2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Hybrid Prototype Contrastive Learning with Cross-Attention for Few-Shot Relation Classification

Zeyu Zhang[1[0009-0004-9107-4327]], Shaowei Wang[1[0009-0001-9028-3344]]
Nana Bu[1[0000-0002-9956-5532]], Junzhe Zhang[1[0000-0001-6543-6073]]
and Yuanyuan Xiao[1[0009-0003-7215-8632]]

[1] School of Information Engineering, Chang'an University, Xi'an 710018, China
2601618515@chd.edu.cn

Abstract. Few-shot relation classification (FSRC) is designed to determine the relation class between entities within a text using a limited set of annotated data. Recently, some studies have focused on optimizing prototype representations by incorporating relation information into the prototype network or applying contrastive learning to alleviate the prediction confusion problem. However, these approaches primarily rely on global instance features and relation information, making it difficult to capture fine-grained local semantic information. This can result in inaccurate evaluations of anomalous samples and the mixing of similar categories. In order to tackle these issues, we introduce a novel hybrid prototype contrastive learning (HPCL) model. Dynamically fusing global and local prototypes through a cross-attention mechanism significantly improves the performance of few-shot relation classification. In addition, HPCL combines a dual contrastive learning strategy (relation-prototype contrastive learning and query-prototype contrastive learning) to effectively enhance intra-class feature sharing and inter-class feature discriminability by optimizing prototype representation. We have conducted comprehensive experiments using the public datasets FewRel 1.0 and FewRel 2.0, and the results show that HPCL not only performs well on traditional datasets but also demonstrates a strong generalization ability in cross-domain adaptation tasks, which can effectively alleviate the challenges brought by data scarcity and insufficient relation description.

Keywords: Few-shot relation classification, Prototype network, Relation information, Cross-attention mechanism, Contrastive learning.

1 Introduction

Relation classification (RC) constitutes a significant area within the broader task of relation extraction (RE). When the entities involved are already identified, relation extraction becomes synonymous with relation classification [1]. RC involves the process of determining the relationship between two entities found within a sentence, typically from a specified collection of relations. This task helps to convert unstructured text into structured information. It is a key task in Natural Language Processing (NLP), which is widely used in NLP applications including knowledge graph construction [2], question-answer systems [3], and information retrieval [4], etc. RC usually requires many manually labeled data for training and prediction. Nevertheless, the process of labeling data by hand is labor-intensive and consumes a significant amount of time, often leading to mistakes influenced by subjective factors. Early studies have made progress in the relation classification task through distant supervision [5], which automatically labels data by combining knowledge graphs. However, the noise introduced by remote supervision methods [6] and the long-tailed distribution of knowledge graph data [7] limit their applications.

To address the challenges posed by distant supervision in relation classification (RC), researchers have leveraged the successful strategies employed in few-shot learning within the domain of computer vision (CV) and introduced the few-shot relation classification (FSRC) task [8]. The objective of FSRC research is to tackle the issue of limited labeled data in practical settings and to efficiently acquire knowledge of the target relation using a minimal set of labeled examples, thereby significantly decreasing the labeling expenses. The generalization ability of the model is mainly improved by methods such as meta-learning [9] or metric learning to enhance the prediction performance in the face of new or unseen relations, thus reducing the dependence on large-scale labeled data. Among them, the prototype network [10] generates class prototypes by calculating the average of samples in each category and classifies them based on the distance between each class sample and the class prototypes, which has become a widely used basic method. Current research based on prototype networks is mainly conducted in two directions: to enhance the prototype representation by introducing external information (such as relation names and descriptions). Specifically, this type of approach incorporates relation-specific semantic information into the prototype representation, thereby enhancing the discriminative ability of the model. For example, Yang et al. [11] propose relation and entity description information to enhance the prototype network. Liu et al. [12] propose a prototype correction module to explicitly correct the original prototype using relation information to assess the preservation and update of prototype and relation features, thereby enhancing the capture of relational semantics. Another research direction aims to mitigate the problem of predictive confusion between similar classes using contrastive learning strategies. Such approaches leverage relational information to enhance intra-class similarity while amplifying inter-class differences, thereby improving the discriminative capability of the model. For example, Wu et al. [13] propose multilevel contrastive learning approach aimed at enhancing the common features among instances within specific classes, while simultaneously emphasizing the distinctions between analogous relational classes via the use of instances and prototypes. Although the above approaches have made significant progress, some limitations remain. First, external information-based approaches are usually practical only under the condition of specific relational descriptions, and their enhancement is limited when relational descriptions are insufficient or of low quality. In addition, although contrastive learning-based methods can mitigate class confusion by positioning instances of the same class nearer to one another and instances from different classes further apart, they often assume that instances within a class have consistent representations, ignoring the fact that different instances may have different semantic biases, leading to a persistent risk of misclassification in hard-to-distinguish relation classes. To mitigate these problems, we introduce an efficient few-shot relation classification model known as HPCL, which dynamically fuses global prototypes (extracted from support samples and relation information) and local prototypes (capturing key contexts) through a cross-attention mechanism so that even if the relation description information is incomplete or of low quality, the model can still rely on the complementary information of global and local features to construct robust relation representations, thus reduce the over-reliance on relation descriptions. Meanwhile, in order to solve the problem of confusing similar relation classes, a dual contrastive learning strategy is proposed, whereby the model can effectively adjust the distribution of instances within a class to better match the real data distribution through the relation-prototype contrastive learning and the query-prototype contrastive learning, avoiding the limitation of over-reliance on a single prototype in the traditional approach, to enhance the discriminative capability of relation classes. This paper makes several contributions:

- We present a novel hybrid prototype contrastive learning (HPCL) model that seamlessly integrates global and local prototypes via a cross-attention mechanism. By leveraging the strengths of both global and local features, HPCL greatly enhances the discriminative power of relation classification. In contrast to conventional approaches, HPCL effectively identifies the intricate features of relations, leading to marked performance gains in few-shot relation classification tasks.
- We adopt a dual contrastive learning strategy of relation-prototype and query-prototype contrastive learning. The former enhances the prototype representation through the comparison between relation descriptions and prototypes, and the latter further improves the ability of the model to discriminate between different relation classes through the comparison between query samples and prototypes.
- The results from our experiments using the FewRel 1.0 and FewRel 2.0 datasets demonstrate that HPCL performs well in all types of few-shot settings, further validating its potential in practical applications.

2 Related work

2.1 Few-Shot Relation Classification

Few-shot relation classification (FSRC) requires models to quickly learn textual features to identify relations between entities in the presence of scarce data. The core challenge of this task lies in the ability of models to generalize effectively when facing new relation classes. Currently, FSRC research is divided into two main classes: (1) Meta-learning methods based on optimization, achieve fast adaptation to new tasks by optimizing model parameters. For example, MAML [9] learns a generalized initialization parameter so that the model can adapt quickly and achieve better performance with only a tiny amount of gradient update when encountering a new task. (2) Methods based on metric learning, which do not require complex network structures, simplify the model architecture by designing appropriate embedding spaces and metrics. Among them, the prototype network [10], which generates class prototypes by calculating the feature means of the support samples of each class and classifies the query samples according to their distance from the class prototypes, has become a widely used basic method.

Methods currently utilized that rely on prototype networks typically incorporate external information, such as relation data, to optimize the prototype representation of FSRC, thus enhancing the discriminative ability and generalization performance of the model. Some of these methods directly incorporate relation information to improve the representation of the prototype [11, 14, 15]. For example, Liu et al. [15] directly added relation information (such as relation name and description) to the prototype network to enhance the prototype. This method does not introduce additional parameters and reduces the risk of overfitting. However, it may not perform as well as more sophisticated fusion methods when the quality of the relational information is low, or the sample size is large. In addition to directly enhancing prototype representations, alternative approaches bolster intra-class similarity and amplify inter-class distinctions via contrastive learning [16, 17, 18]. For example, Dong et al. [18] enhance the discriminative properties of prototypes by introducing multilevel contrastive learning. The method still falls short in terms of computational complexity and Prompt dependency. The above studies show that high-quality prototype representations are crucial for the FSRC task. However, existing methods mainly rely on global features of the text, while our approach further introduces local features to capture fine-grained relationship representations.

2.2 Contrastive Learning

Contrastive learning (CL) focuses on developing discriminative feature representations by bringing positive sample pairs closer together while distancing negative sample pairs. Recently, there has been notable advancement in the FSRC domain [16, 17, 18, 19]. Among these advancements, Peng et al. [19] introduced a framework for pre-training an entity mask utilizing contrastive learning, which improves the model's comprehension of context and relationship type information. Additionally, Han et al. [16] put forward a relation-prototype contrastive learning approach that leverages relational data to differentiate between various prototypes. Borchert et al. [17] improved the alignment between instances and their corresponding relation descriptions using instance-relation description contrastive learning, thereby enhancing the discriminative capacity of classification.

Although these methods have achieved significant results in FSRC tasks, they mainly focus on a single perspective (such as instances, relation prototypes, or relation descriptions) and fail to fully exploit the potential of synergistic optimization among different representations. Specifically, although instance-level contrastive learning can enhance local semantic representation, it fails to fully utilize the high-order semantic constraints of relational categories. Prototype-level contrastive learning can model inter-class differences, but due to the limited support set samples, the prototype representation may be inaccurate, resulting in final classification errors. To avoid these problems, we introduce a dual contrastive learning method to simultaneously optimize the discriminative ability of the model from two perspectives: relation representations and query samples. We utilize relation prototypes as constraints in the feature space to guide the optimization direction of relation representation and query samples. This strategy not only enhances the ability of the model to discriminate between different relation classes but also effectively improves the overall performance of few-shot relation classification. Meanwhile, by utilizing prototypes rather than individual instances as anchors, our dual contrastive learning method effectively mitigates the detrimental effects of intra-class sample differences. Even if there are significant differences within the same class, using prototypes ensures more stable optimization and better generalization.

2.3 Cross-Attention Mechanism

The cross-attention mechanism, a dynamic feature fusion method, shows significant advantages in the FSRC task. The attention mechanism [20] dynamically increases the receptive domains in the network architecture. Unlike the self-attention mechanism, which only focuses on information in the same embedding space, the cross-attention achieves richer feature interactions by asymmetrically combining two independent embedding spaces. Specifically, self-attention calculates attention scores based on query (Q), key (K), and value (V) within a shared labeling space. On the other hand, cross-attention leverages K and V projected from distinct labeling spaces as contextual cues to refine Q, thereby improving the discriminative capability of feature representations.

In FSRC tasks, existing methods are usually based on a single prototype representation level using only global or local features. However, global features can capture the overall semantics of a relation class but tend to ignore fine-grained local information (such as specific interaction patterns between entities). In contrast, local features can reflect detailed information but lack dynamic adaptability to complex contexts. To address this problem, we introduce a cross-attention mechanism that dynamically fuses global prototypes (consisting of the global feature means of the support samples plus the global features of the relations) and local prototypes (consisting of the local feature means of the support samples plus the local features of the relations) to generate hybrid prototype representations with multi-granular discriminative power. This mechanism enables the model to flexibly modify the level of focus on various types of information within the feature space, thus modeling the relation semantics more comprehensively. This strategy not only enhances the ability of the model to discriminate between complex relation classes but also makes more effective use of limited few-shot data for feature learning, thus improving the generalization performance of the FSRC task.

3 Problem Formulation

We employ a conventional N-way-K-shot few-shot relation classification task setup, in which the dataset is segmented into training $\mathcal{D}_{train}$, validation $\mathcal{D}_{val}$, and test sets $\mathcal{D}_{test}$, and the relation classes in these sets do not overlap. Throughout the training phase, we utilize a meta-learning strategy [21] that facilitates the model in acquiring generalization skills from the base relation classes, allowing it to adapt swiftly when faced with novel relation classes. Each N-way-K-shot task includes a support set $S = \left\{s_j^i; i = 1, \cdots, N, j = 1, \cdots, K\right\}$ and a query set $Q = \{q_i; i = 1, \cdots, M\}$. The support set S

comprises N different relations, with each relation represented by K labeled examples, while the query set Q consists of M unlabeled instances, aimed at predicting the respective relation label. Every instance (x, e, y) is made up of a context sentence $x = \{x_1, x_2, \cdots, x_n\}$, where n is the sentence length, an entity pair $e = (e_h, e_t)$ where e_h and e_t representing the head and tail entities, respectively, and a relationship label y.

4 Proposed Method

In this section, we present the specifics of the HPCL methodology. The overall structure is illustrated in Fig. 1. The goal of HPCL is to enhance the effectiveness of few-shot relation classification by refining the prototype fusion technique and optimizing the contrastive learning approach. The input consists of N-way-K-shot tasks, drawn from the FewRel dataset. The HPCL framework primarily includes two components: (1) A cross-attention mechanism is employed to dynamically integrate global prototypes with local ones, ultimately creating more distinctive relation prototypes. This technique not only boosts the expressive capability of the prototypes but also improves the ability to detect subtle differentiations between relation classes. (2) The implementation of relation-prototype contrastive learning and query-prototype contrastive learning serves to optimize the relation representation, enhancing the model's discrimination power, thereby enabling it to differentiate similar classes more effectively and improve its generalization performance.

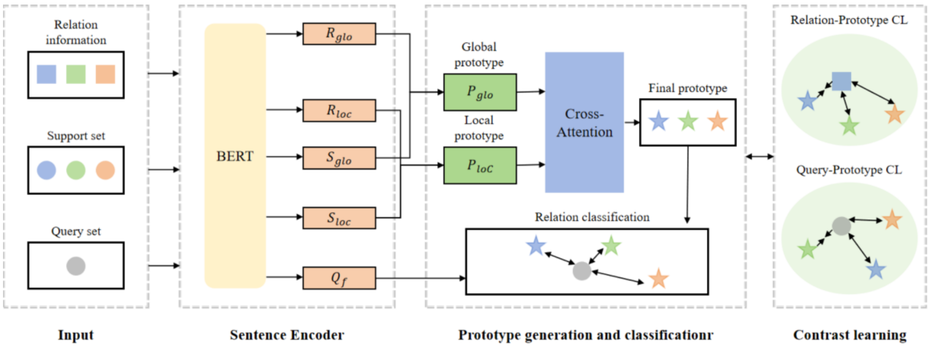

Fig. 1. The overall framework of HPCL. Relation information is represented by squares, circles represent support sets, gray circles represent query sets, and pentagrams represent prototypes. The same color (blue, green, orange) indicates the same relation class.

4.1 Hybrid Prototype

We utilize BERT [23] as an encoder to derive contextual embeddings for the support S and the query set Q. For every relation, we combine the name with the description and feed the resulting sequence into the BERT encoder, which produces the relation embedding $\{R^i \in \mathbb{R}^d; i = 1, \cdots, N\}$, with N representing the total count of relation classes.

For instances in the support S and query sets Q, the global features are obtained by referring to the method of Baldini Soares et al. [22], where the hidden states corresponding to the start tokens of two entity mentions are extracted and spliced in order to form the global representation of the instance $[h_{entity1}; h_{entity2}], h_{entity1}, h_{entity2} \in \mathbb{R}^d$, where d is the dimensionality size of the contextual representations generated by the sentence encoder. The global features of the relation description are extracted from the embedded representation of the corresponding [CLS] tokens. For each relation, we refer to the method of Snell et al. [10] by averaging the global feature of its support samples $\{s_k^i \in \mathbb{R}^{2d}; i = 1, \cdots, N, k = 1, \cdots, K\}$, where K is the number of support samples, and combining it with the global feature of the relation description $\{r_{glo}^i \in \mathbb{R}^{2d}; i = 1, \cdots, N\}$ to construct the final global prototype representation.

$$p_{glo}^i = \frac{1}{K}\sum_{k=1}^{K} s_k^i + r_{glo}^i \in \mathbb{R}^{2d} \tag{1}$$

Global prototypes can capture the overall characteristics of the relation but often ignore fine-grained differences between different instances. Using only global prototypes may reduce ability to distinguish between subtle relation classes. To tackle this issue, we propose the use of local prototypes, which focus on the local information of instances and enhance the ability of the model to perceive fine-grained features. By incorporating local prototypes, the model's discriminative capability is enhanced, leading to increased accuracy in managing relationships that are similar yet distinct.

For relation i, we extract the local features of its k-th support sample to capture its fine-grained semantic information:

$$\hat{s}_k^i = \sum_m^L \alpha_m^s \left[S_k^i\right]_m \in \mathbb{R}^d \tag{2}$$

$$\alpha^s = softmax\left(sum\left(S_k^i\left(S_k^i\right)^T\right)\right) \in \mathbb{R}^L \tag{3}$$

Where L represents the number of tokens for instance, $[\cdot]_m$ is the m-th row of the feature matrix, and the weight α_m^s is obtained by calculating the similarity between the local features. The weights are then summed to form the final local features.

Similarly, we compute the local features of the relation description through the self-attention mechanism:

$$r_{\text{loc}}^i = \sum_m^L \alpha_m^r \, [R^i]_m \in \mathbb{R}^d \tag{4}$$

$$\alpha^r = softmax\left(sum(R^i(R^i)^T)\right) \in \mathbb{R}^L \tag{5}$$

Ultimately, we create a local prototype representation by averaging the local characteristics of the support set and combining them with the local features of the relation description:

$$p_{loc}^i = \frac{1}{K}\sum_{k=1}^{K} \hat{s}_k^i + r_{loc}^i \in \mathbb{R}^d \tag{6}$$

As illustrated in Fig. 2, we utilize a cross-attention mechanism to dynamically combine global and local prototypes, allowing us to capture the interaction between these two types of features, ultimately resulting in a more distinctive hybrid prototype representation. To maintain dimensional consistency, a linear transformation is applied to the local prototype:

$$p_{loc}^{i} = Wp_{loc}^{i} + b \in \mathbb{R}^{2d} \tag{7}$$

where $W \in \mathbb{R}^{2d \times d}$ is a weight matrix and $b \in \mathbb{R}^{2d}$ represent the bias vector.

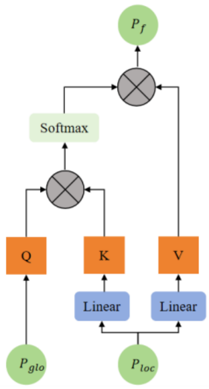

Fig. 2. The fundamental architecture of the cross-attention mechanism. The inputs are transformed linearly to produce queries (Q), keys (K), and values (V). The attention score is computed using the dot product of the query and key, which is then normalized through the Softmax function to create attention weights. These weights are utilized for a weighted summation of the values (V) to yield the final output of attention. The integration of global and local features $\left(p_{glo}\right.$ and $\left.p_{loc}\right)$ and the application of the Softmax function are labeled in the figure.

Then, we compute the hybrid prototype through the cross-attention mechanism:

$$p^{i} = CrossAttention\left(p_{glo}^{i}, p_{loc}^{i}, p_{loc}^{i}\right) \in \mathbb{R}^{2d} \tag{8}$$

$$CrossAttention(Q, K, V) = softmax\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{9}$$

Where Q represents the query, K denotes the key, V signifies the value, and d_k refers the dimension of the key, which serves as a scaling factor to prevent excessively large values.

4.2 Contrastive Learning

In the task of few-shot relation classification task, the model needs to accurately summarize the features of different relation classes using a limited set of labeled data. To address this, we implement a dual contrastive learning approach aimed at enhancing the model's representational capacity. This technique ensures that instances belonging to the same relation class are positioned closer together in the representation space, whereas instances from different relation classes are dispersed further apart.

Specifically, we designed Relation-Prototype Contrastive Learning (RPCL) to use relation descriptions as anchors. The objective is to minimize the distance between them and prototypes of the same relation class while maximizing the distance from prototypes of other classes, thereby enhancing the discriminative capability.

The representation of the relation description is obtained by fusing its global features with local features through a cross-attention mechanism:

$$r^i = CrossAttention\left(r^i_{glo}, r^i_{loc}, r^i_{loc}\right) \in \mathbb{R}^{2d} \tag{10}$$

The relation-prototype contrast loss is calculated as follows:

$$\mathcal{L}_{RPCL} = -\sum_{i=1}^{N} log\left(\frac{exp\left(s\left(r^i, p^i_{hyb}\right)/\tau\right)}{exp\left(s\left(r^i, p^i_{hyb}\right)/\tau\right) + \sum_{j\neq i} exp\left(s\left(r^i, p^j_{hyb}\right)/\tau\right)}\right) \tag{11}$$

Where $s(,)$ is the distance metric function, we use the dot product as a similarity measure between two features, and τ is the temperature coefficient.

We designed Query-Prototype Contrastive Learning (QPCL) with query samples as anchors, where positive samples were selected as the closest relation prototypes, and negative samples as prototypes from other classes. This enables the model to more accurately classify query samples into the correct relational classes.

The query-prototype contrast loss is calculated as follows:

$$\mathcal{L}_{QPCL} = -\sum_{i=1}^{N} log\left(\frac{exp\left(s\left(q^i, p^i_{hyb}\right)/\tau\right)}{exp\left(s\left(q^i, p^i_{hyb}\right)/\tau\right) + \sum_{j\neq i} exp\left(s\left(q^i, p^j_{hyb}\right)/\tau\right)}\right) \tag{12}$$

Where q^i represents the feature representation of the query sample.

4.3 Relation Classification

Upon acquiring N final relation prototypes, we determine the similarity of the query samples with each prototype and utilize the class linked to the relation prototype that exhibits the greatest similarity as the ultimate prediction. The classification loss is measured using the cross-entropy (CE) loss function.

$$\mathcal{L}_{CE} = -log\left(z_y\right) \tag{13}$$

The class label is represented by y, while z_y denotes the predicted probability for class y. The overall loss of our model is defined as:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda\mathcal{L}_{RPCL} + \lambda\mathcal{L}_{QPCL} \tag{14}$$

where λ is the hyperparameter.

5 Experiments

This section assesses the efficacy of the suggested HPCL approach using publicly accessible datasets. Additionally, we perform a set of experiments to investigate how each element of the HPCL method influences the few-shot relation classification task across various settings and thoroughly analyze its contributions to performance.

5.1 Dataset

We utilize two benchmark datasets, FewRel 1.0 [8] and FewRel 2.0 [25], for evaluation purposes. The FewRel 1.0 dataset includes 100 different relation classes, each having 700 instances sourced from Wikipedia domains. These instances are categorized into 64 relations for training, 16 for validation, and 20 for testing. In contrast, FewRel 2.0 introduces a domain adaptation task that builds upon FewRel 1.0. The training set remains the same as that of FewRel 1.0, while the validation set employs the SemEval-2010 task 8 dataset. The testing set is drawn from the biomedical domain (PubMed). Both training and testing for FewRel 1.0 occur within the Wikipedia domain; however, FewRel 2.0 involves training and testing from varied domains, aiming to assess the model's cross-domain adaptability. Additionally, FewRel 1.0 provides names and descriptions of relations, whereas FewRel 2.0 offers only the names, adding an extra layer of complexity to the task.

5.2 Evaluation

A method aligned with the formal evaluations is employed in our approach. Specifically, we randomly select 30,000 episodes from the training dataset, assess 1,000 episodes from the validation dataset, and utilize 10,000 episodes from the test dataset. The metric for evaluation is the mean classification accuracy of the query set. We establish four scenarios: 5-way-1-shot, 5-way-5-shot, 10-way-1-shot, and 10-way-5-shot. We implement the AdamW optimizer, beginning with a learning rate of 2e-5 and a batch size of 4. All experiments were conducted using NVIDIA GeForce RTX 3090 GPUs. Given that the FewRel test set labels are not available to the public, we ensure alignment with previous studies [21, 17] by submitting the model's predictions to the official FewRel leaderboard to acquire the accuracy of the test set.

5.3 Baselines

In order to fully evaluate the performance of HPCL, we compare it with a variety of baseline methods, which are divided into three groups: standard BERT encoder baselines, BERT baselines with external information, and baselines based on the pre-trained relation classification model.

– Standard BERT Encoder Baselines

MAML [9]: A general meta-learning optimization algorithm. GNN [24]: A model combining graph neural networks and meta-learning. Proto-BERT [25]: A prototype network based on BERT for classification by computing instance representations. BERT-PAIR [25]: Splicing query samples and support samples into input sequences and predicting similarity by sequence classification model.

– BERT baselines with external information

REGRAB [26]: A Bayesian meta-learning method using global relation graphs. CTEG [14]: An attention mechanism and confusion-aware training strategy using entity guidance. HCRP [16]: A method for referencing hybrid relation prototypes. SemGL [13]: A method that combines relation graph learning and multilevel contrastive learning through cue enhancement. MultiRep [17]: Combining multiple sentence representations with contrastive learning to enhance the aggregation of instances in different representation spaces. RelPromptCL [18]: Enhances the discriminative properties of relation representations and prototypes by introducing prompt modules and multilevel contrastive learning.

– Baselines based on the pre-trained relation classification model

MTB [22]: A pre-training relation model based on BERT by randomly masking entity mentions. CP [19]: A contrastive learning based on the pre-training framework for entity masks. MapRE [27]: A pre-training model for semantic mapping target relations by combining relation information.

All baseline methods use BERT-base as a sentence encoder to ensure a fair comparison. By comparing with these baseline methods, we are able to comprehensively evaluate the performance of HPCL in the few-shot relation classification task.

5.4 Main results

Table 1 illustrates the performance of HPCL on the FewRel 1.0 dataset for both the validation and test sets. The results from the experiments indicate that HPCL delivers remarkable outcomes in multiple few-shot relation classification tasks.

Table 1. Accuracy (%) of FSRC on FewRel 1.0 validation/test set.

Model	5-way-1-shot		5-way-5-shot		10-way-1-shot		10-way-5-shot	
	val	test	val	test	val	test	val	test
MAML	82.93	89.70	86.21	83.55	73.20	83.17	86.06	88.51
GNN	-	75.66	-	89.06	-	70.08	-	76.93
Proto-BERT	82.92	80.68	91.32	89.60	73.24	71.48	83.68	82.89
BERT-PAIR	85.66	88.32	89.48	93.22	76.84	80.63	81.76	87.02
REGRAB	87.95	90.30	92.54	94.25	80.26	84.09	86.72	89.93
CTEG	84.72	88.11	92.52	95.25	76.01	81.29	84.89	91.33
HCRP	90.90	93.76	93.22	95.66	84.11	89.95	87.79	92.10
SemGL	91.96	**95.11**	**94.70**	96.88	84.98	91.61	89.02	**94.73**
MultiRep	**92.73**	94.18	93.79	96.29	86.12	91.07	88.80	91.98
RelPromptCL	92.12	94.71	94.42	97.38	85.77	91.18	**89.28**	94.48
HPCL	92.53	94.95	94.56	**96.88**	**86.37**	**91.83**	89.14	94.29
CP	-	95.10	-	97.10	-	91.20	-	94.70
MTB	-	91.10	-	95.40	-	84.30	-	91.80
MapRE	-	95.73	-	97.84	-	93.18	-	95.64
HCRP(CP)	94.10	96.42	96.05	97.96	89.13	93.97	93.10	96.46
HPCL(CP)	**96.33**	**96.85**	**97.49**	**98.26**	**93.25**	**94.96**	**95.41**	**96.57**

Compared with the standard baseline methods, HPCL has an average accuracy improvement of at least 11.38%. In the 1-shot setting, the test accuracy of HPCL is 14.27% higher than Proto-BERT; in the 5-shot setting, the accuracy of HPCL is improved by 7.28% compared to Proto-BERT. This result shows that HPCL can effectively model relation prototypes and improve reasoning capabilities in low-resource scenarios by introducing relation information and improving feature representation capabilities.

Compared with other enhancement methods, HPCL relatively outperforms the existing methods. For example, compared with HCRP, HPCL introduces a cross-attention mechanism to fuse global and local prototypes, while HCRP only fuses through splicing. Through this mechanism, HPCL can dynamically aggregate different levels of semantic information, resulting in an average accuracy improvement of at least 1.67%, which verifies the enhancement effect of more fine-grained semantic interactions on relation representation. In addition, HPCL combines a dual contrastive learning strategy, which further enhances the relation discrimination ability of the model compared to MultiRep, which compares local representations (sentence representations) and ignores the contrast constraints between global relation prototypes, resulting in limited ability to distinguish between different relation classes in few-shot samples. In contrast, RPCL and QPCL in HPCL optimize the representation from the relation level and instance level, respectively, which enables the model not only to enhance the discrimination of relation classes in the prototype space but also to strengthen the matching of query samples with the support set of relation prototypes through contrastive learning, thus improving the generalization ability. The accuracy improvement ranges from 0.59% - 2.4% under different task settings. This shows that under few-shot conditions, the ability of prototypes to discriminate between different relation classes can be enhanced by contrastive learning.

Although CP and MapRE rely on large-scale corpora for specialized pre-training, thereby introducing additional prior knowledge, HPCL still achieves highly competitive results without specialized pre-training. Meanwhile, we fine-tuned HPCL on the optimal checkpoints of CP pre-training to obtain HPCL(CP). The experimental results show that HPCL(CP) improves its accuracy by 1.16% - 3.76% compared to CP, further validating its generalization ability under different relation classification tasks.

Table 2. Accuracy (%) of different methods on the FewRel 2.0 test set.

Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
Proto-BERT	40.12	51.50	26.45	36.93
BERT-PAIR	67.41	78.57	54.89	66.85
HCRP	76.34	83.03	63.77	72.94
MTB	74.70	87.90	62.50	81.10
CP	79.70	84.90	68.10	79.80
RelPromptCL	81.24	91.54	69.04	**84.66**
HPCL	**81.63**	**91.77**	**69.75**	83.92

Table 2. demonstrates the effectiveness of HPCL on the cross-domain adaptation tasks in FewRel 2.0. HPCL significantly outperforms the existing methods in all settings, especially in the 10-way-1-shot configuration, which outperforms the current optimal RelPromptCL method by 0.71%. It shows that HPCL has a more substantial

cross-domain transfer capability. However, RelPromptCL still has an advantage in the 10-way-5-shot setup, indicating that prompt-based learning methods may exhibit better adaptability for relation modeling when the support set contains more samples. Notably, HPCL still achieves an average accuracy improvement of 0.15% over RelPromptCL, validating the effectiveness of its cross-attention fusion mechanism and contrastive learning strategy in cross-domain scenarios.

The above analysis results prove the stability and effectiveness of our model. Moreover, the performance increase is mainly due to two aspects: (1) Dynamic fusion of global and local prototypes through the cross-attention mechanism, which enables the model to capture global semantic information and local fine-grained features, thus improving the relation representation ability. (2) Contrastive learning of relation prototypes and contrastive learning of query prototypes makes it easier for the model to distinguish between different relation classes in the reasoning process, thus alleviating the prediction confusion problem.

5.5 Ablation Experiments

Table 3. Ablation experiments under 5-way-1-shot 5-way-5-shot 10 way-1-shot and 10 way-5-shot settings of the FewRel 1.0 validation set.

Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
HPCL	**92.53**	**94.56**	**86.37**	**89.14**
w/o local prototype	90.28	92.17	84.81	87.54
w/o global prototype	88.84	90.93	82.75	86.06
w/o cross attention	91.96	93.84	85.31	88.25
w/o RPCL	92.00	93.92	85.83	88.47
w/o QPCL	92.15	94.14	85.92	88.79

To verify the effectiveness of HPCL, we conducted ablation experiments on the FewRel 1.0 dataset by removing individual parts from the model. Where 'w/o local prototype' removes the local prototype from HPCL, 'w/o global prototype' removes the global prototype, 'w/o cross attention' replaces the cross-attention mechanism with the direct concatenation of global and local prototypes, 'w/o RPCL' disables the relation-prototype contrastive learning module, and 'w/o QPCL' disables the query-prototype contrastive learning module. Table 3. shows the accuracy results for the validation set.

Table 3 illustrates that the removal of any critical element from HPCL results in a reduction in accuracy, indicating that each component plays a role in capturing various aspects of the data. The combination of these components allows the model to acquire more comprehensive and distinctive representations, essential for few-shot relation classification.

It is worth noting that (1) the removal of the global prototype module leads to a maximum drop of 3.69% in 5-way-1-shot accuracy, which constructs a global distribution representation of the category by aggregating high-level semantic features from the samples of the support set, and in few-shot scenarios, the absence of global semantics weakens the ability of the model to generalize to complex semantic structures, leading to a significant increase in sensitivity to local noise. (2) The removal of the local prototype also leads to significant performance degradation, revealing the importance of fine-grained feature identification. The module extracts key fragments from the support set samples through the attention mechanism to construct a prototype representation based on local semantic units. Experiments show that the local prototypes can effectively distinguish relation classes that are globally semantically similar but locally have discriminative differences, emphasizing their essential role in capturing fine-grained relation distinctions. (3) Replacing cross-attention with concatenation leads to a 1.06% accuracy decline in the 10-way-1-shot setting, validating the effectiveness of cross-level interaction. Cross-attention enables dynamic feature fusion by calculating the correlation weights between global and local prototypes. Compared with linear splicing, this nonlinear interaction is more conducive to capturing the complementarity of multi-level features, especially when dealing with long textual relation reasoning, which can effectively alleviate the problem of semantic information dilution. (4) In addition, removing either Relation-Prototype Contrastive Learning (RPCL) or Query-Prototype Contrastive Learning (QPCL) induces slight performance degradation. RPCL enforces contrastive loss constraints between relation descriptions and prototypes to force the model to focus on inter-class discriminative features. QPCL enhances model robustness by constructing contrastive spaces between query samples and prototypes. Experimental results indicate that these mechanisms provide systematic value for optimizing the topological structure of prototype representation spaces.

Overall, these results emphasize that each module in HPCL contributes to the overall effectiveness of the model, with global and local prototypes playing the most critical roles, while the contrastive learning mechanism and cross-attention further enhance the ability of the model to distinguish relations.

6 Conclusion

This paper introduces a novel Hybrid Prototype Contrastive Learning (HPCL) model, which dynamically fuses global and local prototypes using a cross-attention mechanism, thereby significantly improving the representation ability in few-shot relation classification (FSRC) tasks. Additionally, we design relation-prototype contrastive learning (RPCL) and query-prototype contrastive learning (QPCL) strategies to optimize shared intra-class features and strengthen inter-class discriminative power, respectively. Experimental results demonstrate that HPCL exhibits marked superiority in generalization capability and robustness compared to existing baseline methods. Our approach effectively mitigates challenges arising from data scarcity and insufficient relation descriptions and provides a flexible and efficient solution to address semantic bias in relation classification. Future work will extend this methodology to continuous few-shot relation classification tasks to further validate its applicability and scalability.

Acknowledgments. This work was supported by the Funds for Science and technology innovation leading talent of Shaanxi Province, China (No.TZ0336) and the Fundamental Research Funds for the Central Universities, CHD, No.300102244301

References

1. Zhang, Y., Guo, Y.: A Survey of Research on Few-Shot Relation Classification. SSRN 4822235 (2024). https://ssrn.com/abstract=4822235
2. Dai, G., Wang, X., Zou, X., Liu, C., Cen, S.: MRGAT: Multi-Relational Graph Attention Network for knowledge graph completion. Neural Netw. **154**, 234–245 (2022)
3. Kim, J.-J., Lee, D.-G., Wu, J., Jung, H.-G., Lee, S.-W.: Visual question answering based on local-scene-aware referring expression generation. Neural Netw. **139**, 158–167 (2021)
4. Lawrie, D., Yang, E., Oard, D.W., Mayfield, J.: Neural Approaches to Multilingual Information Retrieval. In: Kamps, J., Goeuriot, L., Crestani, F., Maistro, M., Joho, H., Davis, B., Gurrin, C., Kruschwitz, U., Caputo, A. (eds.) Advances in Information Retrieval, pp. 521–536. Springer, Cham (2023)
5. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP, pp. 1003–1011. ACL, Singapore (2009)
6. Ye, Q., Liu, L., Zhang, M., Ren, X.: Looking Beyond Label Noise: Shifted Label Distribution Matters in Distantly Supervised Relation Extraction. In: Proceedings of EMNLP-IJCNLP 2019, pp. 3841–3850. ACL, Hong Kong (2019)
7. Zhang, N., Deng, S., Sun, Z., Wang, G., Chen, X., Zhang, W., Chen, H.: Long-tail Relation Extraction via Knowledge Graph Embeddings and Graph Convolution Networks. In: Proceedings of NAACL-HLT 2019, pp. 3016–3025. ACL, Minneapolis (2019)
8. Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., Liu, Z., Sun, M.: FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation. In: Proceedings of EMNLP 2018, pp. 4803–4809. ACL, Brussels (2018)
9. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of ICML 2017, pp. 1126–1135. JMLR.org (2017)
10. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Proceedings of NIPS 2017, pp. 4080–4090. Curran Associates, Red Hook (2017)
11. Yang, K., Zheng, N., Dai, X., He, L., Huang, S., Chen, J.: Enhance prototypical network with text descriptions for few-shot relation classification. In: Proceedings of CIKM 2020, pp. 2273–2276. ACM (2020)
12. Liu, Y., Hu, J., Wan, X., Chang, T.-H.: Learn from Relation Information: Towards Prototype Representation Rectification for Few-Shot Relation Extraction. In: Findings of NAACL 2022, pp. 1822–1831. ACL, Seattle (2022)
13. Wu, H., He, Y., Chen, Y., Bai, Y., Shi, X.: Improving few-shot relation extraction through semantics-guided learning. Neural Netw. **169**, 453–461 (2024)
14. Wang, Y., Verspoor, K., Baldwin, T.: Learning from unlabelled data for clinical semantic textual similarity. In: Proceedings of ClinicalNLP 2020, pp. 227–233. ACL (2020)
15. Liu, Y., Hu, J., Wan, X., Chang, T.-H.: A Simple yet Effective Relation Information Guided Approach for Few-Shot Relation Extraction. In: Findings of ACL 2022, pp. 757–763. ACL, Dublin (2022)
16. Han, J., Cheng, B., Lu, W.: Exploring task difficulty for few-shot relation extraction. In: Proceedings of EMNLP 2021, pp. 2605–2616. ACL (2021)
17. Borchert, P., De Weerdt, J., Moens, M.-F.: Efficient Information Extraction in Few-Shot Relation Classification through Contrastive Representation Learning. In: Proceedings of NAACL 2024 (Short Papers), pp. 638–646. ACL (2024)
18. Dong, Y., Yang, R., Liu, J., Qin, X.: Few-shot relation extraction through prompt with relation information and multi-level contrastive learning. IEEE Access 12, 123352–123361 (2024). https://doi.org/10.1109/ACCESS.2024.3452720

19. Peng, H., Gao, T., Han, X., Lin, Y., Li, P., Liu, Z., Sun, M., Zhou, J.: Learning from Context or Names? An Empirical Study on Neural Relation Extraction. In: Proceedings of EMNLP 2020, pp. 3661–3672. ACL (2020)
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Proceedings of NIPS 2017, pp. 6000–6010. Curran Associates (2017)
21. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., Wierstra, D.: Matching networks for one shot learning. In: Proceedings of NIPS 2016, pp. 3630–3638. Curran Associates, Barcelona (2016)
22. Soares, L.B., FitzGerald, N., Ling, J., Kwiatkowski, T.: Matching the Blanks: Distributional Similarity for Relation Learning. In: Proceedings of ACL 2019, pp. 2895–2905. ACL, Florence (2019)
23. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of NAACL-HLT 2019, pp. 4171–4186. ACL, Minneapolis (2019)
24. Satorras, V.G., Estrach, J.B.: Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043 (2018)
25. Gao, T., Han, X., Zhu, H., Liu, Z., Li, P., Sun, M., Zhou, J.: FewRel 2.0: Towards More Challenging Few-Shot Relation Classification. In: Proceedings of EMNLP-IJCNLP 2019, pp. 6250–6255. ACL, Hong Kong (2019)
26. Qu, M., Gao, T., Xhonneux, L.P.A.C., Tang, J.: Few-shot relation extraction via bayesian meta-learning on relation graphs. In: Proceedings of ICML 2020, pp. 7867–7876. PMLR (2020)
27. Dong, M., Pan, C., Luo, Z.: MapRE: An effective semantic mapping approach for low-resource relation extraction. In: Proceedings of EMNLP 2021, pp. 2694–2704. ACL (2021)