Leveraging Large Language Models and Cross-Attention Mechanism for Zero-Shot Relation Extraction with Contrastive Learning

Anonymous EMNLP submission

Abstract

In the zero-shot relation extraction (ZSRE) task, large language models (LLMs) have shown remarkable capabilities in predicting unknown relations, offering significant improvements in efficiency and flexibility over traditional methods. However, the probabilistic nature of the generation process in LLMs may lead to the occurrence of hallucinations, causing inaccurate relation triples be generated. To relieve this problem, this paper proposes a novel model, Cross-Attention Contrastive Relation Extraction (CACRE), which aims at detecting erroneous relation triples generated by LLMs and then effectively distinguishing valid ones. The CACRE model leverages contrastive learning and cross-attention mechanisms. Specifically, contrastive learning is applied to distinguish between positive and negative relation triples, enhancing the model's feature extraction capability by learning discriminative features. Subsequently, a cross-attention mechanism is employed to capture the semantic associations between texts and triples, thereby improving the model's ability to understand and extract information from the input content. Experimental results on the DuIE2.0 Chinese dataset demonstrate that CACRE significantly outperforms baseline models in zero-shot scenario with an average 12% improvement in precision.

1 Introduction

The objective of zero-shot relation extraction (ZSRE) is to automatically identify and extract relations between entities from text, neither relying on predefined relation labels nor domain-specific annotated data. Traditional relation extraction (RE) approaches depend heavily not only on given explicit relation labels (Miwa and Bansal, 2016; Han et al., 2020), but also on substantial annotated data, which restricts their ability to generalize across unknown relation types. Recently, the rapid advancements in large language models (LLMs) (Bubeck et al., 2023) such as GPT (Radford et al., 2018), Qwen (Bai et al., 2023) and DeepSeek (Bi et al., 2024) have propelled the development of ZSRE tasks. By leveraging their exceptional reasoning capabilities, LLMs can infer relations (Tang et al., 2023) for previously unseen types without the need for additional task-specific training. This capability significantly reduces the dependence on annotated datasets and highlights the strong potential of LLMs for ZSRE tasks (Wei et al., 2023), presenting a novel approach for advancing RE research.

However, despite their strong generalization capabilities, LLMs face obvious limitations in extracting relations, particularly when processing texts with complex relation descriptions or semantic ambiguities, which can easily cause hallucinations (Adlakha et al., 2024; Lin et al., 2024; Zhou et al., 2024). The incorrect outputs, which include content that is either irrelevant to the input text or factually inaccurate (Li et al., 2023), often closely similar with correct data in syntactic structure, making them extremely hard to distinguish, thereby affecting the reliability of ZSRE. To address this issue, this paper proposes a Cross-Attention Contrastive Relation Extraction (CACRE) model, which achieves semantic alignment between texts and triples through a cross-attention mechanism, and introduces contrastive learning to further enhance the model's ability to distinguish correct from incorrect relational triples. This model can effectively filter out incorrect triples generated by LLMs, improving the overall accuracy of ZSRE.

The proposed CACRE model involves four main steps. As shown in Figure 1, first, relation triples are generated by designing specific prompts to guide LLMs in extracting triples from textual datasets, which are categorized into anchors, positives, and negatives to form structured training data. The negatives are required to have a high similarity to the anchors or positives. Second, a cross-attention mechanism (Niu et al., 2021) is ap-

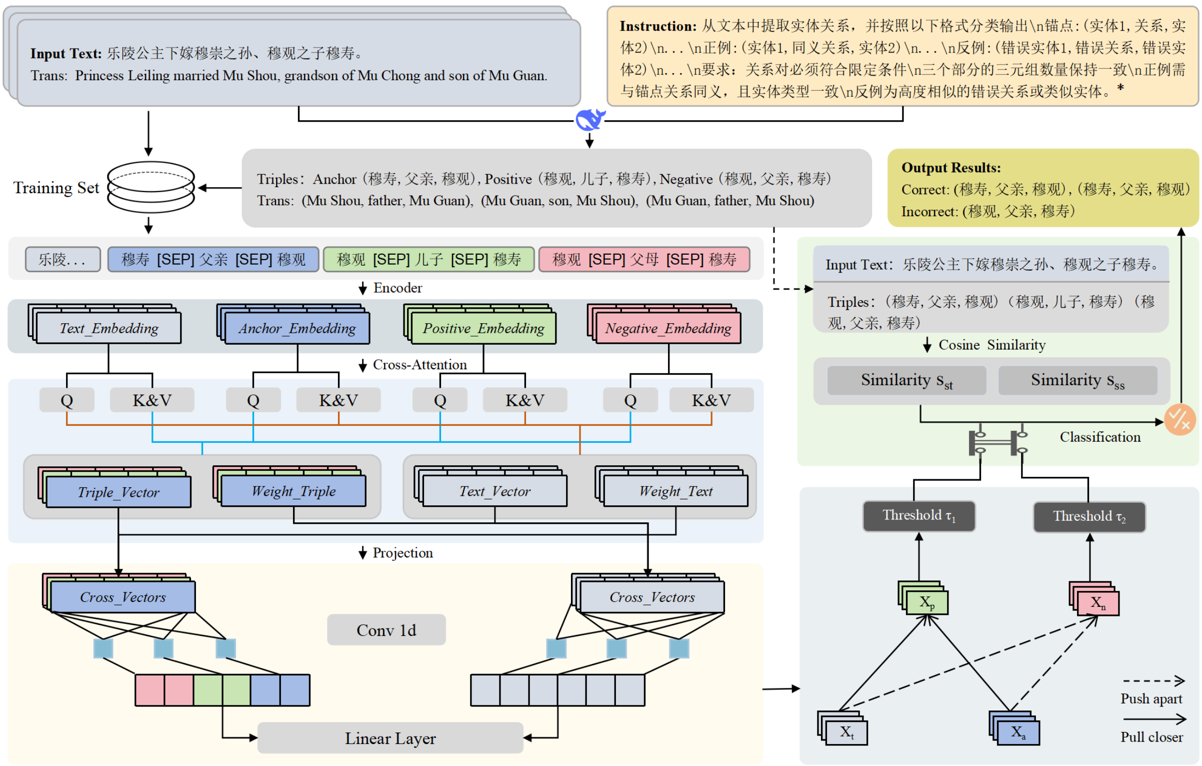

Figure 1: Model framework.[1]

plied to capture the information interaction of the source text with those of the relation triples, producing interaction vectors that capture nuanced correlations between the source text and inferred relations, thereby improving the model's representation of relation features. Third, CACRE is optimized through the contrastive learning, which is an unsupervised learning strategy (Hastie et al., 2009), enabling the model to differentiate between correct and erroneous triples. Finally, distinguishes valid relations and entities from LLM-generated relation triples. By employing the aforementioned methods, this model can effectively mitigate hallucination and bias issues in LLMs outputs.

The contributions of this paper are summarized as follows.

- The proposed CACRE model significantly enhances the precision and reliability of ZSRE performed by LLMs, exhibiting exceptional capabilities in identifying erroneous relations, particularly in accurately distinguishing incorrect triples that closely resemble correct ones.

- The proposed cross-attention mechanism facilitates bidirectional information exchange between relation triples and text embeddings, effectively enhancing the semantic representation of relations.

- Building upon the concept of contrastive learning, a projection network module and a fusion function are designed to effectively calculate the text and relation triples, integrating local and global semantics to capture the feature differences between correct and incorrect examples.

2 Methodology

This section introduces the four main modules in the CACRE model. First, task-specific instructions guide LLMs to extract relation triples, categorized as anchors, positives, and negatives. Second, the text and triples are encoded using the RoBERTa-Chinese-base (Liu et al., 2019), and then, these embeddings are joined using a cross-attention mechanism, which enhances their interactions to produce

[1]The translation marked with * in Figure 1: Extract entity relations from text and categorize the output in the following format\n Anchor: (Entity 1, Relation, Entity 2)\n... \n Positive: (Entity 1, Synonymous Relation, Entity 2)\n...\n Negative: (Wrong Entity 1, Wrong Relation, Wrong Entity 2)\n...\n Requirements: The relation pair must meet the following qualifiers\n The number of triples in the three parts is the same\n The positive example must have a synonymous relation and the same entity type as the anchor\n The negative example should involve a highly similar incorrect relation or entity.

fine-grained and enriched embeddings. Third, the model is trained with a contrastive learning framework that aligns positive samples with the source text while distancing negatives in the embedding space. Finally, the trained model predicts the effectivity of relations and entities based on the text and the corresponding relation triples.

2.1 RE via LLM

As shown in Figure 1, a custom-designed instruction guides the DeepSeek V3 (DeepSeek-AI, 2024) model in producing high-quality structured data. The instruction was designed to guide DeepSeek V3 output three types of relation triples—anchors, positives, and negatives—from each text input. The instruction constraint specifies that the DeepSeek V3 should extract relations and entities from a predefined relation set.

Given a text T, a relation triple $R = \{(e_1, r, e_2)\}$ is defined, where e_1 and e_2 are entities, and r represents the relation between them. Further defined anchor relation triples R_a, positive relation triples R_p, and negative relation triples R_n. The DeepSeek V3 is applied to extract relation triples R and a function f_{extract} is defined to express the extraction of relations, resulting in a set of triples R_t:

$$R_t = f_{\text{extract}}(T) = \{R_{a_m}, R_{p_m}, R_{n_m}\}, \quad (1)$$

where m is the number in one of the three types. So, for a single sample $S = \{T, R_t\}$, R_a serve as anchors, while R_p are valid semantic correlations to R_a and are designed to strengthen the model's learning of correct relations. In contrast, R_n introduce deliberate errors at the relation triples compared to R_a or R_p, yet remain highly similar in form to correct triples.

2.2 Feature Extraction using Encoding and Cross-Attention Mechanisms

This paper adopts the RoBERTa-Chinese-base to encode the text T and relation triples R_t. S is transformed into its encoded representation $S' = \{\mathcal{E}(T), \mathcal{E}(R_t)\}$, where $\mathcal{E}(T)$ is the encoded T, $\mathcal{E}(R_t)$ is the encoded R_t. Then a cross-attention mechanism is introduced to compute the semantic associations between $\mathcal{E}(T)$ and $\mathcal{E}(R_t)$, enabling the model to capture finer-grained and multidimensional representations of semantic relations. Therefore, the text vector V_{T} and the triple-based vector V_{R} is separately calculated by the attention mechanism. Meanwhile, the attention mechanism computes weight matrixes W_{T} and W_{R}, quantifying the fine-grained alignment between the text and the triples.

$$V_{\text{vector}} = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \quad (2)$$

where V_{vector} denotes the specific interaction vector, and when calculating V_{T}, $\mathcal{E}(R_t)$ is used as the Key and Value, $\mathcal{E}(T)$ is used as Query, while calculating V_{R}, $\mathcal{E}(R_t)$ is used as the Query and $\mathcal{E}(T)$ is used as the Key and Value. $\top$ denotes the transpose operation, d_k is the dimension of the key vectors.

Further, the $\tilde{V}_{\text{T}}$ and the $\tilde{V}_{\text{R}}$ is dynamically aggregated by follow formula with weight matrix W.

$$W_{\text{avg}} = \frac{1}{n_{\text{heads}}}\sum_{i=1}^{n_{\text{heads}}} W_i, \quad (3)$$

$$\tilde{V} = \text{bmm}(W_{\text{avg}}, V_{\text{vector}}), \quad (4)$$

where W_{i} represents the attention weight matrix W for the i-th head, W_{avg} is the average of these attention weights across all heads. The bmm represents batch matrix multiplication.

2.3 Contrastive Learning

Contrastive learning (Li et al., 2020) is an unsupervised learning approach (Giorgi et al., 2021) that optimizes the spatial distribution of embeddings by encouraging semantically similar samples to cluster closely while pushing dissimilar samples farther apart. This paper adopts the SimCLR framework (Chen et al., 2020), which excels in semantic representation learning. To tailor SimCLR for RE tasks, the projection network module is enhanced with specific optimizations. The projection network module is employed to map tensors into a projection space, which consists of 1D convolution, activation functions, and a fully connected layer.

For tensors $\tilde{V}_{\text{T}}$ and $\tilde{V}_{\text{R}}$, denoted as input $G = \{\tilde{V}_{\text{T}}, \tilde{V}_{\text{R}}\}$, the process is carried out through through a 1D convolutional layer for local feature extraction, followed by linear projection and nonlinear activation, ultimately yielding the projected output:

$$X_{\text{proj}} = F(\text{Linear}(\text{Conv1D}(G))), \quad (5)$$

where $F(\cdot)$ is defined as

$$\begin{aligned} F(x) = \text{LayerNorm}(&\text{LeakyReLU}(\\ &\text{Linear}(W_p \cdot x + b_p) \\ &\cdot W_y + b_y)), \end{aligned} \quad (6)$$

where x is the input, W_p and b_p are the parameters for the first linear transformation, and W_y and b_y are the parameters for the final transformation. Meanwhile, a residual connection is incorporated to obtain the final output $X = \{X_{\mathrm{T}}, X_{\mathrm{R}}\} = \{X_{\mathrm{T}}, X_{\mathrm{a}}, X_{\mathrm{p}}, X_{\mathrm{n}}^{(1)}, X_{\mathrm{n}}^{(2)}, \ldots, X_{\mathrm{n}}^{(N)}\}$. If the input dimension H matches the projection dimension P, the residual is directly added. Otherwise, the input is linearly transformed to match the projection dimension:

$$X = X_{\text{proj}} + \lambda \cdot (V_{\mathrm{w}} W_{\text{res}} + b_{\text{res}}), \tag{7}$$

where $W_{\text{res}} \in \mathbb{R}^{H \times P}$, V_{w} and $b_{\text{res}} \in \mathbb{R}^{P}$ are the weights and biases for the residual connection, and λ is a learnable scaling factor controlling the contribution of the residual.

As shown in Figure 1, the CACRE is optimized by two key perspectives: enhancing semantic alignment between the text and triples, and refining semantic distinctions within the triples themselves. Additionally, emphasis is placed on enhancing semantic differentiation within triples to strengthen the model's ability to discern subtle semantic details. To achieve this, a novel multi-granularity InfoNCE loss function is proposed, extending the traditional InfoNCE loss (Oord et al., 2018) by incorporating both local and global feature similarities. To further address challenging samples and enhance the learning capacity of CACRE, a ratio-based hard negative sample selection strategy and a dynamic margin adjustment mechanism are integrated to optimize training efficiency.

First, X_{T} is designated as an anchor X_{a} to maximize its similarity with the positive sample X_{p} while minimizing similarity with negative samples X_{n}. Then, similarly, the same applies to X_{a} itself. The local similarity is computed between the anchor X_{a}, the positive X_{p}, and the negatives $X_{\mathrm{n}}^{(i)}$:

$$S_{\text{local}}^{+} = \frac{1}{\tau} \cos\left(X_{\mathrm{a}},\, X_{\mathrm{p}}\right), \tag{8}$$

$$S_{\text{local}}^{-}(i) = \frac{1}{\tau} \cos\left(X_{\mathrm{a}},\, X_{\mathrm{n}}^{(i)}\right), \quad i = 1, \ldots, N_{\text{neg}}, \tag{9}$$

where τ is the temperature parameter. The local similarity is averaged over the sequence length L to obtain the final local similarity scores:

$$\bar{S}_{\text{local}}^{+} = \frac{1}{L} \sum_{j=1}^{L} S_{\text{local}}^{+}[j], \tag{10}$$

$$\bar{S}_{\text{local}}^{-}(i) = \frac{1}{L} \sum_{j=1}^{L} S_{\text{local}}^{-}(i)[j], \quad i = 1, \ldots, N_{\text{neg}}, \tag{11}$$

where j represents the time step in the sequence L. The global representations are obtained by mean-pooling over the sequence dimension:

$$\mu_{\mathrm{a}} = \frac{1}{L} \sum_{j=1}^{L} X_{\mathrm{a}}[j], \qquad \mu_{\mathrm{p}} = \frac{1}{L} \sum_{j=1}^{L} X_{\mathrm{p}}[j], \tag{12}$$

$$\mu_{\mathrm{n}}^{(i)} = \frac{1}{L} \sum_{j=1}^{L} X_{\mathrm{n}}^{(i)}[j], \quad i = 1, \ldots, N_{\text{neg}}. \tag{13}$$

Using these mean-pooled representations, the global similarities are computed as:

$$S_{\text{global}}^{+} = \frac{1}{\tau} \cos\left(\mu_{\mathrm{a}},\, \mu_{\mathrm{p}}\right), \tag{14}$$

$$S_{\text{global}}^{-}(i) = \frac{1}{\tau} \cos\left(\mu_{\mathrm{a}},\, \mu_{\mathrm{n}}^{(i)}\right), \quad i = 1, \ldots, N_{\text{neg}}. \tag{15}$$

The final similarity scores are a weighted combination of the local and global similarities:

$$S^{+} = w \cdot \bar{S}_{\text{local}}^{+} + (1 - w) \cdot S_{\text{global}}^{+}, \tag{16}$$

$$S^{-}(i) = w \cdot \bar{S}_{\text{local}}^{-}(i) + (1 - w) \cdot S_{\text{global}}^{-}(i), \tag{17}$$

where w controls the contribution of local versus global similarity.

To further enhance the discriminative ability against hard negative samples, CACRE selects hard negative samples in one batch. Let $K = \lceil \text{top}_k \cdot N_{\text{neg}} \rceil$, top_k is a ratio coefficient used to determine the number of hard negative samples K and define:

$$\{S_{\text{hard}}^{-}(i)\}_{i=1}^{K} = \text{topk}\Big[\{S^{-}(i)\}_{i=1}^{N_{\text{neg}}},\, K\Big]. \tag{18}$$

In addition, a dynamic margin is introduced to adjust the difficulty of negatives:

$$m = \text{base\_margin} + \beta\Big[S_{\text{global}}^{+} - \frac{1}{N_{\text{neg}}} \sum_{i=1}^{N_{\text{neg}}} S_{\text{global}}^{-}(i)\Big], \tag{19}$$

where m is the dynamic margin, base_margin is the initial margin, and β is a scaling factor.

The logits are then defined by concatenating the positive similarity score with hard negative similarity scores, followed by subtraction of m:

$$\text{logits} = \Big[S^+,\ S^-_{\text{hard}}(1) - m,\ S^-_{\text{hard}}(2) - m,\ \ldots,\ S^-_{\text{hard}}(K) - m\Big]. \tag{20}$$

Finally, the multi-granularity InfoNCE loss is computed as the negative log-likelihood of the positive similarity:

$$\mathcal{L} = -\log\left[\frac{\exp(S^+)}{\exp(S^+) + \sum_{i=1}^{K} \exp\big(S^-_{\text{hard}}(i) - m\big)}\right] \tag{21}$$

2.4 Data Prediction through CACRE

In unsupervised learning, prediction tasks fundamentally rely on feature vectors derived from model outputs. This paper leverages LLMs to extract relation triples, which are subsequently fed into the CACRE model. The CACRE model processes the text and relation triples to generate corresponding feature vectors. Subsequently, the similarity between the text and the triples, as well as the similarity among the triples themselves, is computed and compared against thresholds optimized during training.

The similarity between the text and the triple is denoted as S_{st}, while S_{tt} represents the internal similarity among the components of the triple. In Figure 1, when the S_{st} is greater than or equal to threshold τ_1, the relation triple is deemed preliminarily reliable, when the S_{tt} is greater than or equal to threshold τ_2, the relation triple is regarded as accurate.

3 Experiments

3.1 Dataset

This experiment uses the DuIE2.0 dataset (Li et al., 2019), an open-source Chinese dataset for entity relation extraction. Because the test set of this dataset is not publicly available, this paper compares model improvement performance by selecting 10% of the training data for model training and 2% for validation, and the validation set of the original dataset is used as the test set. The original DuIE2.0 dataset and experimental details are shown in Table 1.

	Dataset	#Sentences	#Triples	#Relations
Original	Train	171293	310709	48
	Validation	20674	37825	48
Experiment	Train	18618	128823	48
	Validation	3499	24468	48
	Test	20674	37825	48

Table 1: Statistics of DuIE2.0 dataset.

3.2 Experimental Environment and Sets

The experiments were conducted on a computing system equipped with two NVIDIA A800 80GB PCIe GPUs, providing a total of 160GB of memory. The system operated on Ubuntu 20.04.6 LTS. Table 2 shows the hyperparameter configurations that were utilized during the model training process. Due to DeepSeek V3's excellent ability in ZSRE demonstrated in Table 3, it was selected as the RE model to generate training data.

Hyperparameter	Value
Pre-training model	RoBERTa-Chinese-base
Max-sequence length	128
Learning rate	1×10^{-4}
Batch size	32
Projection dimension	512
Epochs	70
Temperature	0.07
Dropout rate	0.2
Base margin	0.08
Top$_k$	0.8
Weight w	0.7

Table 2: Hyperparameter settings for training functions.

3.3 Evaluation Metric

This experiment evaluates model performance using the metrics of precision, recall, and F1-score.

3.4 Compared Models

DEPR: The model proposes a dual-head framework for entity and relation prediction, aiming to jointly tackle entity recognition and RE (Xiao et al., 2023).

CasRelBLCF: The model addresses overlapping triples through entity mapping and leverages deep reinforcement learning to filter distant supervision noise (Tang et al., 2024).

CECRel: CECRel is a contrastive learning-based unified model for entity and relation extraction. It enhances information extraction by leveraging data augmentation and feature enhancement (Tong et al., 2025).

Learning Setting	Model	Compared Models			CACRE		
		P	R	F1	P	R	F1
Non-Zero-shot	DEPR	71.1	65.4	68.1	-	-	-
	CasRelBLCF	74.0	68.6	71.2	-	-	-
	CECRel	76.8	79.7	74.1	-	-	-
	Electra-based Joint Model	**78.9**	71.2	**74.8**	-	-	-
Zero-shot	Chen (Chen and Liu, 2024)	-	-	73.4	-	-	-
	LLaMA 3:70b	45.6	41.4	43.4	57.9 (+12.3)	40.7	47.8 (+4.4)
	Qwen-Plus	61.0	69.0	65.0	70.9 (+9.9)	68.6	69.7 (+4.7)
	ERNIE 4.0	64.1	74.4	69.0	**78.9** (+14.8)	72.3	75.5 (+6.5)
	DeepSeek V3	65.2	**80.1**	71.9	76.6 (+11.3)	**78.0**	**77.4** (+5.5)

Table 3: Main experiment results of different models.

Electra-based Joint Model: The model employs a joint learning approach to mitigate the issue of entity overlap, thereby improving the accuracy of RE (Zhu et al., 2020).
Chen (Chen and Liu, 2024): They proposed a multi-agent cooperative framework that employs multiple specialized agents to improve LLM performance in constructing knowledge graphs.
Qwen-Plus: Qwen-Plus is an advanced pre-trained large language model developed by Alibaba Cloud, providing powerful natural language processing capabilities (Bai et al., 2023).
ERNIE 4.0: ERNIE 4.0 is a pre-trained language model developed by Baidu, offering powerful natural language processing capabilities by integrating multiscale knowledge and structured information (Zhang et al., 2019).
LLaMA 3: LLaMA 3 is the latest language model developed by Meta, offering efficiency and scalability (Dubey et al., 2024).
DeepSeek V3: DeepSeek V3 is a model specifically optimized for deep learning tasks, equipped with powerful feature extraction and classification capabilities.

3.5 Main Results

Table 3 displays the main results of comparative experiments. The results of LLMs based on the experimental dataset. For different learning sets, baseline models are categorized into two types: zero-shot and non-zero-shot learning.
Zero-shot: Compared to LLMs that performing ZSRE, CACRE improves precision by an average of 12% and recall by 5.3%. Further, directly applying LLMs for result extraction underperforms compared to Chen (Chen and Liu, 2024). After integrating CACRE, DeepSeek's F1 score increased by 4% over Chen's, while ERNIE 4.0 achieved a 2.1% improvement. However, the experimental results indicate that CACRE has a adverse effect on recall rate that the average decline is 1.1%, primarily due to the model's reliance on training data generated by LLMs, which contains hallucinated outputs and causes error propagation.
Non-Zero-shot: Experimental results demonstrate that although directly applying LLMs for RE does not universally surpass supervised learning methods, the CACRE yields significant improvements. Compared to the Electra-based Joint Model, DeepSeek-V3 and ERNIE 4.0 with CACRE achieve 2.6% and 0.7% higher recall rates, respectively. It indicates that CACRE can effectively filter incorrect relations and entities through contrastive learning, thereby substantially enhancing RE task performance.

3.5.1 Experimental Analysis

In order to effectively distinguish negative samples, CACRE can be viewed as a binary classification task for positive and negative samples. During the validation phase, Receiver Operating Characteristic Curve (ROC) as the auxiliary evaluation metric, as it provides a comprehensive assessment of the model's performance across various decision thresholds.

As shown in Figure 2, this experiment presents two different ROC curves for comparison: the Anchor ROC curve reflects the model's ability to differentiate similarities within relation triples and aims to evaluate its ability to identify positive and negative samples in relation triples; whereas the Text ROC curve focuses on evaluating the model's

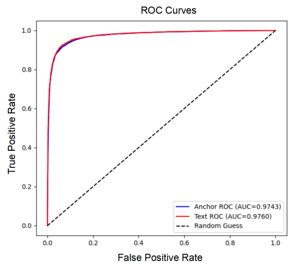

Figure 2: ROC for CACRE performance evaluation.

ability to differentiate between positive and negative samples in text and relation triples. The experimental results show that the CACRE model demonstrated significant advantages in distinguishing positive and negative samples.

3.5.2 Distribution Visualization of Data

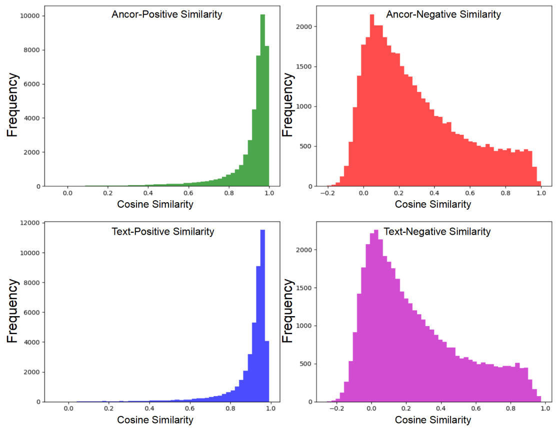

Figure 3: Similarity distributions of anchor and text samples.

To intuitively demonstrate the learning effectiveness of the CACRE model, a visual analysis of the similarity distribution was conducted during the validation phase. Specifically, the similarity was calculated the similarity of the text and the anchor with the positive and the negative samples, respectively.

As shown in Figure 3, the distribution of Anchor-Positive similarity is concentrated near 1, significantly higher than the distribution of Anchor-Negative similarity. Similarly, the distribution of Text-Positive similarity tends to show higher similarity values, further confirming the high similarity between the text and relation triples. In contrast, the similarity distribution of Text-Negative is lower, with a notable difference in similarity between positive and negative samples, indicating that CACRE can effectively distinguish between positive and negative samples.

3.6 Ablation Study

This paper includes ablation experiments on each module of CACRE, with the results summarized in Table 4.

Model	**P**	**R**	**F1**
DeepSeek V3	65.2	**80.1**	71.9
CACRE-att	67.0	70.0	68.5
CACRE-pro	73.2	74.9	74.0
CACRE	**76.6**	78.0	**77.4**

Table 4: Submodule performance comparison.

DeepSeeK V3 as a based model, CACRE-pro refers to the version of the model trained without the projection layer, while CACRE-att indicates the removal of the cross-attention mechanism during training. The CACRE-att model demonstrated relatively weaker performance, primarily due to the incorrect exclusion of numerous true positive samples, which significantly reduced its recall rate. CACRE-att showed a 1.8% improvement in precision, but a 10.1% decrease in recall. CACRE-pro achieved a 6.5% increase in F1 score compared to CACRE-att. CACRE demonstrates a substantial improvement in terms of both accuracy and F1 score.

3.7 Case Study

The performance of CACRE on samples was exhibited in Table 5. The table shows two samples with their similarity scores, regardless of whether it concerns Text 1 or Text 2, indicate a significantly higher similarity between the text and positive examples compared to that between the text and negative examples. Similarly, the similarity between the anchor and positive examples is markedly higher than that between the anchor and negative examples. These findings provide that the CACRE model effectively discriminates between negative samples that are highly similar to the correct triples but are actually incorrect.

4 Related Work

ZSRE is an important research direction due to its ability to identify and extract relations without the need for annotated data. Traditional methods, several approaches (Socher et al., 2013) proposed cross-modal transfer learning methods for zero-shot learning, achieving joint embedding through contrastive learning of textual representations. Levy et al. (Levy et al., 2017) formulate

Text	Relation Triple	Similarity Score
Input Text 1: 秦始皇的母亲是赵姬。 Trans: The first being the famous **mother** of the **First Emperor of Qin**, **Zhao Ji**.	**Anchor:** [秦始皇,母亲,赵姬] Trans: [First Emperor of Qin, mother, Zhao Ji] **Positive:** [赵姬,儿子,秦始皇] Trans: [Zhao Ji, son, First Emperor of Qin] **Negative:** [赵姬,母亲,秦始皇] Trans: [Zhao Ji, mother, First Emperor of Qin]	**Anchor-Negative:** 0.6720 **Positive-Anchor:** 0.8497 **Positive-Text:** 0.8767 **Negative-Text:** 0.7112
Input Text 2:《救世神棍》是一部由李志毅导演、梁朝伟等人主演的剧情片。 Trans: "**Savior Charlatan**" is a drama film directed by Lee Chi-Ngai and **starring Tony Leung** and others.	**Anchor:** [救世神棍,主演,梁朝伟] Trans: [Savior Charlatan, starred in, Tony Leung] **Positive:** [救世神棍,演员,梁朝伟] Trans: [Savior Charlatan, actor, Tony Leung] **Negative:** [救世神棍,导演,梁朝伟] Trans: [Savior Charlatan, director, Tony Leung]	**Anchor-Negative:** 0.5691 **Positive-Anchor:** 0.8273 **Positive-Text:** 0.9272 **Negative-Text:** 0.6488

Table 5: Samples analysis.

RE as a machine reading comprehension task by transforming it into question-answering problems. Additionally, Shin et al. (Shin et al., 2020) based on templates are suitable for relations with well-defined rules but are less flexible. Other methods use dictionaries and knowledge bases (Trisedya et al., 2019) for inference rely on existing resources, but they are limited by the scope of the knowledge base and similar resources. Recently, LLMs including BERT (Devlin et al., 2019), GPT (Radford et al., 2018), and T5 (Raffel et al., 2020), have shown great promise for ZSRE by leveraging vast pre-trained knowledge and strong natural language reasoning capabilities. However, a notable challenge in applying LLMs to ZSRE is hallucination outputs, where models generate semantically incorrect yet syntactically plausible relations (Chen and Li, 2021).

To mitigate hallucination bias in ZSRE, several works have introduced contrastive learning methods. Theodoropoulos et al. (Theodoropoulos et al., 2021) and Chen et al. (Chen et al., 2020) utilized contrastive learning frameworks to refine relation representations, ensuring that the model can distinguish between valid and erroneous relations. Additionally, Zhou et al. (Zhou et al., 2019) and Wang et al. (Wang et al., 2022) explored reinforcement and multi-task learning techniques to guide the extraction process, reducing hallucination by aligning model outputs with auxiliary tasks or reward signals. In addition, cross-attention mechanisms have been explored to capture complex relation dependencies more accurately. Huang et al. (Huang et al., 2022) demonstrated the effectiveness of cross-attention layers in fusing information between different sources for RE, while Wu et al. (Wu and Shi, 2021) extended this idea by introducing a cross-type attention mechanism to jointly extract entities and relations. More recent approaches have continued to build on these ideas. Luo et al. (Luo et al., 2023) proposed a hierarchical attention mechanism to further enhance relation representation, allowing for better extraction from nested or complex structures. These methods help alleviate hallucination outputs by grounding RE with structured background knowledge.

5 Conclusion

This paper proposes the CACRE model, which designed to address the common issue of hallucinatory outputs in ZSRE tasks based on LLMs. By employing a cross-attention mechanism, CACRE can quantify the fine-grained semantic relations and feature representations between the text and the triples. A multi-granularity fusion function is used to apply contrastive learning, which enables the model to capture subtle differences and enhance its learning performance. CACRE demonstrates robust discriminative capability by effectively identifying and filtering out negative relation triples, thereby significantly improves the precision of LLMs in the task of ZSRE.

Limitations

Despite the demonstrated effectiveness of CACRE in mitigating hallucinated relation triples in ZSRE, several limitations remain. First, the model's performance is inherently contingent upon the quality and distribution of candidate triples generated by the LLMs, excessive noise or bias in these candidates may constrain the upper bound of CACRE's filtering capability. Second, the model exhibits sensitivity to the selection of decision thresholds during inference, which introduces challenges in achieving optimal precision-recall trade-offs across diverse datasets and application scenarios. Furthermore, the contrastive learning framework presupposes the availability of sufficiently informative positive and negative samples, which may not always be guaranteed in practical zero-shot settings. Addressing these limitations is essential for further enhancing the robustness and generalizability of CACRE in applications.

References

Vaibhav Adlakha, Parishad BehnamGhader, Xing Han Lu, Nicholas Meade, and Siva Reddy. 2024. Evaluating correctness and faithfulness of instruction-following models for question answering. *Transactions of the Association for Computational Linguistics*, 12:681–699.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *ArXiv preprint*, abs/2309.16609.

Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Zhe Fu, et al. 2024. Deepseek llm: Scaling open-source language models with longtermism. *ArXiv preprint*, abs/2401.02954.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *ArXiv preprint*, abs/2303.12712.

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, Online. Association for Computational Linguistics.

Gui Chen and Xianhui Liu. 2024. A multi-agent collaborative framework for constructing knowledge graphs from text. In *2024 IEEE International Conference on Knowledge Graph (ICKG)*, pages 9–16.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR.

DeepSeek-AI. 2024. Deepseek-v3 technical report. *ArXiv preprint*, abs/2412.19437.

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, Minneapolis, Minnesota. Association for Computational Linguistics.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *ArXiv preprint*, abs/2407.21783.

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, Online. Association for Computational Linguistics.

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, Suzhou, China. Association for Computational Linguistics.

Trevor Hastie, Robert Tibshirani, Jerome Friedman, Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. Unsupervised learning. *The elements of statistical learning: Data mining, inference, and prediction*, pages 485–585.

Xiaofeng Huang, Zhiqiang Guo, Jialiang Zhang, Hui Cao, and Jie Yang. 2022. Reca: Relation extraction based on cross-attention neural network. *Electronics*, 11(14):2161.

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, Vancouver, Canada. Association for Computational Linguistics.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. 2020. Prototypical contrastive learning of unsupervised representations. *ArXiv preprint*, abs/2005.04966.

Junyi Li, Xiaoxue Cheng, Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. 2023. HaluEval: A large-scale hallucination evaluation benchmark for large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6449–6464, Singapore. Association for Computational Linguistics.

Shuangjie Li, Wei He, Yabing Shi, Wenbin Jiang, Haijin Liang, Ye Jiang, Yang Zhang, Yajuan Lyu, and Yong Zhu. 2019. Duie: A large-scale chinese dataset for information extraction. In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*, pages 791–800. Springer.

Zichao Lin, Shuyan Guan, Wending Zhang, Huiyan Zhang, Yugang Li, and Huaping Zhang. 2024. Towards trustworthy llms: a review on debiasing and dehallucinating in large language models. *Artificial Intelligence Review*, 57(9):243.

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

Haoran Luo, Haihong E, Yuhao Yang, Yikai Guo, Mingzhi Sun, Tianyu Yao, Zichen Tang, Kaiyang Wan, Meina Song, and Wei Lin. 2023. Hahe: Hierarchical attention for hyper-relational knowledge graphs in global and local level. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8095–8107, Toronto, Canada. Association for Computational Linguistics.

Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1105–1116, Berlin, Germany. Association for Computational Linguistics.

Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. 2021. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *ArXiv preprint*, abs/1807.03748.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4222–4235, Online. Association for Computational Linguistics.

Richard Socher, Milind Ganjoo, Christopher D. Manning, and Andrew Y. Ng. 2013. Zero-shot learning through cross-modal transfer. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 935–943.

Hongmei Tang, Dixiongxiao Zhu, Wenzhong Tang, Shuai Wang, Yanyang Wang, and Lihong Wang. 2024. Research on joint model relation extraction method based on entity mapping. *Plos one*, 19(2):e0298974.

Ruixiang Tang, Xiaotian Han, Xiaoqian Jiang, and Xia Hu. 2023. Does synthetic data generation of llms help clinical text mining? *ArXiv preprint*, abs/2303.04360.

Christos Theodoropoulos, James Henderson, Andrei Catalin Coman, and Marie-Francine Moens. 2021. Imposing relation structure in language-model embeddings using contrastive learning. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 337–348, Online. Association for Computational Linguistics.

Yetao Tong, Jijun Tong, Shudong Xia, Qingli Zhou, and Yuqiang Shen. 2025. Cecrel: A joint entity and relation extraction model for chinese electronic medical records of coronary angiography via contrastive learning. *Journal of Biomedical Informatics*, page 104792.

Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, and Rui Zhang. 2019. Neural relation extraction for knowledge base enrichment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 229–240, Florence, Italy. Association for Computational Linguistics.

An Wang, Ao Liu, Hieu Hanh Le, and Haruo Yokota. 2022. Towards effective multi-task interaction for entity-relation extraction: A unified framework with selection recurrent network. *ArXiv preprint*, abs/2202.07281.

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

Hui Wu and Xiaodong Shi. 2021. Synchronous dual network with cross-type attention for joint entity and relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2769–2779, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yanbing Xiao, Guorong Chen, Chongling Du, Lang Li, Yu Yuan, Jincheng Zou, and Jingcheng Liu. 2023. A study on double-headed entities and relations prediction framework for joint triple extraction. *Mathematics*, 11(22):4583.

Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced language representation with informative entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1441–1451, Florence, Italy. Association for Computational Linguistics.

Bo Zhou, Daniel Geißler, and Paul Lukowicz. 2024. Misinforming llms: vulnerabilities, challenges and opportunities. *ArXiv preprint*, abs/2408.01168.

Kai Zhou, Xiangfeng Luo, Hao Wang, and Richard Xu. 2019. Multi-task learning for relation extraction. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1480–1487.

Mingda Zhu, Jiqing Xue, and Gaoyuan Zhou. 2020. Joint extraction of entity and relation based on pretrained language model. In *2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, volume 2, pages 179–183. IEEE.