

Article

Interaction Information Guided Prototype Representation Rectification for Few-Shot Relation Extraction

Xiaoqin Ma ^{1,2}, Xizhong Qin ^{1,2,*}, Junbao Liu ^{1,2} and Wensheng Ran ³¹ College of Information Science and Engineering, Xinjiang University, Urumqi 830049, China² Xinjiang Key Laboratory of Signal Detection and Processing, Urumqi 830049, China³ Xinjiang Uygur Autonomous Region Product Quality Supervision and Inspection Institute, Urumqi 830049, China

* Correspondence: qinxz@xju.edu.cn

Abstract: Few-shot relation extraction aims to identify and extract semantic relations between entity pairs using only a small number of annotated instances. Many recently proposed prototype-based methods have shown excellent performance. However, existing prototype-based methods ignore the hidden inter-instance interaction information between the support and query sets, leading to unreliable prototypes. In addition, the current optimization of the prototypical network only relies on cross-entropy loss. It is only concerned with the accuracy of the predicted probability for the correct label, ignoring the differences of other non-correct labels, which cannot account for relation discretization in semantic space. This paper proposes an attentional network of interaction information to obtain a more reliable relation prototype. Firstly, an inter-instance interaction information attention module is designed to mitigate prototype unreliability through interaction information between the support set and query set instances, utilizing category information hidden in the query set. Secondly, the similarity scalar, which is defined by the mixed features of the prototype and the relation and is added to the focal loss to improve the attention of hard examples. We conducted extensive experiments on two standard datasets and demonstrated the validity of our proposed model.

Keywords: few-shot relation extraction; prototypical network; relation prototype

Citation: Ma, X.; Qin, X.; Liu, J.; Ran, W. Interaction Information Guided Prototype Representation Rectification for Few-Shot Relation Extraction. *Electronics* **2023**, *12*, 2912. <https://doi.org/10.3390/electronics12132912>

Academic Editor: Arkaitz Zubiaga

Received: 1 June 2023

Revised: 29 June 2023

Accepted: 30 June 2023

Published: 3 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Relation extraction (RE) is an essential information extraction component in natural language processing [1]. Its purpose is to extract semantic relations among entities from natural language texts. The structured data captured by RE can support downstream tasks such as knowledge graph construction [2], machine reading comprehension [3], and question-answering systems [4]. With the development of deep learning, neural models have been widely applied to relation extraction and have achieved significant results. However, the performance of these models relies heavily on a large amount of high-quality annotated data, whose obtainment is often time-consuming and human intensive. To alleviate this issue, inspired by the ability of humans to acquire new knowledge from a few instances through prior knowledge, few-shot learning was proposed and quickly became a viable approach for accomplishing various tasks [5–7].

The main objective of the few-shot relation extraction task is to utilize a small number of support instances for learning the features of relation classes and use these features to determine the relations of query instances [8]. The details are shown in Table 1. At present, the prototypical network is the most popular algorithm for few-shot relation extraction [9]. Its main idea is that each relation class has a prototype representation in an embedding space. Specifically, it is designed to extract the mean of the samples in each class as the prototype in the embedding space by nonlinearly mapping the input into the embedding space. All query instances are classified using the nearest distant rule to find the class prototype closest to the prototype.

Table 1. The 3-way-2-shot task is used for few-shot relation extraction. In this task, the head and tail entities are represented in red and blue, respectively. The relationship categories in the training and testing phases are non-overlapping.

Data	Relation	Instances
Support set	R1: country of citizenship	Instance 1: Charles Gniette was a Belgian field hockey player who competed in the 1920 Summer Olympics. Instance 2: Catherine Loyola (born 1986) is a fashion model and beauty queen from the Philippines.
	R2: mother	Instance 1: Ariston had three other children by Perictione: Glaucon, Adeimantus, and Potone. Instance 2: Dylan and Caitlin brought up their three children, Aeronwy, Llewellyn, and Colm.
	R3: developer	Instance 1: In the mid-1980s Microsoft developed a multitasking version of DOS. Instance 2: The expansion uses Valve Corporation’s Steam to download and install updates.
Query set	R1, R2, or R3	Rugby League Live2 followed in 2012, again developed by Big Ant Studios.

In order to improve the performance of prototype-based few-shot relation extraction tasks based on the prototypical network, existing works mainly focus on the following three aspects. The first category is introducing external information to compensate for the lack of information. REGRAB [10] proposes a Bayesian meta-learning approach that effectively learns the posterior distribution of class prototypes by constructing a global relation graph from Wikidata. ConceptFERE [11] suggests a novel approach for enhancing few-shot relation extraction. This method leverages the inherent concepts of entities to provide information for relation prediction, thereby improving the performance of relation classification. In [12], TD-proto provides auxiliary supporting evidence for relation classification by utilizing textual descriptions of relations and entities to enhance the prototype network. The second category is to improve model structure or optimize training strategy so that the model can learn better prototypes, i.e., intra-class similarity and inter-class dissimilarity. EGNN-Proto [13] develops a neural graph network into the framework of the prototypical network, allowing the meta-learned feature embeddings to adapt to new tasks quickly. A new Interactive Attention Network [14] (IAN) is introduced that leverages the interaction information of inter-instance and intra-instance to classify relations. Considering both inter-class and intra-class distances simultaneously, PROTOTYPICAL-RELATION NETS [15] proposes a novel prototype relation network. The third category involves improving models through contrastive learning or pretraining, aiming to make similar instances closer and dissimilar instances farther apart. CP [16] introduces a contrastive pretraining framework for relation extraction to enhance the capability of capturing entity types and extracting relevant facts from the context. For mitigating the inductive bias of source classes and forcing the network to learn more distinctive information, InfoPatch [17] introduces an advanced contrastive training scheme to mitigate the inductive bias of source classes, forcing the network to learn more distinctive information. PCL [18] proposes a prototypical contrastive learning approach that addresses the fundamental limitations of instance-wise contrastive learning.

However, there are two limitations in existing works based on prototypical network. Firstly, prototypical network-based methods often rely on averaging the representations of support instances within each class to construct class prototypes while neglecting the interaction information between support and query instances. Therefore, the obtained prototypes are unreliable [19]. Secondly, the current optimization of the prototypical network only relies on the cross-entropy loss, focusing merely on the accuracy of predicted

probability for the correct labels, while ignoring the differences among other non-correct labels. For this reason, the learned features are scattered [20].

To tackle the mentioned issues, we develop an FSRE model called QAFPR, which integrates interaction information attention and adaptive focal loss. Specifically, our model leverages the rich interaction information between support and query instances and obtains more accurate relation prototype representations through an interaction information attention module. Furthermore, we introduce adaptive focal loss with dynamic task-level weights, enabling the model to learn how to focus on hard examples adaptively [18], while allowing the model to concentrate better on intra-class semantic relations. By adopting this approach, our model achieves more reliable prototype representations without using excessive parameters. The contributions of this paper are primarily reflected in the following aspects:

1. In prototypical networks, interaction information extracts support set prototypes through instance-level attention, mitigates biases arising from fewer instances, and generates more reliable prototypes.
2. We employ an adaptive focal loss as the loss function, assigning varying weights to each task to prioritize hard examples.
3. We conducted experiments on two benchmark datasets, FewRel1.0 [21] and FewRel2.0 [22]. The experimental results show that our model achieves competitive performance compared to existing baselines.

The rest of this article is as follows: In Section 2, we described the related work, including methods and techniques in relation extraction, few-shot relation extraction, and prototype-based few-shot relation extraction. In Section 3, we describe our proposed model framework in detail, including the application of the interaction information attention module and adaptive focal loss. In Section 4, our experimental setup and the results of the experiment are given. In Section 5, we analyze the experimental results and demonstrate the performance and advantages of our model. Finally, we conclude with a summary of our work and give an outlook for the future.

2. Related Work

In this section, we briefly review relation extraction, few-shot relation extraction, and prototype-based few-shot relation extraction methods.

2.1. Relation Extraction

Relation extraction is an essential task in natural language processing, which identifies predefined relations between two target entities in each utterance and provides the basis for constructing structured knowledge (e.g., knowledge graphs) [23]. The current mainstream deep learning models used for this task heavily rely on a large amount of supervised data, leading to the model's generalizability being dependent on the quantity and quality of the labelled data [24–26]. Although regularization techniques are widely used to mitigate overfitting in deep learning models, they do not provide additional supervised information for the model. Therefore, when the amount of labelled data is insufficient, simply applying regularization may not effectively solve the generalization problem [27]. To address the issue of limited training data and reduce the manual annotation cost, Mintz et al. [28] first proposed using distant supervision for automatic data annotation. This approach assumes that “if two entities have a known relation label in a knowledge base, then sentences containing these two entities should express a similar relation to some extent”. Heuristically, the assumption aligns the target entities in the sentences with entities in the knowledge base for labeling the relations automatically. However, this assumption also raises some issues: (1) It is true that using the distant-supervised method can generate noisy data because the same entity pair may imply different semantic relations in different sentences [29]. (2) The knowledge base of many domains still needs to be completed (e.g., food safety domain). In addition, most relations exist in a long-tailed distribution. The data

available for training via assumption is insufficient [30]. In order to solve the problems mentioned above, few-shot learning was applied and born [31].

2.2. Few-Shot Relation Extraction

Few-shot learning enables models to learn with only a few training samples and exhibits good generalization capability. Although existing few-shot learning methods are predominantly developed in computer vision [32–34], their success has inspired researchers to explore the application of few-shot learning in natural language processing. Few-shot relation extraction is designed to predict new relations from several tagged instances. Han et al. [21] introduced few-shot learning for the first time into relation extraction and created the FewRel1.0 dataset. They also experimented with several typical few-shot learning methods, laying the foundation for further exploration by subsequent researchers.

Research on FSRE predominantly focuses on two standard algorithms: optimization-based [35] and metric-based approaches [9]. Finn et al. [36] tested MetaNet, a meta-learning approach that learns meta-level knowledge across tasks and enables rapid parameterization for generalization to new tasks. Munkhdalai et al. [37] proposed a model-agnostic meta-learning method for training initialization parameters, making the model achieve optimal performance after a few gradient steps. Inspired by the MAML approach, Dong et al. [38] established a connection between instance-based information and semantic-based information to attain more effective initialization and faster adaptation.

The metric-based approach finds the neighbouring category to determine the classification result of a sample to be classified by calculating the distance between the sample and a known classified sample. Gao et al. [39] used hybrid attention to increase the diversity of few-shot tasks and enhance robustness in noisy samples. Considering local-level and instance-level matching information, Ye et al. [40] encoded the interaction between query instances and each support set interactively. Xie et al. [41] reduced the interference of noisy samples on the model by employing a heterogeneous graph network combined with adversarial training. Han et al. [42] proposed a two-stage approach for supervised contrastive learning and an instance-level prototypical network, such that semantically related relational representations are close to each other, and other representations are far away. Gao et al. [43] introduced a novel guiding method that learns the similarity between instances from existing relations and their labelled data by utilizing a Siamese network.

2.3. Prototypical Networks for Few-Shot Relation Extraction

Parameter-based optimization methods have shown relatively poor performance compared to metric-based methods. Therefore, most researchers have focused their research on metric-based learning methods, such as Siamese networks [44], matching networks [1], graph neural networks [23], prototypical networks [9], and so on. Among them, the prototypical network has become the dominant method for FSRE tasks due to its high efficiency.

To enhance the expressive power of the semantic space, Sun et al. [45] propose a hierarchical attention network for few-shot text classification. Fan et al. [46] employed a large-margin prototype network with fine-grained features which are supposed to generalize to long-tail relations effectively. Wen et al. [47] integrated the transformer into the prototypical network to achieve better relational level feature extraction. Hui et al. [48] applied a context attention-based prototypical network that designs context attention to highlight critical instances within the support set, aiming to generate promising prototypes. Wang et al. [49] used two mechanisms to decouple easily-confused relations. Wang and Zheng et al. [50] proposed a Discriminative Rule-based Knowledge method that alleviates word overlap confusion through rule-based incorporation of ontological knowledge graphs (KG). Yu et al. [51] proposed a novel multi-prototype embedding network model to jointly extract the composition of relational triples. Utilizing this method re-initializes the memory network by using prototypes of all observed relations in the current learning phase, Cui et al. [52] proposed a continual relation extraction model based on relation prototypes which can alleviate the problem of catastrophic forgetting. Xiao et al. [53] explored a novel

method that was based on label words and joint representation learning. This approach effectively leverages the information from relation labels to learn improved representations. Zhang and Lu [54] proposed a novel method called Label Prompt Dropout, which randomly drops label descriptions during learning. Li et al. [55] employed a joint training approach to learning prototype encoders from relation definitions.

However, the studies mentioned above tend to construct class prototypes by averaging the representations of support instances for each class and ignoring the interaction between support and query instances. Thus, the studies fail to obtain reliable relation prototypes [16]. Additionally, the optimization methods of the prototypical network ignore the intra-class compactness and inter-class separability in the semantic space of relations [14]. Therefore, we propose a new prototypical-network-based model (QAFPR) for few-shot relation extraction tasks. First, inspired by the hidden category information in unlabeled query instances [56], we design an interaction information attention module that incorporates query information and support set information to match the support features and effectively reduces the deviation between the obtained and expected relation prototypes. Secondly, we use mixed features of rectification prototypes and relation information to calculate a similarity matrix. Then, we obtain similarity scalars from the matrix and add them into the focal loss, allowing the model to pay more attention to hard examples. Our method effectively captures consistent features between query and support information, enabling the matching of support features to obtain more reliable relation prototypes.

3. Methodology

The prototypical network is designed to learn a representation space for classification by computing the distance to each class prototype. However, the obtained relation prototypes may have biases. We propose a new prototypical network-based model for few-shot relation extraction to obtain more reliable relation prototypes. With the help of unlabeled instances in the query set, each relation prototype is obtained through the interaction information attention between query information and support information, fused with the original prototype rather than the centroid of the support instances for that relation. Additionally, we employ an adaptive focal loss to encourage the prototype network to learn reliable instances among different relations, leading to close intra-class relations.

In this section, we will provide a detailed description of the details of the method we propose. As shown in Figure 1, we introduce our proposed model framework consisting of the following modules: (1) A sentence encoding layer, using the pre-training model BERT to encode each instance as a support set, a query set, and relational information as low-dimensional embedding. (2) The relational representation layer takes the token embedding of a given relational representation. It fuses the relational representations of the two views by adding them directly to obtain an embedding of the same dimension as the entity embedding. (3) An interaction information module that takes the word embeddings of a given support instance and query instance and fuses them into an attention weight by measuring the semantic correlation vectors of the support and query instances and uses that attention weight to better correct the prototype. (4) A prototype fusion layer that will fuse the rectification prototype embedding and the final embedding of the relational representation, given the embedding of the rectification prototype and the final embedding of the relational representation for each class, into a final prototype. (5) The adaptive focal loss improves attention to complex tasks through a mixture of prototypes and relational features for different tasks.

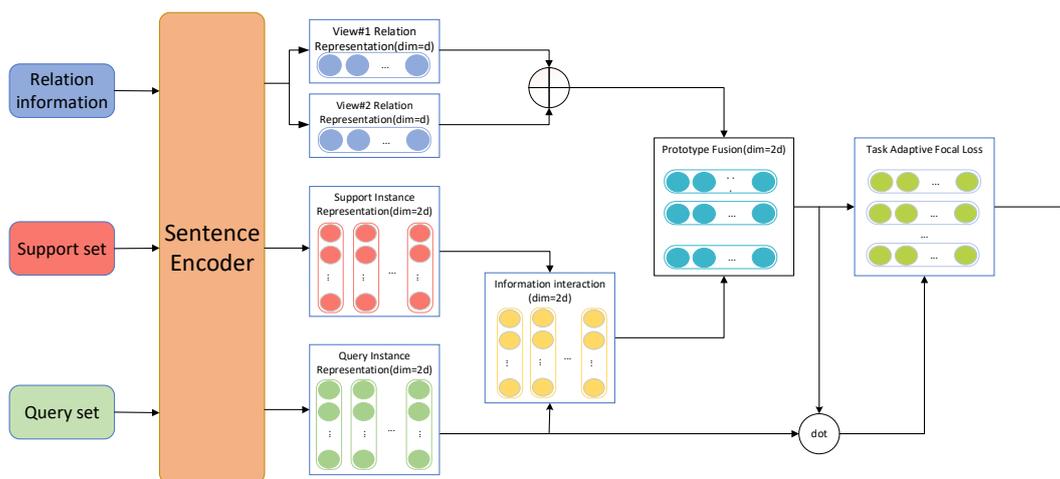


Figure 1. The framework of our proposed QAFPR model. The input instances, sentence encoder, interaction information attention module, prototype fusion module, and adaptive focal loss module are shown from left to right. \oplus indicates direct addition. Blue denotes relational information, red denotes support set instances, and green denotes query set instances.

3.1. Task Definition

Given a relation set, R , and a support set, S , the purpose of the FSRE model is to classify the entity pair (h, t) mentioned in the query instance q into the most likely relation $r (r \in R)$. S is defined as:

$$S = \left\{ (x_1^1, h_1^1, t_1^1, r_1), \dots, (x_1^{n_1}, h_1^{n_1}, t_1^{n_1}, r_1), \dots, (x_m^1, h_m^1, t_m^1, r_m), \dots, (x_m^{n_m}, h_m^{n_m}, t_m^{n_m}, r_m) \right\}, (r_1, r_2, \dots, r_m) \in R \quad (1)$$

where $(x_i^j, h_i^j, t_i^j, r_i)$ means that the entity pair (h_i^j, t_i^j) mentioned in the instance x_i^j expresses the semantics of the relation r_i . x_i^j contains all the entities of the sum, and each instance is composed of a word sequence $\{w_1, w_2, \dots\}$. In a few-shot learning scenario, a relation extraction model must learn features from a support set, S , and predict the relation r , for a given query instance, q . In this task, only a tiny number of relation instances are typically available. The N-way-K-shot approach is a commonly used few-shot learning method. N represents the number of related categories, and K represents the number of instances used for learning per category.

3.2. Sentence Encoder

This module extracts feature representations of the support set, S , and the query set, Q . We use a BERT [57] as the encoder, which allows us to capture the semantic information of the support set and the query set. For instance, in the support and query sets, the middle state is obtained by connecting the hidden states corresponding to the start token mentioned by the two entities. i.e.,

$$h_I = [h_{entity1}; h_{entity2}], h_{entity1}, h_{entity2} \in \mathbb{R}^d \quad (2)$$

where h_I denotes the representation of the relation between instance, I , and the two given entities $h_{entity1}, h_{entity2}$, and d is the hidden size.

For details, only a few words in the context are relevant for relation extraction, and most words introduce a significant amount of noise. To eliminate the influence of context on entities, we decided to separate the CLS token and the entities from the sentence and only use the entities for feature extraction. We use BERT-base for entity feature extraction, as shown in Figure 2. Given a sentence $S = \{w_1, w_2, \dots, w_n\}$ and entity pair, entity1 and entity2, where w_i represents the i -th word in the sentence, we use the sentence as input

to the encoder to obtain sentence features with positional information that captures the contextual interaction around the entity:

$$output = [w_1, w_2, \dots, entity1, \dots, entity2, \dots, w_n] \tag{3}$$

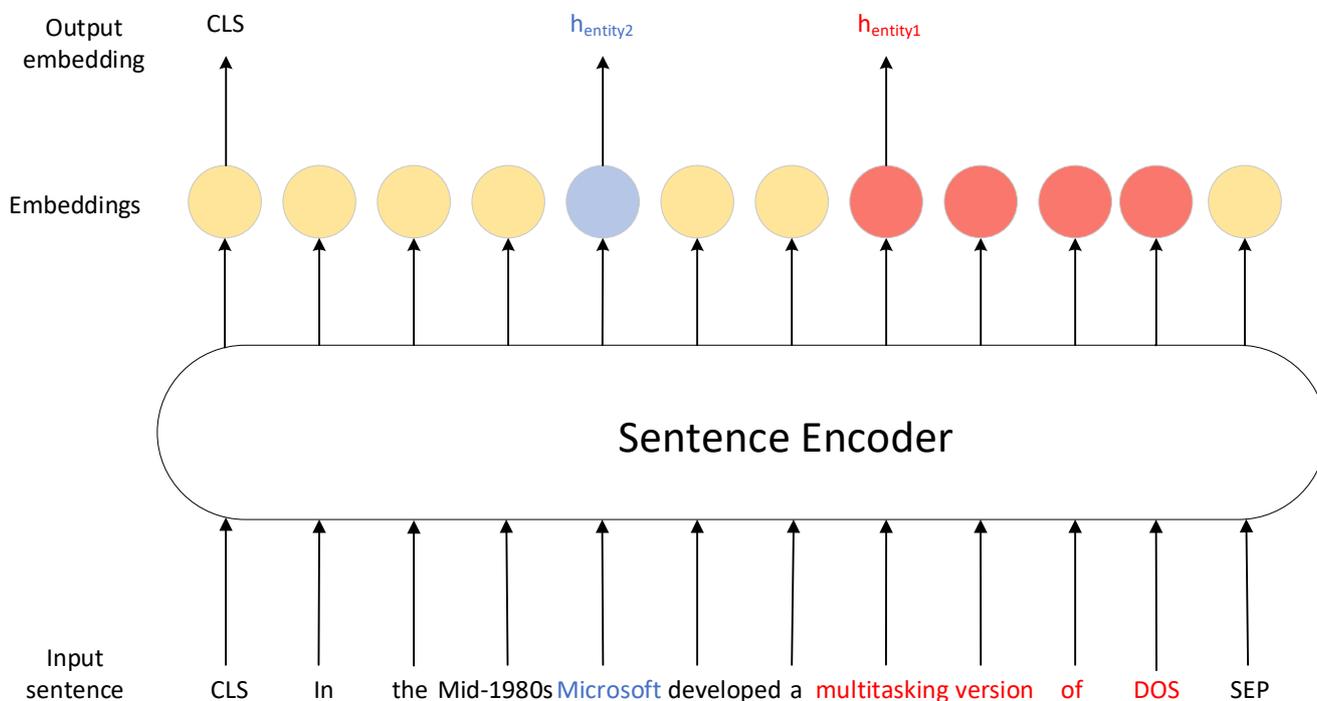


Figure 2. The feature encoder we adopted. Yellow represents markers, red represents the head entity, and blue represents the tail entity. CLS and SEP tokens indicate the beginning and end of the input sentence, which consists of l_x words.

Then, we extract the entity based on its position information in the sentence.

We concatenate the name and description for each relation, then input the sequence into BERT. We extract the embeddings of the $[cls]$ token, i.e., $\{\mathcal{R}_i^{view1} \in \mathbb{R}^d, i = 1, 2, \dots, n\}$, which represents the entire sequence, and the average of all token embeddings, i.e., $\{\mathcal{R}_i^{view2} \in \mathbb{R}^d, i = 1, 2, \dots, n\}$, to represent the relation from two different views.

3.3. Relation Representation

As described in Section 4.1, the relation is represented as $\mathcal{R}_i^{view1}, \mathcal{R}_i^{view2} \in \mathbb{R}^d$. We combine \mathcal{R}_i^{view1} and \mathcal{R}_i^{view2} using the concatenation operation \oplus , which enables the relation representation to have the same dimensional features as the relation prototype. This ensures that the relation representation has the same dimensional features as the relation prototype, as shown in the following equation:

$$\mathcal{R}^{final} = \mathcal{R}^{view1} \oplus \mathcal{R}^{view2} \tag{4}$$

where \mathcal{R}^{final} represents the final information representation and $\mathcal{R}^{final} \in \mathbb{R}^{2d}$.

3.4. Interaction Information Attention Module

To generate relation representations, we will send all instances in the support set into an instance encoder to obtain the relation representation for each instance, as described

in Section 3.2. The average value of the relation represented belonging to the relation i is defined as the relation prototype:

$$p_i = \frac{1}{K} \sum_{k=1}^K s_k^i \tag{5}$$

where p_i represents the relational prototype of relation i , K represents the number of instances for the relation i , and s_k^i denotes the embedding of the k -th supporting instance of relation i .

However, due to the limited number of instances in the support set, there is a deviation between the obtained and expected relation prototypes. Additionally, the weighted average of the class prototype overlooks the category information hidden in the query set and the interaction between instances in the support set and the query set. As a result, the obtained prototype is unreliable, as shown in Figure 3a. Therefore, to utilize more instance information and generate reliable relation prototypes, we propose an interaction information attention module to rectify the prototype and fully exploit the sample resources of few-shot learning. As shown in Figure 3b, this module aims to leverage the semantic correlation between support instances and query instances to aid in calculating more reliable prototype representations. Specifically, the rectified prototype can be represented as follows:

$$P_i = \sum_{k=1}^K \alpha_k^i s_k^i \tag{6}$$

where P_i is the prototype of relation i based on interaction information attention. s_k^i represents the embedding of the k -th support instance for relation i . α_k^i is the weight that represents the semantic relevance between the K support instances and the query instance for relation i . The weights are determined by:

$$\alpha_k^i = \text{softmax} \left(-\frac{1}{|Q|} \sum_{j=1}^{|Q|} d(s_k^i, q_j) \right) \tag{7}$$

where s_k^i represents the embedding of the k -th support instance for relation i . q_j is the embedding of the j -th query instance in the query set. $|Q|$ is the total number of instances in Q and $d(\cdot)$ represents the Euclidean distance.

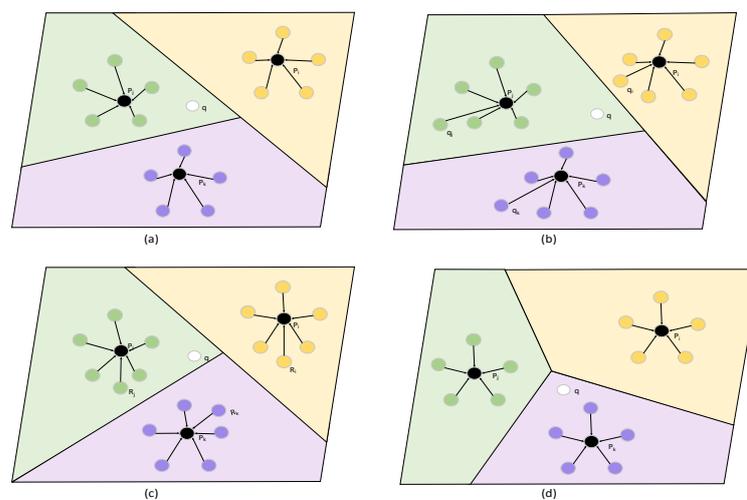


Figure 3. The diagram illustrates the prototype rectification mechanism proposed. Let us assume there exists a query instance belonging to the class. (a) The initial prototype closest to the query sample q . (b) Considers the interaction between the query information and the support set. (c) Obtains the final prototype through a fusion mechanism. (d) The prototype closest to the query is obtained with the updated positional distribution, resulting in the correct classification.

3.5. Prototype Fusion

Inspired by the fusion method proposed by Liu [19] for integrating relational information and instance prototypes, we adopt a direct addition fusion mechanism, as shown in Figure 3c. We obtain the final prototype by directly adding the relational information and the rectified prototype:

$$P = p_i + \mathcal{R}^{final} \tag{8}$$

where P represents the final prototype representation and $P \in \mathbb{R}^{2d}$.

The model calculates the distance between each relation prototype, P , and the query instance, Q , using the vector dot product and selects the shortest distance relation class as the final prediction.

3.6. Adaptive Focal Loss

Typically, the training objective of metric-based few-shot relation extraction models is to minimize the cross-entropy loss, which is the negative log-likelihood of the true labels:

$$\mathcal{L}_{ce} = -\log(p_t) \tag{9}$$

$$p_t = P(T = t|Q, S) \tag{10}$$

where T represents the class label, and t represents the true label. Cross-entropy loss only focuses on the accuracy of predicting the probability for the correct label while ignoring the differences among other non-correct labels. As a result, it cannot ensure intra-class compactness and inter-class separability. Lin et al. [58] proposed focal loss to mitigate the imbalance of hard examples and easy examples:

$$\mathcal{L}_F = -(1 - p_t)^\gamma \log(p_t) \tag{11}$$

where $\gamma \geq 0$ adjusts the rate of weights under easy instances. For an easy instance, p_t is almost 1 and the factor becomes 0.

Adaptive focal loss is an improved version of focal loss, as shown in Figure 4. It introduces dynamic task-level weights to pay more attention to hard examples. Specifically, we estimate the weights from different classes for an N-way-K-shot task by calculating the similarity between categories. We concatenate the mixed features of prototypes and relations to represent each class $c^i = [r_h^i, p_h^i]$, and then we define a task similarity matrix $S^\tau \in \mathbb{R}^{M \times M}$, for $i, j \in \{1, \dots, M\}$,

$$S_{ij}^\tau = \frac{c^i \cdot c^j}{\|c^i\| \times \|c^j\|} \tag{12}$$

where $\|\cdot\|$ represents the Euclidean norm. The scalar task similarity is obtained from the following equation:

$$S^\tau = softmax(\|S^\tau\|_F) \tag{13}$$

where $\|\cdot\|_F$ represent the Frobenius norm, and τ corresponds to the number of tasks in the mini-batch. Task-level scalars are added to the focal loss, which emphasizes the importance of each instance-level task and assigns different weights to each task. Formally, the adaptive focal loss is defined as follows,

$$\mathcal{L}_{TF} = -S^\tau(1 - p_t)^\gamma \log(p_t) \tag{14}$$

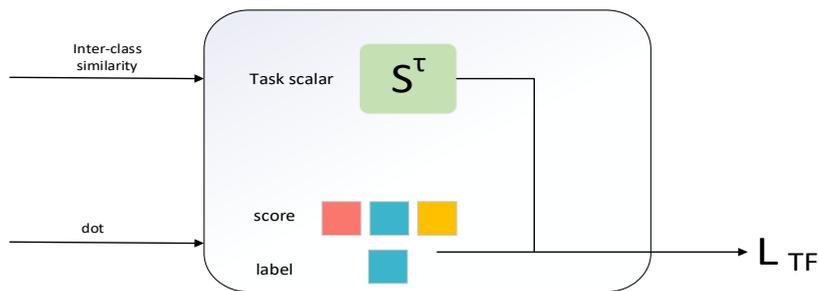


Figure 4. Adaptive focal loss.

4. Experiment

In this section, we describe the experimental setup, first presenting the dataset and parameter settings and the model implementation results.

4.1. Datasets

For each task, we evaluate our model on two benchmark datasets, and the details are shown in Table 2.

Table 2. Dataset details.

Dataset	Source	Apply	Relation Number	Instance Number
FewRel1.0	Wiki	Train	64	44,800
		Val	16	11,200
		Test	20	14,000
FewRel2.0	Wiki PubMed	Train	64	44,800
		Val	10	7000
		Test	15	10,500

FewRel1.0 [21] The dataset comprises 100 relations, with each relation containing 700 instances. Our experimental setup adheres to the official benchmark split, which involves allocating 64 classes for training, 16 classes for validation, and 20 new classes for testing. Note that the training/validation/test set relations do not overlap. Due to the unknown labels of the FewRel1.0 test set, we submit our model’s prediction results from CodaLab to obtain the accuracy of the test set.

FewRel2.0 [22] Since all FewRel1.0 data are from Wikipedia, i.e., contained in the same domain, we evaluate our model on FewRel2.0, which considers cross-domain issues. The training set of FewRel2.0, like FewRel1.0, is sourced from Wikipedia and consists of 64 relations. We use the PubMed subset of FewRel2.0 for validation and testing, including 10 and 15 relations sourced from biomedical literature databases. As a result, the training, validation, and testing sets are in different domains.

4.2. Implementation Details

We used BERT-base-uncased and CP [16] as the sentence encoder. The BERT-base model consists of a 12-layer transformer module with the CP having the same structure but further post-trained by comparative learning. In the training process, we set the training iterations to 30,000 and the validation iterations to 1000. The batch size was set to 4. The AdamW optimizer was used to reduce losses. We manually adjusted the hyperparameters according to the performance of the validation data, as shown in Table 3. We utilized the same set of hyperparameter values for both datasets, except for the learning rate (LR). For Fewrel1.0, we concatenated the name and description of each relation as input, and the LR was 1×10^{-5} . For Fewrel2.0, we entered only the name of the relation, with the LR

adjusted to 3×10^{-6} . Adhering to the official evaluation settings, we used N-way-K-shot to measure the model's performance on the validation and test sets. We performed five different random experiments for all datasets to reduce randomness and reported the average performance as the final experimental result. All our experiments were conducted on a computer with an Intel Core i9 13900K/F CPU@5.8 GHz and a GeForce RTX 3090 GPU card with 24 GB of VRAM.

Table 3. Hyperparameter settings.

Component	Parameter	Value
BERT/CP	Type	Base-uncased
	Hidden size	768
	Max length	128
Training	Learning rate	$1 \times 10^{-5}/3 \times 10^{-6}$
	Batch size	4
	Max iterations	30,000
Loss	γ	1

4.3. Compared Methods

To assess the validity of QAFPR, we compared it with the following baseline methods.

Proto-HATT [39]: a prototypical network to obtain more accurate relational prototypes using attentional mechanisms.

MLMAN [40]: a prototypical network to obtain more accurate relational prototypes using matching methods.

BERT-PAIR [22]: uses BERT as the instance encoder and classifies each query instance based on the distance to the relation prototype.

Prototype-BERT [21]: a primary prototypical network, instance embedding using the BERT model.

REGRAB [10]: a Bayesian meta-learning acquisition method with external global relation graphs to study the relations between entity pairs.

TD-Proto [12]: an importance distribution of generic content words learned through memory networks.

CTEG [49]: a model which is equipped with two mechanisms to learn to decouple these confusing relations.

ConceptFERE [11]: proposes a novel approach for enhancing few-shot relation extraction by incorporating inherent entity concepts.

HCRP [59]: a contrastive learning method which is proposed to learn better representations using relation-labelled information.

SimpleFSRE [19]: embedding relational descriptions directly into the prototype representation.

MTB [26]: a post-training task called matching the blank which was designed using contrast learning.

CP [16]: a post-training comparison framework for entity masking which is proposed using external information from knowledge graphs for contrastive learning.

To provide a fair comparison with existing BERT-based and different baselines for post-training tasks, we provide BERT-based and CP-based results for our proposed model, respectively.

4.4. Overall Results

To assess the validity of our model, we compared performance with a strong baseline model on FewRel1.0 and FewRel2.0 in the 5-way and 10-way settings. The results of the implementation of the FewRel1.0 are shown in Table 4, and the results of the FewRel2.0 are shown in Figure 5. Based on the experimental results on both datasets, the performance improvement of our model is more significant when there are fewer support instances.

Specifically, we divided the results of FewRel1.0 into two parts based on the encoder type: CNN-based and BERT-based methods. The BERT-based approach is divided into two parts: the first is to use the original BERT, and Proto-BERT represents the original baseline of our model. The second approach uses an additional contrast learning CP on BERT to obtain a better contextual representation. Our model has used BERT and CP as encoders for our model for apparent comparisons.

Table 4. Experimental results on the FewRel1.0 validation/test set, where N-w-K-s is an abbreviation for N-way-K-shot. Please note that the results of comparison models are taken from the paper, or the results reported by CodaLab. Bold is the highest result, underlined is the second highest. * Indicates the initial baseline.

Encoder	Model	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s	
CNN	Proto_HATT	72.65/74.52	86.15/88.40	60.13/62.38	76.20/80.45	
	MLMAN	75.01/___	87.09/90.12	62.48/___	77.50/83.05	
BERT	BERT_PAIR	85.66/88.32	89.48/93.22	76.84/80.63	81.76/87.02	
	Proto_BERT *	84.77/89.33	89.54/94.13	76.85/83.41	83.42/90.25	
	REGRAB	87.95/90.30	92.54/94.25	80.26/84.09	86.72/89.93	
	TD-Proto	___/84.76	___/92.38	___/74.32	___/85.92	
	CTEG	84.72/88.11	92.52/95.25	76.01/81.29	84.89/91.33	
	ConceptFERE	___/89.21	___/90.34	___75.72	___/81.82	
	HCRP	90.90/93.76	93.22/95.66	84.11/89.95	87.79/92.10	
	SimpleFSRE	91.29/94.42	94.05/96.37	86.09/90.73	89.68/93.47	
	QAFPR	92.26/94.95	94.56/96.98	87.45/91.58	89.02/93.72	
	MTB	___/91.10	___/95.40	___/84.30	___/91.80	
	CP	___/95.10	___/97.10	___/91.20	___/94.70	
	HCRP(CP)	94.10/96.42	96.05/97.96	89.13/93.97	93.10/96.46	
	SimpleFSRE(CP)	96.21/96.63	97.07/97.93	93.38/94.94	95.11/96.39	
	QAFPR(CP)	96.01/97.30	97.72/98.24	93.01/95.13	95.57/96.51	
		Δ	+5.62	+2.76	+8.17	+3.47
		Δ (CP)	+2.20	+1.14	+3.93	+1.81

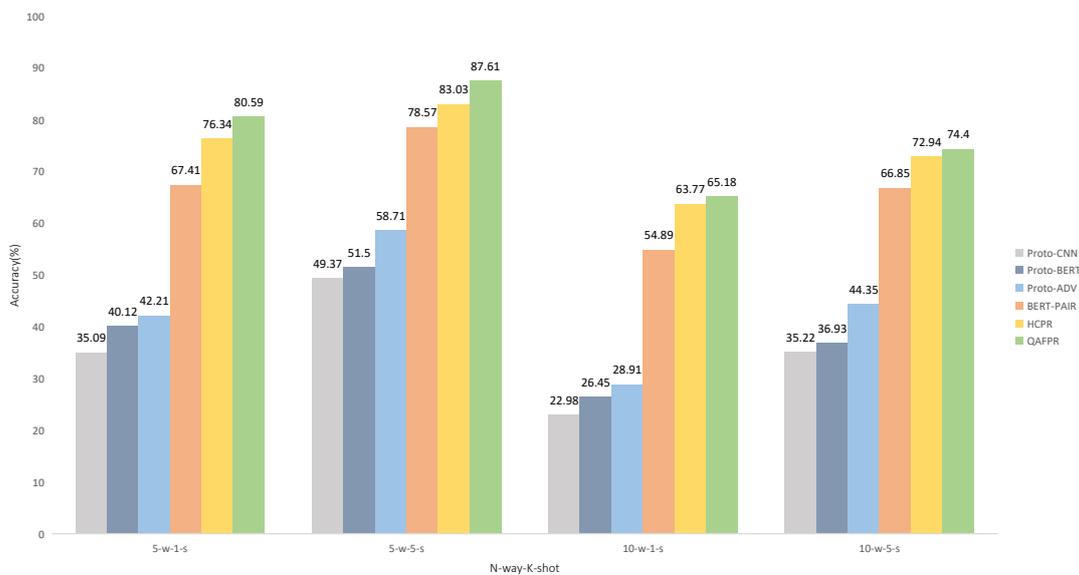


Figure 5. Experimental results on the FewRel2.0 domain adaptation test set, where N-w-K-s is an abbreviation for N-way-K-shot. The results of the comparison model are taken from the paper or from the results reported by CodaLab.

From Table 4, we can obtain three conclusions:

1. Our method QAFPR (BERT) outperforms the state-of-the-art when using BERT as the backend model, as shown in the first part of the BERT-based model in Table 4.
2. When QAFPR (CP) adopts CP as the backend model, our method shows improvements compared to the state-of-the-art SimpleFSRE (CP) method, indirectly reflecting that our method is more suitable for few-shot scenarios. Additionally, compared to the base model CP, our model achieves improvements of 2.20%, 1.14%, 3.93%, and 1.81% in the four N-way-K-shot settings, respectively.
3. Compared with the baseline model Proto-BERT, our model shows improvements of 5.62%, 2.76%, 8.17%, and 3.47% in the four N-way-K-shot settings, as shown in the last two rows of Table 4. These observation results validate the effectiveness of our proposed method.

Our method also performs better on FewRel2.0, as illustrated in Figure 5, demonstrating our model's stability and validity. Performance gains come from three main aspects: (1) Utilizing the interaction information of the instances to mine the valuable information in the instances for obtaining a more accurate prototype of the relation. (2) The prototype fusion module uses relational information to further obtain reliable prototypes. (3) A task adaptive focal loss module that learns different weights for different tasks, noting the different hard tasks.

5. Analysis

In this section, we analyze the experimental results in the following areas: (1) Ablation experiments for interaction information and adaptive focal loss, aiming to evaluate the accuracy of the obtained prototype. (2) Compared to the current CP-based HCRP and SimpleFSRE, our evaluation shows that our model performs better under the same encoder. (3) The visualization case study aims at visually evaluating the effectiveness of our model.

5.1. Ablation Study

We perform an ablation study on BERT-based 5-way-1-shot, 5-way-5-shot, 10-way-1-shot, and 10-way-5-shot with the test set to validate the effectiveness of the proposed interaction information attention module, adaptive focal loss, and prototype fusion module (abbreviated as QPR, FPR, and RPR, respectively). We consider three ablation experiments, including QPR, FPR, and RPR, as shown in Table 5.

Table 5. Results of Ablation study on FewRel1.0 test set (%).

Model	5-1	5-5	10-1	10-5
QAFPR	94.95	96.98	91.58	93.72
QPR	94.81	96.81	91.35	93.60
FPR	94.16	96.63	89.48	92.98
RPR	89.57	95.66	84.08	92.30

From Table 5, we can observe the following conclusions. Firstly, in this variant of QPR, we replace the adaptive focal loss with cross-entropy loss to validate the effectiveness of the interaction information attention module. Experimental results show performance degradation in all four settings. Next, in the variant of FPR, we remove the interaction information module to validate the effectiveness of the adaptive focal loss. We found that the performance suffers a loss of 0.79% to 2.10%, particularly in the 10-way-1-shot scenario where the performance loss is 2.10%. Thirdly, in the variant of RPR, we remove the interaction information module and replace the adaptive focal loss with cross-entropy loss. The performance drops sharply under the 5-way-1-shot and 10-way-1-shot settings, indicating that the prototype fusion module performs poorly under low-resource conditions.

5.2. Comparison with HCRP and SimpleFSRE

We compare our method with CP, HCRP, and SimpleFSRE on the test set, as shown in Figure 6. From the graph, it can be observed that our method based on the CP encoder shows improvement in all four N-way-K-shot settings, with more significant improvement under 5-way-1-shot and 5-way-5-shot settings. These results validate that our model can also achieve good performance across different encoders.

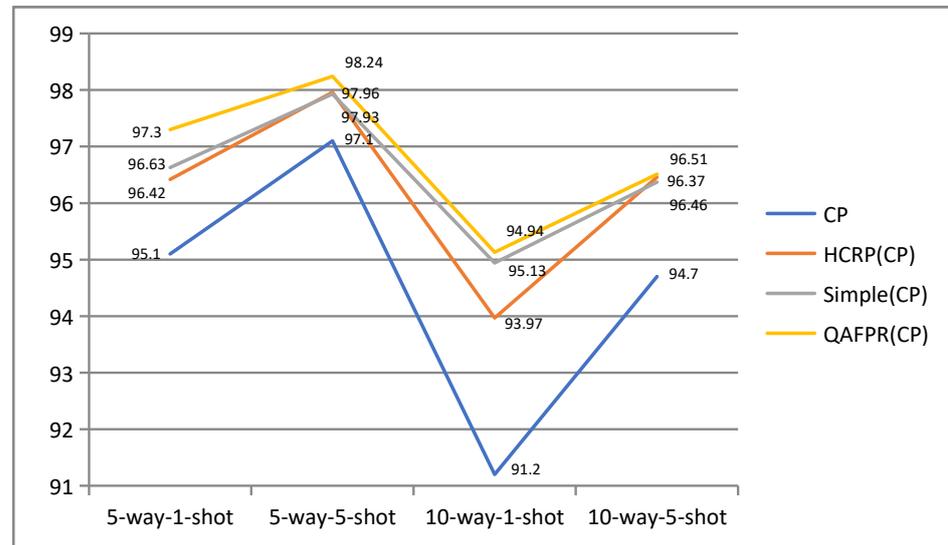


Figure 6. Comparison with HCRP and SimpleFERS on the test set.

5.3. Case Study

We conducted a case study on the 5-way-1-shot task, where the validation set contains instances of five types of relations, namely “constellation,” “main subject,” “mother,” “child,” and “spouse,” to demonstrate the performance of our model. We randomly selected 500 instances that are from the validation set of FewRel1.0 for each relation. This case study compares the representation spaces of SimpleFSRE and QAFPR mappings with the same inputs, and the results are analyzed in two ways:

1. We visualized the representation space of the embedding using T-SNE to intuitively describe the relational representation of the validation instances as shown in Figure 7. Figure 7a shows the mapping result when no interaction information is added, and Figure 7b shows the mapping result when interaction information is added. It is clear from the graphs that the relations “mother” and “child” are of better quality than those generated by SimpleFSRE, with less overlap among classes and more distinct boundaries. The relation representation of “member of” of QAFPR’s aggregates in one space compared to SimpleFSRE’s.
2. We visualize the effect of the hard examples in SimpleFSRE and QAFPR as shown in Figure 8. Figure 8a shows the effect of SimpleFSRE, and Figure 8b shows the effect of QAFPR, with darker colors indicating higher classification accuracy. Our model is significantly more accurate than SimpleFSRE in classifying the classes “P25: mother” and “P40: child”, verifying the effectiveness of adaptive focal loss.

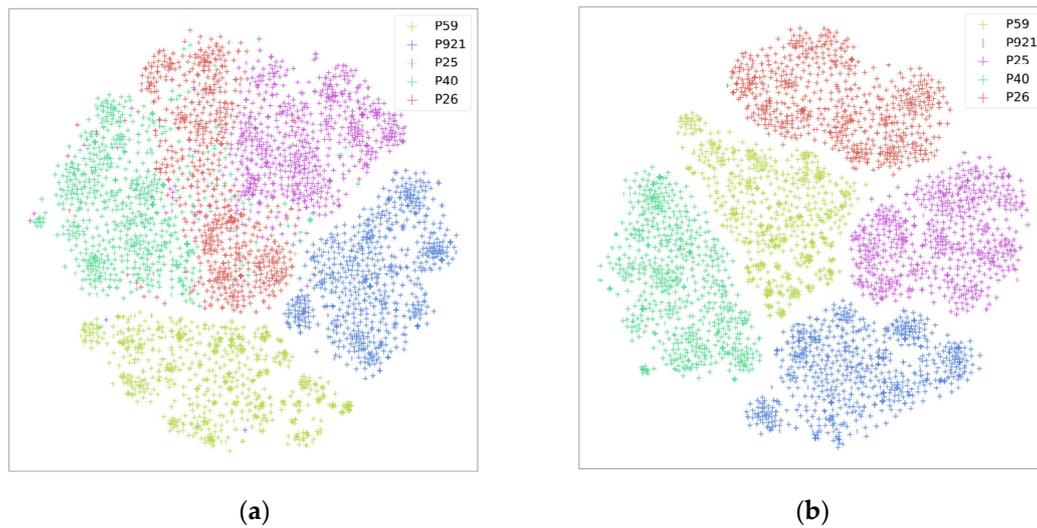


Figure 7. T-SNE plots of instance embeddings with and without (w/o) interaction information. A total of 5 relations were sampled out of 500 samples (P59: constellation, P921: main subject, P25: mother, P40: child, and P26: spouse). (a) shows the results of SimpleFSRE and (b) shows the results of QAFPR.

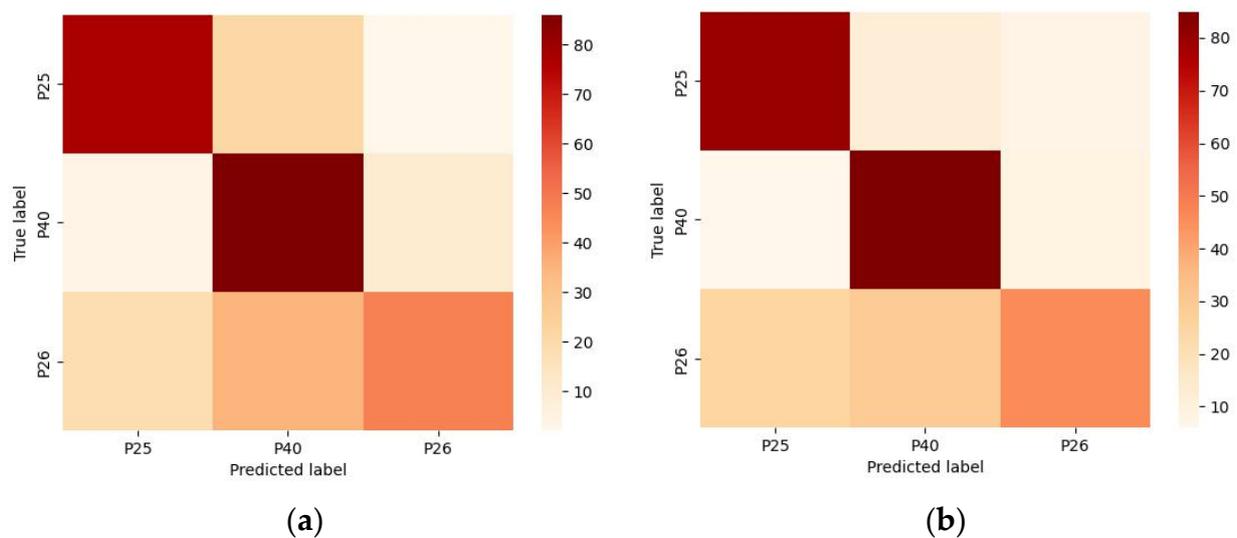


Figure 8. Example of a 3-way-1-shot hard task. (a) shows the accuracy of SimpleFSRE and (b) shows the accuracy of QAFPR.

6. Conclusions and Outlook

This study focuses on few-shot relation extraction, which involves identifying semantic relation between entity pairs using only a small number of annotated instances. Existing prototype-based methods have limitations in generating prototypes and handling challenging tasks. To address these issues, we propose an attention network called QAFPR, which consists of an interaction information attention module and an adaptive focal loss. To obtain more reliable relation prototypes, we utilize category information hidden in the query set and correct the prototypes through interaction among instances. To make the model more attentive to hard tasks, we introduce dynamic task-level weights using an adaptive focal loss that treats the corrected prototypes and similarity information of relation as hyperparameters, enabling the model to learn how to focus on hard tasks adaptively.

In the future, we will continue to explore tasks related to new-shot relation extraction. On the one hand, we plan to explore further cross-domain few-shot scenarios; on the other

hand, we would like to introduce few-shot scenarios into realistic scenarios where training and validation samples are scarce.

Author Contributions: Conceptualization and methodology, X.M.; software, X.M.; validation, X.M., X.Q. and J.L.; formal analysis, X.M.; investigation, X.M.; resources, X.M.; data curation, X.Q.; writing—original draft preparation, X.M.; writing—review and editing, X.Q. and W.R.; visualization, J.L.; supervision, W.R.; project administration, X.Q.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Major science and technology special projects of Xinjiang Uygur Autonomous Region (2020A03001) and its sub-program Key technology development and application demonstration of integrated food data supervision platform in Xinjiang region (2020A03001-2).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, X.; Gao, T.; Lin, Y.; Peng, H.; Yang, Y.; Xiao, C.; Liu, Z.; Li, P.; Zhou, J.; Sun, M. 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, Suzhou, China, 4–7 December 2020; pp. 745–758.
2. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [[CrossRef](#)] [[PubMed](#)]
3. Kambhatla, N.; Born, L.; Sarkar, A. CipherDAug: Ciphertext based Data Augmentation for Neural Machine Translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 201–218.
4. Zhang, L.; Lin, C.; Zhou, D.; He, Y.; Zhang, M. A Bayesian end-to-end model with estimated uncertainties for simple question answering over knowledge bases. *Comput. Speech Lang.* **2021**, *66*, 101167. [[CrossRef](#)]
5. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [[CrossRef](#)]
6. Zhang, X.; Qiang, Y.; Sung, F.; Yang, Y.; Hospedales, T.M. RelationNet2: Deep comparison columns for few-shot learning. *arXiv* **2018**, arXiv:1811.07100.
7. Chen, W.Y.; Liu, Y.C.; Kira, Z.; Wang, Y.C.F.; Huang, J.B. A closer look at few-shot classification. *arXiv* **2019**, arXiv:1904.04232.
8. Sabo, O.; Elazar, Y.; Goldberg, Y.; Dagan, I. Revisiting few-shot relation classification: Evaluation data and classification schemes. *Trans. Assoc. Comput. Linguist.* **2021**, *9*, 691–706. [[CrossRef](#)]
9. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
10. Meng, G.; Tianyu, G.; Louis-Pascal, X.; Tang, J. Few-shot Relation Extraction via Bayesian Meta-learning on Relation Graphs. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual Event, 13–18 July 2020; pp. 7867–7876.
11. Yang, S.; Zhang, Y.; Niu, G.; Zhao, Q.; Pu, S. Entity Concept-enhanced Few-shot Relation Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Online, 1–6 August 2021; pp. 987–991.
12. 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 the 29th ACM International Conference on Information & Knowledge Management (CIKM'20), New York, NY, USA, 19–23 October 2020; pp. 2273–2276.
13. Lyu, C.; Liu, W.; Wang, P. Few-shot text classification with edge-labeling graph neural network-based prototypical network. In Proceedings of the 28th International Conference on Computational Linguistics, Online, 8–13 December 2020; pp. 5547–5552.
14. Han, Y.; Qiao, L.; Zheng, J.; Kan, Z.; Feng, L.; Gao, Y.; Tang, Y.; Zhai, Q.; Li, D.; Liao, X. Multi-view interaction learning for few-shot relation classification. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Online, 1–5 November 2021; pp. 649–658.
15. Liu, X.; Zhou, F.; Liu, J.; Jiang, L. Meta-learning based prototype-relation network for few-shot classification. *Neurocomputing* **2020**, *383*, 224–234. [[CrossRef](#)]
16. Peng, H.; Gao, T.; Han, X.; Lin, Y.; Li, P.; Liu, Z.; Sun, M.; Zhou, J. Learning from Context or Names? In An Empirical Study on Neural Relation Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 19–20 November 2020; pp. 3661–3672.
17. Liu, C.; Fu, Y.; Xu, C.; Yang, S.; Li, J.; Wang, C.; Zhang, L. Learning a few-shot embedding model with contrastive learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 8635–8643.
18. Li, J.; Zhou, P.; Xiong, C.; Hoi, S.C. Prototypical contrastive learning of unsupervised representations. *arXiv* **2020**, arXiv:2005.04966.

19. 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 the Association for Computational Linguistics: ACL 2022*; Association for Computational Linguistics: Dublin, Ireland, 2022; pp. 757–763.
20. Ren, H.; Cai, Y.; Lau, R.Y.K.; Leung, H.-F.; Li, Q. Granularity-aware area prototypical network with bimargin loss for few shot relation classification. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 4852–4866. [[CrossRef](#)]
21. 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 the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 31 October–4 November 2018; pp. 4803–4809.
22. 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 the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 3–7 November 2019; pp. 6250–6255.
23. Yu, T.; He, S.; Song, Y.Z.; Xiang, T. Hybrid graph neural networks for few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Online, 22 February–1 March 2022; Volume 36, pp. 3179–3187.
24. Geng, Z.; Chen, G.; Han, Y.; Lu, G.; Li, F. Semantic relation extraction using sequential and tree-structured LSTM with attention. *Inf. Sci.* **2020**, *509*, 183–192. [[CrossRef](#)]
25. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany, 7–12 August 2016; pp. 207–212.
26. Soares, L.B.; FitzGerald, N.; Ling, J.; Kwiatkowski, T. Matching the blanks: Distributional similarity for relation learning. *arXiv* **2019**, arXiv:1906.03158.
27. Puspitaningrum, D. Improving Performance of Relation Extraction Algorithm via Leveled Adversarial PCNN and Database Expansion. In *Proceedings of the 2019 7th International Conference on Cyber and IT Service Management (CITSM)*, Jakarta, Indonesia, 6–8 November 2019; Volume 7, pp. 1–6.
28. Mintz, M.; Bills, S.; Snow, R.; Jurafsky, D. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore, 2–7 August 2009; pp. 1003–1011.
29. Zhu, H.; Lin, Y.; Liu, Z.; Fu, J.; Chua, T.-S.; Sun, M. Graph neural networks with generated parameters for relation extraction. *arXiv* **2019**, arXiv:1902.00756.
30. Kang, B.; Xie, S.; Rohrbach, M.; Yan, Z.; Gordo, A.; Feng, J.; Kalantidis, Y. Decoupling representation and classifier for long-tailed recognition. *arXiv* **2019**, arXiv:1910.09217.
31. Song, Y.; Wang, T.; Cai, P.; Mondal, S.K.; Sahoo, J.P. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. *ACM Comput. Surv.* **2023**, accepted. [[CrossRef](#)]
32. He, J.; Hong, R.; Liu, X.; Xu, M.; Zha, Z.J.; Wang, M. Memory-augmented relation network for few-shot learning. In *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, WA, USA, 12–16 October 2020; pp. 1236–1244.
33. Bai, J.; Yuan, A.; Xiao, Z.; Zhou, H.; Wang, D.; Jiang, H.; Jiao, L. Class incremental learning with few-shots based on linear programming for hyperspectral image classification. *IEEE Trans. Cybern.* **2020**, *52*, 5474–5485. [[CrossRef](#)] [[PubMed](#)]
34. Lu, J.; Jin, S.; Liang, J.; Zhang, C. Robust few-shot learning for user-provided data. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 1433–1447. [[CrossRef](#)] [[PubMed](#)]
35. Vuorio, R.; Sun, S.H.; Hu, H.; Lim, J.J. Multimodal model-agnostic meta-learning via task-aware modulation. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.
36. Munkhdalai, T.; Yu, H. Meta networks. In *Proceedings of the International Conference on Machine Learning*, Sydney, Australia, 6–11 August 2017; pp. 2554–2563.
37. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 1126–1135.
38. Dong, B.; Yao, Y.; Xie, R.; Gao, T.; Han, X.; Liu, Z.; Lin, F.; Lin, L.; Sun, M. Meta-information guided meta-learning for few-shot relation classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, Online, 8–13 December 2020; pp. 1594–1605.
39. Gao, T.; Han, X.; Liu, Z.; Sun, M. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'19/IAAI'19/EAAI'19)*, Honolulu, HI, USA, 27 January–1 February 2019; pp. 6407–6414.
40. Ye, Z.-X.; Ling, Z.-H. Multi-Level Matching and Aggregation Network for Few-Shot Relation Classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 28 July–2 August 2019; pp. 2872–2881.
41. Xie, Y.; Xu, H.; Li, J.; Yang, C.; Gao, K. Heterogeneous graph neural networks for noisy few-shot relation classification. *Knowl. Based Syst.* **2020**, *194*, 105548. [[CrossRef](#)]
42. Han, J.; Cheng, B.; Nan, G. Learning Discriminative and Unbiased Representations for Few-Shot Relation Extraction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM'21)*, Queensland, Australia, 1–5 November 2021; pp. 638–648. [[CrossRef](#)]

43. Gao, T.; Han, X.; Xie, R.; Liu, Z.; Lin, F.; Lin, L.; Sun, M. Neural snowball for few-shot relation learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 7772–7779.
44. Chen, X.; He, K. Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 15750–15758.
45. Sun, S.; Sun, Q.; Zhou, K.; Lv, T. Hierarchical attention prototypical networks for few-shot text classification. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 476–485.
46. Fan, M.; Bai, Y.; Sun, M.; Li, P. Large margin prototypical network for few-shot relation classification with fine-grained features. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2353–2356.
47. Wen, W.; Liu, Y.; Ouyang, C.; Lin, Q.; Chung, T. Enhanced prototypical network for few-shot relation extraction. *Inf. Process. Manag.* **2021**, *58*, 102596. [[CrossRef](#)]
48. Hui, B.; Liu, L.; Chen, J.; Zhou, X.; Nian, Y. Few-shot relation classification by context attention-based prototypical networks with BERT. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 118. [[CrossRef](#)]
49. Wang, Y.; Bao, J.; Liu, G.; Wu, Y.; He, X.; Zhou, B.; Zhao, T. Learning to decouple relations: Few-shot relation classification with entity-guided attention and confusion-aware training. *arXiv* **2020**, arXiv:2010.10894.
50. Wang, M.; Zheng, J.; Cai, F.; Shao, T.; Chen, H. DRK: Discriminative Rule-based Knowledge for Relieving Prediction Confusions in Few-shot Relation Extraction. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 2129–2140.
51. Yu, H.; Zhang, N.; Deng, S.; Ye, H.; Zhang, W.; Chen, H. Bridging Text and Knowledge with Multi-Prototype Embedding for Few-Shot Relational Triple Extraction. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 6399–6410.
52. Cui, L.; Yang, D.; Yu, J.; Hu, C.; Cheng, J.; Yi, J.; Xiao, Y. Refining sample embeddings with relation prototypes to enhance continual relation extraction. 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), Virtual Event, 1–6 August 2021; pp. 232–243.
53. Xiao, Y.; Jin, Y.; Hao, K. Adaptive prototypical networks with label words and joint representation learning for few-shot relation classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 1406–1417. [[CrossRef](#)] [[PubMed](#)]
54. Zhang, P.; Lu, W. Better Few-Shot Relation Extraction with Label Prompt Dropout. *arXiv* **2022**, arXiv:2210.13733.
55. Zhen, L.; Zhang, Y.; Nie, J.-Y.; Li, D. Improving Few-Shot Relation Classification by Prototypical Representation Learning with Definition Text. In *Findings of the Association for Computational Linguistics: NAACL 2022*; Association for Computational Linguistics: Seattle, WA, USA, 2022; pp. 454–464.
56. Zhang, Y.; Cen, M.; Wu, T.; Zhang, H. RAPS: A Novel Few-Shot Relation Extraction Pipeline with Query-Information Guided Attention and Adaptive Prototype Fusion. *arXiv* **2022**, arXiv:2210.08242.
57. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
58. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *99*, 2999–3007.
59. Han, J.; Cheng, B.; Lu, W. Exploring Task Difficulty for Few-Shot Relation Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana, Dominican Republic, 7–11 November 2021; pp. 2605–2616.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.