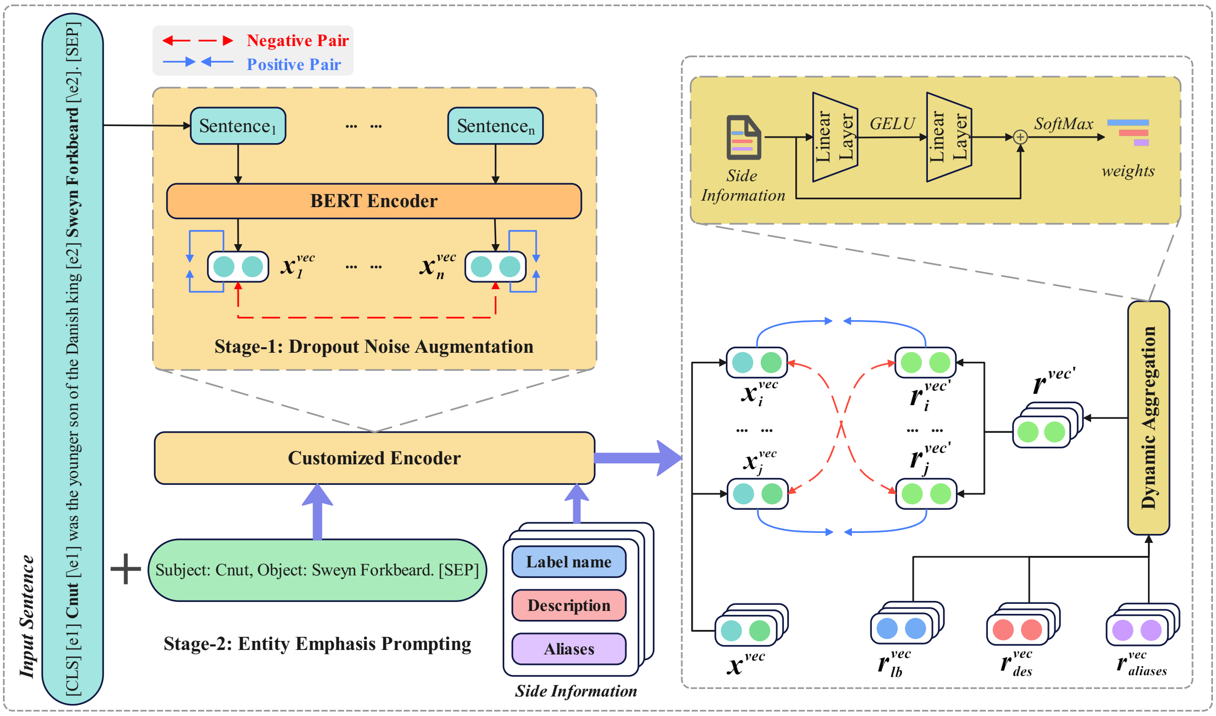

Figure 2: The overview of our CE-DA framework. It contains two modules: (1) The **Custom Embedding Module** which consists of two stages: the *dropout noise augmentation* and the *entity emphasis prompting*; (2) The **Dynamic Aggregation Module**, which is designed to effectively aggregate various side information of each relation.

accuracy. AlignRE (Li et al., 2024c) leverages side information besides relation descriptions to construct comprehensive relation prototypes, but its static weighting method may overlook the associations between different types of information. ***Classification-based Methods*** transform the relation extraction task into classification networks for other pivot tasks to unify low-shot relation extraction (Obamuyide and Vlachos, 2018; Sainz et al., 2021; Liu et al., 2022; Lv et al., 2023). ***Generation-based Methods*** prompt language models to generate auxiliary data, such as structured texts (Chia et al., 2022), to solve ZSRE. Recently, exploring the LLMs-based ZSRE method has gained significant attention (Wei et al., 2023; Zhou et al., 2024; Zhang et al., 2023). SUMASK (Li et al., 2023) recursively uses LLMs to convert RE inputs into an effective QA format. Li et al. (2024a) fine-tunes LLMs for zero and few-shot RE through a meta-training framework and directly generates relations using tabular prompting. Unlike previous prototype-based methods, we propose a two-stage training framework that applies unsupervised contrastive learning and entity emphasis prompting to enhance sentence representation quality. Additionally, we further improve the quality of relation prototype representations by applying dynamic aggregation on side information.

3 Method

3.1 Problem Formulation

Given seen dataset D_s and unseen dataset D_u. The seen relation set $R_s = \{r_s^1, ..., r_s^n\}$ and unseen relation set $R_u = \{r_u^1, ..., r_u^n\}$ are predefined, where $n = |R_s|$ and $m = |R_u|$ are the size of seen and unseen label sets respectively. These label sets are disjoint, i.e., $R_s \cap R_u = \emptyset$, and only the samples from D_s are available during training.

In zero-shot relation extraction, our goal is to learn from D_s and generalize to D_u. During training, for N examples $D_s = \{(x_i, e_h^i, e_t^i, r_s^j)|i = 1, ..., N\}$, where x_i is input sentence, e_h^i is the head entity, e_t^i is the tail entity, and $r_s^j \in R_s$ is corresponding relation. We train a model M to predict relations: $M(D_s^i) \rightarrow r_s^j \in R_s$. During testing, given an input sentence x and entity pair $e_h, e_t \in D_u$, using the model M to predict an unseen relation $r_u \in R_u$ between e_h and e_t: $M(D_u^i) \rightarrow r_u^j \in R_u$.

3.2 Overview

An overview of our proposed CE-DA is shown in Figure 2, which consists of two modules: (1) The **Custom Embedding Module** includes a two-stage training framework to obtain sentence representation. In the first stage, we use sentences from training dataset D_s as its training data, through unsupervised training with *Dropout Noise Augmentation*, a customized encoder M_E for the ZSRE task is obtained. In the second stage, we introduce the *Entity Emphasis Prompting*, where the head and tail entities in the input sentences are specifically emphasized during the ZSRE task, enhancing the interaction between the entities and the sentence context. (2) The **Dynamic Aggregation Module** also starts by using the customized encoder M_E to generate representations of relation side information (including the label name, description, and aliases of the relation). Then, we dynamically assign different weights to each type of side information, producing a comprehensive relation representation that integrates the side information. Finally, we compute the similarity between the enhanced sentence representation and the aggregated prototype representation and select the highest as the relation prediction result.

3.3 Custom Embedding Module

The Custom Embedding Module consisting of two stages aims to obtain high-quality representations of the input sentences. We introduce *dropout noise augmentation* in the first stage to train the encoder model M_E, and *entity emphasis prompting* in the second stage to highlight the role of entities within sentences at the semantic level to further improve the quality of sentence representation.

Sentence Encoder. We utilize the BERT-base (Devlin et al., 2019) as the pre-trained encoder to generate the representation of each sentence. Given a sentence $x_s = \{w_1^x, ..., w_n^x\} \in D_s$ and input it into the pre-trained encoder. We use the last hidden states of special token [CLS] as context representation, which is formulated as follows:

$$h_0^x, h_1^x, \ldots, h_n^x = \mathrm{BERT}(w_0^x, w_1^x, \ldots, w_n^x), \quad (1)$$

$$x^{vec} = h_0^x, \quad (2)$$

where $x^{vec} \in \mathbb{R}^d$, d is the hidden dimension.

Dropout Noise Augmentation. Given a collection of sentences $\{x_i\}_i^n \in D_s$, we generate positive pairs by randomly sampling dropout masks in the pre-trained language model, thereby introducing noise into the input sentences through dropout. We denote $h_i^z = f_\theta(x_i, z)$, where θ is the learnable parameter, and z is a random mask for dropout.

During the training phase, the same sample is input twice into the same encoder, each time with a different dropout mask (denoted as z and z'), resulting in two distinct representation embeddings $\mathbf{h}_i^{z_i}$ and $\mathbf{h}_i^{z'_i}$ as positive pairs. In-batch negatives are used by randomly sampling another input from the batch as negative pairs.

The training objective function of M_E is:

$$\mathcal{L}_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_i^{z'_i})/\tau}}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i^{z_i}, \mathbf{h}_j^{z'_j})/\tau}}, \quad (3)$$

where τ is a temperature hyperparameter, N is the number of input sentences, $i \in [1, N]$, and $\mathrm{sim}(\mathbf{h}_1, \mathbf{h}_2) = \frac{\mathbf{h}_1 \mathbf{h}_2}{\|\mathbf{h}_1\| \cdot \|\mathbf{h}_2\|}$ is the cosine similarity.

The optimization objective is that embeddings for positive sentences should stay close, and embeddings for random sentences should distributed as uniformly as possible. Note that the positive pair takes the same sentence, and their embeddings only differ in dropout masks. Therefore, by generating positive samples through this method of changing the dropout mask, the model can better handle various noises, thereby exhibiting stronger generalization ability and obtaining more distinguishable representations of input sentences.

Entity Emphasis Prompting. To emphasize the roles of head and tail entities in the sentence and enhance their interaction with the context, we introduce the *entity emphasis prompting* method, which explicitly highlights the subject and object in the sentence.

Specifically, we append the prompt "`Subject: [head entity], Object: [tail entity]`" at the end of the input sentence, which allows the model to focus more on the entities during the embedding process, resulting in embeddings that better capture the potential relation between the subject and object in the sentence.

3.4 Dynamic Aggregation Module

The Dynamic Aggregation Module is designed to obtain relation prototype representation by effectively integrating various side information of a relation, including the label name, description, and aliases. By calculating the importance score for each type of information corresponding to a relation, these feature embeddings are dynamically weighted and aggregated.

First, we aggregate the feature embeddings of the three types of information to form the relation feature matrix r^{vec}:

$$r^{vec} = \{r^{vec}_{lb}, r^{vec}_{des}, \frac{1}{n}\sum_{j=1}^{n} r^{vec}_{alias_j}\} \tag{4}$$

where r^{vec}_{lb} and r^{vec}_{des} represent the feature embeddings for the label name and description respectively, while $r^{vec}_{alias_j}$ denotes the embedding representation of the j-th alias, n is the number of aliases contained in the dataset.

To calculate the importance score α_i for each embedding, we design a two-layer feedforward neural network to compute α_i:

$$\alpha_i = W_2 \cdot \text{GELU}(W_1 r_i + b_1) + b_2, \tag{5}$$

where W_1 and W_2 are learnable parameters, b_1 and b_2 are biases, $r_i \in r^{vec}$.

Additionally, we add a bias term $Sum(\cdot)$, which represents the sum of all elements for r_i. By adding this to the importance score α_i, the model can also focus on the original semantics of each side information embedding, rather than solely relying on the results of a linear transformation. Then we normalize the final importance scores $\tilde{\alpha}_i$ using the SoftMax function to obtain the final weights for each embedding, $\tilde{\alpha}_i$ are calculated as follows:

$$\tilde{\alpha}_i = \alpha_i + Sum(r_i). \tag{6}$$

Finally, we use the importance weights to perform a weighted sum of the three types of side information embeddings, generating the aggregated relation prototype representation:

$$r^{vec'} = \sum_{j=1}^{3} \tilde{\alpha}_i r_i. \tag{7}$$

3.5 Train and Test

We utilize the model M_E as the input sentence encoder to obtain the sentence representation x^{vec} and the aggregated relation representation $r^{vec'}$. The training objective is to minimize the distance between input sentence embedding $x^{vec}_{s_i}$ and the corresponding aggregated embedding $r^{vec'}_{s_i}$ and maximize the distance from $r^{vec'}_{s_j}(j \neq i)$.

$$\mathcal{L}'_i = -\log \frac{e^{\text{sim}(x^{vec}_{s_i}, r^{vec'}_{s_i})/\tau}}{\sum_{j=1}^{N} e^{\text{sim}(x^{vec}_{s_i}, r^{vec'}_{s_j})/\tau}}, \tag{8}$$

where τ is a temperature hyperparameter, N is the number of samples and $i \in [1, N]$.

During the testing phase, we use the embedding M_E to embed the input sentence x_u and the unseen relation r_u, obtaining the enhanced sentence representation x^{vec}_u and the aggregated prototype representation $r^{vec'}_u$. The prediction of unseen relations is performed using the nearest neighbor search:

$$P(x_{u_i}) = \arg\max \text{sim}(x^{vec}_{u_i}, r^{vec'}_{u_j}). \tag{9}$$

The function P returns the predicted relation of a new input sentence x_{u_i}.

4 Experiments Setup

4.1 Datasets

We conduct experiments on two widely used zero-shot relation extraction datasets. FewRel (Han et al., 2018), collected from Wikipedia and subsequently hand-annotated by crowd workers, comprises 56,000 sentences across 80 distinct relation types, with each type containing 700 sentences. WikiZSL (Chen and Li, 2021) is derived from Wiki-KB (Sorokin and Gurevych, 2017) through distant supervision and contains rich textual information along with category labels. The dataset consists of 93,383 sentences covering 113 relation types.

We split the whole dataset into training, validation, and testing data. We randomly select 5 relations for the validation set, $m \in \{5, 10, 15\}$ novel relations as the unseen relations for the testing set, the remaining ones are considered as the seen relations for training. To ensure the reliability of the experiment results, we chose five random seeds for dataset partition that remain consistent with Zhao et al. (2023) and report the average results across different selections.

4.2 Baselines

We compare our model with three categories of baseline methods. For **prototype-based** methods, we select ZS-BERT (Chen and Li, 2021) and RE-Matching (Zhao et al., 2023), AlignRE (Li et al., 2024c), and the recent competitive method EMMA (Li et al., 2024b). For **classification-based** methods, we choose PromptMatch (Sainz et al., 2021), while for **generation-based** methods, we select REPrompt (Chia et al., 2022) and LLMs-based methods SUMASK (Li et al., 2023) and MICRE (Li et al., 2024a).

4.3 Implementation Details

We utilize the BERT-base as the pre-trained encoder in the first stage of the Custom Embedding

Unseen Labels	Method	Wiki-ZSL			FewRel		
		Prec.	*Rec.*	F_1	*Prec.*	*Rec.*	F_1
$m = 5$	ZS-BERT (Chen and Li, 2021)	71.54	72.39	71.96	76.96	78.86	77.90
	PromptMatch (Sainz et al., 2021)	77.39	75.90	76.63	91.14	90.86	91.00
	REPrompt (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
	MICRE (Li et al., 2024a)	76.46	78.53	77.48	89.34	91.88	90.59
	AlignRE (Li et al., 2024c)	83.11	80.30	81.64	93.30	92.90	93.09
	EMMA (Li et al., 2024b)	**91.32**	**90.65**	**90.98**	94.87	94.48	94.67
	CE-DA	88.01	87.02	87.51	**95.26**	**95.08**	**95.17**
$m = 10$	ZS-BERT (Chen and Li, 2021)	60.51	60.98	60.74	56.92	57.59	57.25
	PromptMatch (Sainz et al., 2021)	71.86	71.14	71.50	83.05	82.55	82.80
	REPrompt (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
	MICRE (Li et al., 2024a)	72.36	74.88	73.60	80.67	82.31	81.48
	AlignRE (Li et al., 2024c)	75.00	73.26	74.10	86.41	85.14	85.75
	EMMA (Li et al., 2024b)	**86.00**	**84.55**	**85.27**	87.97	86.48	87.22
	CE-DA	82.54	81.82	82.16	**88.61**	**87.60**	**88.10**
$m = 15$	ZS-BERT (Chen and Li, 2021)	34.12	34.38	34.25	35.54	38.19	36.82
	PromptMatch (Sainz et al., 2021)	62.13	61.76	61.95	72.83	72.10	72.46
	REPrompt (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	65.74	73.80	73.52	73.66
	SUMASK (Li et al., 2023)	43.55	40.27	41.85	44.76	41.13	42.87
	MICRE (Li et al., 2024a)	67.14	68.87	67.99	73.74	75.83	74.77
	AlignRE (Li et al., 2024c)	69.01	67.52	68.26	77.63	77.00	77.31
	EMMA (Li et al., 2024b)	**78.51**	**77.63**	**78.07**	80.47	79.73	80.10
	CE-DA	75.33	74.94	75.13	**84.03**	**82.60**	**83.31**

Table 1: Main results on two zero-shot relation extraction datasets. **Bold** marks the highest score, underline marks the second-best score. All baseline results are sourced from the original papers.

Module, with a dropout rate of 0.2 and a batch size of 64. In the second stage, we employ the customized embedding model obtained from the first stage as the encoder for the input sentences. We set AdamW (Loshchilov, 2017) as the optimizer with an initial learning rate of 2e-5 and a batch size of 32. The model is trained for 5 epochs with a warm-up of 100 steps. The temperature τ for the loss $\mathcal{L}_i$ in the first stage and $\mathcal{L}_i^{'}$ in the second stage is set to 0.02.

5 Results and Analysis

5.1 Main Results

We report the main results on FewRel and WikiZSL datasets, where three evaluation metrics, *Prec.*, *Rec.*, and F_1, denote Precision, Recall, and F1, respectively. The results are presented in Table 1, which shows that our method outperforms previous SOTA methods in terms of F1 on the FewRel dataset and achieves comparable results to the recent SOTA method on the Wiki-ZSL dataset when targeting different numbers of unseen relations.

Moreover, the superiority of CE-DA becomes more pronounced when m = 15. CE-DA achieves an F1 improvement of 3.21 on the FewRel when m = 15 compared to EMMA. Such results indicate that the improvement of CE-DA grows as m increases. Prediction becomes more challenging as the number of unseen relations increases, leading to the overlap and interaction between various relations. This suggests that our method effectively captures the semantic information of both the sentence and the unseen relations, demonstrating that our model can generate higher-quality sentence and relation prototype representations.

Additionally, our method performs well without relying on external resources. Compared to methods that depend on manually crafted relation descriptions (Zhao et al., 2023) or prompt templates (Zhang et al., 2023), CE-DA significantly reduces

Dataset	Method	*Prec.*	*Rec.*	F_1
FewRel	w/o DNA	82.93	82.27	82.60
	w/o EEP	82.87	81.23	82.04
	w/o DA	82.07	80.14	81.09
	Ours	**84.03**	**82.60**	**83.31**

Table 2: Ablation study on FewRel ($m = 15$). **DNA** and **EEP** denote the Dropout Noise Augmentation in the first stage and the Entity Emphasis Prompting in the second stage of the Custom Embedding Module, respectively, while **DA** denotes the Dynamic Aggregation Module.

labor costs while maintaining high performance.

5.2 Ablation Study

We report the impact of different components on CE-DA with the setting of m = 15 on the FewRel dataset, and the results are shown in Table 2.

We use BERT-base to encode the input sentences, and when the Dropout Noise Augmentation (DNA) is removed, it leads to a decline in model performance. Our analysis suggests that this is primarily due to DNA enhancing sentence representation by applying random dropout masks, allowing the model to learn diverse information from different contexts of the same sentence. This enables the model to better handle noise and variations across different contexts.

The Entity Emphasis Prompting (EEP) helps the model clearly distinguish the subject (head entity) and object (tail entity) by explicitly marking them in the sentence. Experimental results show that, after removing it, the model loses the emphasis on the interaction information between entities, making the generated embeddings less effective at capturing the interaction between entities, which subsequently impacts the model's prediction accuracy.

The core of the Dynamic Aggregation (DA) module lies in dynamically weighted different types of information such as relation labels, descriptions, and aliases. This module generates more expressive relation representations by learning the connections between different types of side information. When this module is removed (w/o DA), the model cannot fully utilize various relation information, leading to a decline in the quality of relation representations, which in turn affects the accuracy of relation prediction.

5.3 Analysis of Training Methods

As outlined in Section 3.3, the two-stage training process in our CE-DA framework involves training the customized embedding model through Dropout Noise Augmentation in the first stage. In the second stage, this model is employed as a pre-trained encoder to train for the ZSRE task with Entity Emphasis Prompt.

Training method	*Prec.*	*Rec.*	F_1
joint training	83.58	82.32	82.94
two-stage training	**84.03**	**82.60**	**83.31**

Table 3: Comparison of *joint training* and *two-Stage training* methods on FewRel ($m = 15$).

Instead of using our two-stage training method, when adopting the joint training process, that is, we simultaneously train the Dropout Noise Augmentation stage and the ZSRE task by introducing an Entity Emphasis Prompt into the sentence within a unified framework, with the loss from the first stage being backpropagated to the ZSRE task training stage.

As shown in Table 3, it can be observed that the two-stage training performs better than joint training. Our two-stage framework focuses on optimizing specific objectives. In the first stage, the goal of Dropout Noise Augmentation is to improve the quality of sentence representations while enabling the model to handle the sentence with noise more effectively. In the second stage, we introduce the Entity Emphasis Prompting method to enhance the interaction between entities and their context in the sentence. This two-stage training framework avoids potential conflicts from training multiple tasks simultaneously and improves the robustness.

5.4 The Impact of Entity Emphasis Prompt

To analyze the influence of various prompts on the model's performance, we conduct experiments using different prompts. As shown in the Table 4, directly specifying "Subject" and "Object" in the sentence structure, as in P3, proved more effective than using parentheses, as in P1. P2 incorporates the entities into the sentence and highlights their relation, but the results were unsatisfactory. This may be due to the added verbosity, which interfered with the model's ability to identify the relation.

The P3 template demonstrates strong performance across various metrics. This suggests that appending the subject and object as independent prompts at the end of a sentence enhances the model's ability to understand entity relations within the sentence. Meanwhile, this appending method

Entity Emphasis Prompt	*Prec.*	*Rec.*	F_1
P1 ... e^h (Subject) ... e^t (Object) ...	82.10	81.52	81.80
P2 ... e^h ... e^t ... e^h and e^t have a "relation" in this sentence	81.71	81.54	81.63
P3 ... e^h ... e^t ... Subject: e^h, Object: e^t	**84.03**	**82.60**	**83.31**

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

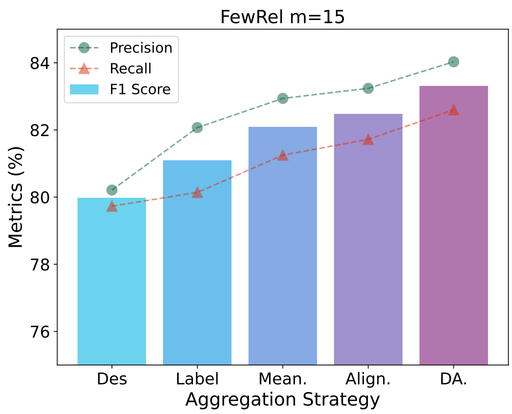

Figure 3: Effects on varying the aggregation strategies with $m = 15$ on the FewRel dataset.

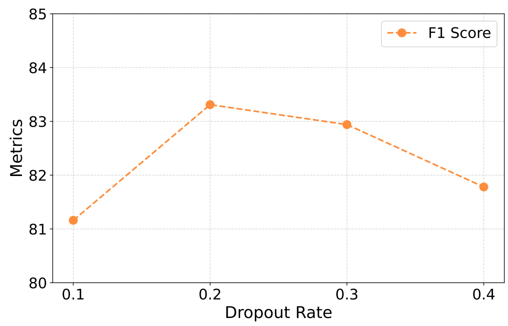

Figure 4: Analysis of dropout rates in the first stage of Custom Embedding Module ($m = 15$).

does not disrupt the original syntax and structure of the sentence, whereas the parentheses in P1 might break the original sentence's grammatical structure. This is particularly relevant for language models, as sentence structure and grammar are important features. If the sentence's natural fluency is affected, the model's performance may also decline accordingly. This suggests that prompts should be designed without disrupting the syntactic structure.

5.5 Aggregation Strategy Analysis

As shown in Figure 3, we analyze the impact of various relation side information aggregation methods on representation quality. Des Strategy and Label Strategy represent only matching the description and the relation label, respectively. Mean Strategy represents the aggregation of embeddings from different side information via mean pooling, Align Strategy represents the semantic alignment-based aggregation approach, which calculates the weight of each side information based on similarity (Li et al., 2024c), while our Dynamic Aggregation (DA) method dynamically assigns importance weights to different side information based on the training objective, further enhancing the interaction between the sentence and relation side information.

The results indicate that the aggregation methods perform better than individual strategies, and our DA Strategy demonstrates a pronounced advantage compared to the Align Strategy.

5.6 Performance on Different Dropout Rates

The dropout rate in Dropout Noise Augmentation of the Custom Embedding Module is an important hyper-parameter in our framework. We conduct experiments on the FewRel dataset to study the impact of the dropout rate on the final performance of the model. Figure 4 illustrates the results of our experiments across different dropout rates on the FewRel dataset under m=15.

When the dropout rate is set to 0.2, the model achieves the best performance. As the dropout rate increases, the model's performance tends to decline, which may be due to excessive dropout causing information loss, and the enhanced samples differ significantly from the original samples. Conversely, when the dropout rate is set to 0.1, the noise added is too small, making the differences between samples less distinct, and the model may struggle to effectively distinguish particularly similar samples, thus affecting its performance.

6 Conclusion

In this paper, we present a novel ZSRE framework CE-DA, which primarily consists of two modules: Custom Embedding and Dynamic Aggregation. To enhance sentence representations, we propose a new two-stage training framework to generate customized embeddings for sentences. We train a custom embedding model by Dropout Noise Augmentation in the first stage and introduce the Entity Emphasis Prompting method in the second stage. We also propose Dynamic Aggregation, a module that assigns different weights to different types of side information to enhance prototype representation. Extended experiments demonstrate the effectiveness of our method on ZSRE.

Limitations

Although our framework has achieved excellent performance in the ZSRE task by simultaneously improving the quality of sentence representation and prototype representation, there are still some limitations and areas for improvement.

Firstly, our method only outperforms the recent SOTA approach on FewRel, its performance on WikiZSL still slightly lags behind the SOTA approach, indicating that our method still requires improvement when handling rare data categories in large-scale datasets. Moreover, in terms of training overhead, we need to first train a custom embedding model, and this two-stage training process incurs additional computational cost than previous methods. Furthermore, the aggregation method is limited by existing data sources, and noise in the data may affect the quality of relation prototypes. We believe that optimizing the representation of relation prototypes through richer data will yield more benefits.

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

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.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

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.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910.

John Giorgi, Osvald Nitski, Bo Wang, and Gary Bader. 2021. Declutr: Deep contrastive learning for unsupervised textual representations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 879–895.

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.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 9726–9735. IEEE Computer Society.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342.

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*.

Shilong Li, Ge Bai, Zhang Zhang, Ying Liu, Chenji Lu, Daichi Guo, Ruifang Liu, and Sun Yong. 2024b. Fusion makes perfection: An efficient multi-grained matching approach for zero-shot relation extraction. In *Proceedings of the 2024 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), pages 79–85.

Zehan Li, Fu Zhang, and Jingwei Cheng. 2024c. 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.

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 (Volume 1: Long Papers)*, pages 5785–5795.

I Loshchilov. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

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.

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.

Sachin Pawar, Girish K Palshikar, and Pushpak Bhattacharyya. 2017. Relation extraction: A survey. *arXiv preprint arXiv:1712.05191*.

Nils Rethmeier and Isabelle Augenstein. 2023. A primer on contrastive pretraining in language processing: Methods, lessons learned, and perspectives. *ACM Computing Surveys*, 55(10):1–17.

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.

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*, pages 1784–1789.

Dianbo Sui, Xiangrong Zeng, Yubo Chen, Kang Liu, and Jun Zhao. 2023. Joint entity and relation extraction with set prediction networks. *IEEE Transactions on Neural Networks and Learning Systems*.

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*, pages 2361–2364.

Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. Luke: Deep contextualized entity representations with entity-aware self-attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6442–6454.

Kai Zhang, Bernal Jimenez Gutierrez, and Yu Su. 2023. Aligning instruction tasks unlocks large language models as zero-shot relation extractors. In *ACL*.

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.

Sizhe Zhou, Yu Meng, Bowen Jin, and Jiawei Han. 2024. Grasping the essentials: Tailoring large language models for zero-shot relation extraction. *arXiv preprint arXiv:2402.11142*.