



Adaptive Prototype Network with Common and Discriminative Representation Learning for Few-Shot Relation Extraction

Wenyue Hu¹, Jiang Zhong¹(✉) , Yang Xia², Yangmei Zhou³,
and Rongzhen Li¹

¹ Chongqing University, Chongqing, China

{huwenyue, zhongjiang, lirongzhen}@cqu.edu.cn

² Chongqing Academy of Big Data Co., Ltd., Chongqing, China

³ Chongqing Academy of Science and Technology, Chongqing, China

Abstract. The task of few-shot relation extraction presents a significant challenge as it requires predicting the potential relationship between two entities based on textual data using only a limited number of labeled examples for training. Recently, quite a few studies have proposed to handle this task with task-agnostic and task-specific weights, among which prototype networks have proven to achieve the best performance. However, these methods often suffer from overfitting novel relations because every task is treated equally. In this paper, we propose a novel methodology for prototype representation learning in task-adaptive scenarios, which builds on two interactive features: 1) common features are used to rectify the biased representation and obtain the relative class-centered prototype as much as possible, and 2) discriminative features help the model better distinguish similar relations by the representation learning of the entity pairs and instances. We obtain the hybrid prototype representation by combining common and discriminative features to enhance the adaptability and recognizability of few-shot relation extraction. Experimental results on FewRel dataset, under various few-shot settings, showcase the improved accuracy and generalization capabilities of our model.

Keywords: Relation extraction · Few-shot learning · Adaptive prototype network

1 Introduction

Relation extraction plays a vital role within the domain of natural language processing by identifying and extracting the relationships between different entities mentioned in textual data. It has significant applications in various downstream tasks, including text mining [25], knowledge graph [18] and social networking [21]. This problem is commonly approached as a supervised classification

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

X. Yang et al. (Eds.): ADMA 2023, LNAI 14179, pp. 63–77, 2023.

https://doi.org/10.1007/978-3-031-46674-8_5

task. However, the process of labeling large datasets of sentences for relation extraction is not only time-consuming but also expensive. This limitation often leads to a scarcity of labeled data available for training relation extraction models, hindering the progress and evolution of this task. To address this challenge, researchers have delved into a new field of study called few-shot relation extraction (FSRE). FSRE is viewed a model aligned with human learning and is considered the most promising method for tackling this task. Existing studies [6, 19, 22, 23] have achieved remarkable results and surpassed human levels in general tests. Some researchers [2, 11], however, observe that most proposed models are ineffective in real applications and are looking into specific reasons such as task difficulty. Most task-agnostic and task-specific models [16] actually struggle to handle the FSRE task adaptively with varying difficulties. Figure 1 depicts a process of FSRE augmented by the relation instances. Due to the high similarity of relationships and the limited number of training instances, the existing methods are hard to identify the relation *mother* between entity pair (*Isabella Jagiellon*, *Bona Sforza*) for the query instance.

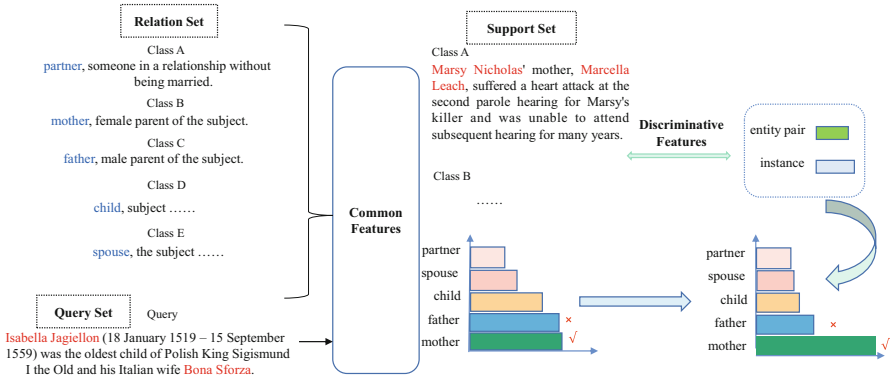


Fig. 1. A novel 5-way 1-shot relation extraction pipeline. First, use the relation set (left) and then enhance the relation representation with the support set (right). Finally, self-adapt to obtain true value for similar relationships in difficult tasks.

The FSRE task is commonly tackled through two prominent approaches: meta-learning based methods and metric-learning based methods [15, 32]. Meta-learning, with its emphasis on the acquisition of learning abilities, focuses on training a task-agnostic model with various subtasks, enabling it to learn how to predict scores for novel classes. In contrast, metric-based approaches aim to discover an improved metric for quantifying distribution discrepancies and effectively discerning different categories. Among them, the prototype network [11, 27] is widely used as a metric-based approach and has demonstrated remarkable performance in FSRE scenarios. Compared to meta-learning based methods, metric-based methods in FSRE task usually focus on training task-specific models that have poorer adaptability to new tasks. Conversely, the meta-learning

based methods may struggle with the discriminative features of relations between entity pairs during the training process, leading to underwhelming performance. This can be attributed primarily to the scarcity of training instances for each subtask in FSRE scenarios. As shown in Fig. 1, when only using the common features from relation instances as the metric index, it is difficult to distinguish the pair (*IsabellaJagiellon*, *BonaSforza*) in the query sentence from the five similar relations. While the discriminative features of the support instances are introduced, the features have more obvious distinctions, and the score of the *mother* is higher than others. In order to address these challenges in few-shot relation extraction, we present a novel adaptive prototype network representation that incorporates both relation-meta features and various instance features, called AdaProto. Firstly, we introduce relation-meta learning to obtain the common features of the relations so that the model does not excessively deviate from its relational classes, thus making the model have better adaptability. Besides, we propose entity pair rectification and instance representation learning to gain discriminative features of the task, thus increasing the discriminability of the relations. Finally, we employ a multi-task training strategy to achieve a model of superior quality.

We can summarize our main contributions into three distinct aspects: **1)** We present a relation-meta learning approach that enables the extraction of common features from relation instances, resulting in enhanced adaptability of the model. **2)** Our model combines common features with discriminative features learned from the support set to enhance the recognizability of relational classification. **3)** We assess the performance of our proposed model on the extensively utilized datasets FewRel through experiments employing the multi-task training strategy. The experimental results confirm the model’s excellent performance.

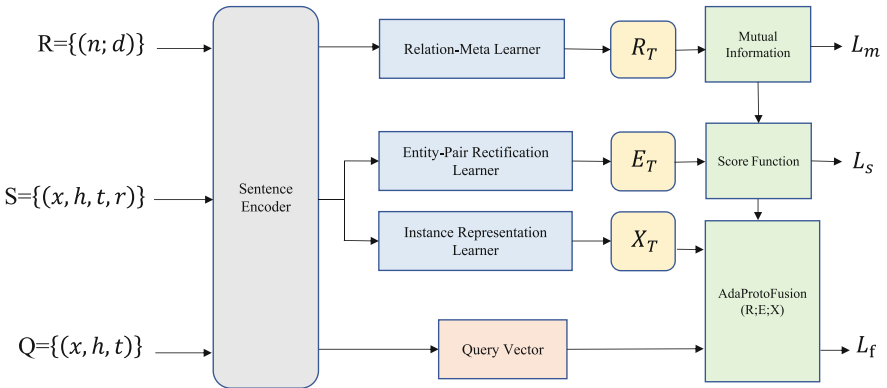


Fig. 2. The overall framework of AdaProto. The relation-meta learner exploits relation embeddings and relation meta to transfer the relative common features. To capture the discriminative features, E_t and X_t represent the entity pair rectification and instance embeddings, respectively. Finally, multi-task training adaptively fuses the above two features into different tasks.

2 Related Work

2.1 Few-Shot Relation Extraction

Relation extraction, a fundamental task within the domain of natural language processing (NLP), encompassing a wide range of applications, including but not limited to text mining [25] and knowledge graph [18]. Its primary objective is to discern the latent associations among entities explicitly mentioned within sentences. Few-shot relation extraction (FSRE), as a specialized approach, delves into the prediction of novel relations through the utilization of models trained on a restricted set of annotated examples. Han et al. [13] introduce a comprehensive benchmark called FewRel, which serves as a large-scale evaluation platform specifically designed for FSRE. Expanding upon this initiative, Gao et al. [9] presented FewRel 2.0, an extended iteration of the dataset that introduces real-world challenges, including domain adaptation and none-of-the-above detection. These augmentations render FewRel 2.0 a more veracious and efficacious milieu for the comprehensive evaluation of FSRE techniques.

Meta-learning based and metric-learning based methods are the two predominant approaches widely employed in solving the FSRE task. Meta-learning, with its emphasis on the acquisition of learning abilities, strives to educate a task-agnostic model can acquire knowledge and making predictions for novel classes through exposure to diverse subtasks. The model-agnostic meta-learning (MAML) algorithm was introduced by Finn et al. [7]. This algorithm is an adaptable technique applicable to a broad spectrum of learning scenarios by leveraging gradient descent-based training. Dong et al. [5] introduced a meta-information guided meta-learning framework that utilizes semantic notions of classes to enhance the initialization and adaptation stages of meta-learning, resulting in improved performance and robustness across various tasks and domains. Lee et al. [15] proposed a domain-agnostic meta-learning algorithm that specifically targets the challenging problem of cross-domain few-shot classification. This innovative algorithm acts as an optimization strategy aimed at improving the generalization abilities of learning models. However, the meta-learning based approaches can be hindered by the lack of sufficient training instances for each subtask, making it vulnerable to the discriminative features of the relationships between entity pairs. As a result, this limitation can negatively impact the performance of the model on the FSRE task.

The primary objective of metric-based approaches is to identify a more effective metric for measuring the discrepancy in distribution, allowing for better differentiation between different categories. Among them, the prototypical network [29] is widely used as one of metric-based approaches to few-shot learning and has demonstrated its efficacy in addressing the FSRE task. In this approach, every example is encoded as a vector by the feature extractor. The prototype representation for each relation is obtained by taking the average of all the exemplar vectors belonging to that relation. To classify a query example, its representation is compared to the prototypical representations of the candidate relations, and it is assigned a class based on the nearest neighbor rule. There have been many

previous works that enhanced the model performance of different problems in the FSRE task based on prototypical networks. Ye and Ling [36] introduced a modified version of the prototypical network that incorporates multi-level matching and aggregation techniques. Gao et al. [8] proposed a novel approach called hybrid attention-based prototypical networks. This innovative approach aims to mitigate the impact of noise and improve the overall performance of the system. Yang et al. [34] proposed an enhancement to the prototypical network by incorporating relation and entity descriptions. And Yang et al. [35] introduced the inherent notion of entities to offer supplementary cues for relationship extraction, consequently augmenting the overall effectiveness of relationship prediction. Ren et al. [26] introduced a two-stage prototype network that incorporates prototype attention alignment and triple loss, aiming to enhance the performance of their model in complex classification tasks. Han et al. [11] devised an innovative method that leverages supervised contrastive learning and task adaptive focal loss, specifically focusing on hard tasks in FSRE scenario. Brody et al. [2] conducted a comprehensive analysis of the effectiveness of robust few-shot models and investigated the reliance on entity type information in FSRE models. Despite these notable progressions, the task adaptability of FSRE model is still under-explored. In this study, we propose an adaptive prototype network representation with relation-meta features and various instance features, which enhances the adaptability of the model and yields superior performance on relationships with high similarity.

2.2 Adaptability in Few-Shot Learning

Few-shot learning (FSL) is a prominent approach aimed at training models to comprehend and identify new categories using a limited amount of labeled data. However, the inherent limitations of the available information from a restricted number of samples in the N -way- K -shot setting of FSL pose a significant challenge for both task-agnostic and task-specific models to adaptively handle the FSL task. Consequently, the performance of FSL models can exhibit substantial variations across different environments, particularly when employing metrics-based approaches that typically train the task-specific model. One of the challenges in FSL is to enable models to quickly adapt to dynamic environments and previously unseen categories. Simon et al. [28] proposed a framework that incorporates dynamic classifiers constructed from a limited set of samples. This approach enhances the few-shot learning model’s robustness against perturbations. Lai et al. [14] proposed an innovative meta-learning approach that captures a task-adaptive classifier-predictor, enabling the generation of customized classifier weights for few-shot classification tasks. Xiao et al. [33] presented an adaptive mixture mechanism that enhances generation of interactive class prototypes. They also introduced a loss function for joint representation learning, which seamlessly adapts the encoding process for each support instance. Their approach is also built upon the prototypical framework. Han et al. [12] proposed an adaptive instance revaluing network to tackle the biased representation problem. Their proposed method involves an enhanced bilinear instance representatio

and the integration of two original structural losses, resulting in a more robust and accurate regulation of the instance revaluation process. Inspired by these works, we introduce relation-meta learning to obtain the common features that were used to rectify the biased representation, making the model generalize better and have stronger adaptability.

3 Methodology

The general architecture of the AdaProto model, proposed by us, is depicted in Fig. 2. This model is designed to tackle the task of extracting relationships between entity pairs in few-shot scenarios. The model consists of four distinct modules:

3.1 Relation-Meta Learning

To obtain relation-meta information from relation instances, we introduce a relation-meta learner, inspired by MetaR [3], that learns the relative common features, thus maintaining the task-specific weights for the relational classes. Unlike the previous method, our approach obtains the relation-meta directly from the relation set, not from the entity pairs, and transfers the common features to entity pairs for rectifying the relation representations. Firstly, we generate a template called relation set by combining the relation name and description as “*name:description*”(n; d), and feed them into the encoder to produce the feature representation. To enhance the relation representation, we employ the mean value of the hidden states of the sentence tokens and concatenate it with their corresponding [CLS] token to represent the relation classes $\{r_i \in R^{2d}; i = 1, \dots, N\}$. Next, the relation-meta learner compares all the N relation representations obtained from the encoder and generates the only representation specific to each relational label in the task. Finally, to represent the common features through the relation-meta learner, we design a nonlinear two-layer fully connected neural network and a mutual information mechanism to explore the feature representations by training the relation-meta learner.

$$R_m = GELU(H_{r_i}W_1 + b_1)W_2 + b_2 \quad (1)$$

where R_m is the representation of relation-meta, H_{r_i} is an output of the encoder corresponding to r_i , and W_i, b_i are the learning parameters. The hidden embeddings are further exploited by the fully-connected two layers network with $GELU(\cdot)$ activation function. We assume that the ideal relation-meta R_m only retains the relation specific to the i -th class r_i and is orthogonal to other relations. Therefore, for all relational categories R , when $i \neq j$, the classes should satisfy their mutual information $MI(R_i, R_j)=0$, and the relative discriminative representation of each relation is independent. To achieve this goal, we design a training strategy based on mutual information, which constrains the relative

common feature representation of each relation that only contains the information specific to the relational classes by minimizing the mutual information loss function.

$$L_m = \sum_{1 \leq i, j \leq N, i \neq j} MI(R_i, R_j) \quad (2)$$

3.2 Entity-Pair Rectification Learning

To transfer the common features and obtain discriminative features, following the previous transE [1], we design score function to evaluate the interaction of entity pairs and common features learned from the relation-meta learner. The function can be formulated as the interaction between the head entity, tail entity, and relations to generate the rectified relation representation:

$$s(h_i, t_i) = \|R_m + h_i - t_i\| \quad (3)$$

where R_m represents the relation-meta, h_i is the head entity embedding and t_i is the tail entity embedding. The L_2 norm is denoted by $\|\cdot\|$. With regards to the relational classes in the given task, the entity-pair rectification learner evaluates the relevance of each instance in the support set. To efficiently evaluate the effectiveness of entity pair, inspired by MetaR [3], we design loss function to update the learned relational classes R_m with the score of the entity pairs and common features in the following way:

$$L_s(S_r) = \sum_{h, t \in S_r} (\lambda + s(h_i, t_i) - s(h_i, t'_i)) \quad (4)$$

where λ is the margin hyperparameter, e.g., $\lambda=1$, t'_i is the negative example of t_i , and $s(h_i, t'_i)$ is the negative instance score corresponding to current positive entity pair $s(h_i, t_i) \in S_r$. And the rectified relations E_r can be represented as follows:

$$E_r = R_m - \nabla_{R_m} L_s(S_r) \quad (5)$$

where ∇ is the gradient.

3.3 Instance Representation Learning

Textual context is the main source of the relation classification, while the instances vary differently in the randomly sampled set, resulting in the discrepancy in relation features, especially for the task with similar relations. The keywords play a vital role in discriminating different relations in a sentence, so we design attention mechanisms to distinguish the relations from instance representation. We allocate varying weights to the tokens based on the similarity between instances with the relations. The k -th relational feature representation R_i^k is obtained by computing the similarity between each instance embedding s_i^k and the relation embedding r_i .

$$R_i^k = \sum_{n=1}^{l_r} \alpha_r^n r_i^n \quad (6)$$

where $k = 1, \dots, K$ is k -th instance, l_r is the length of r_i , and in the matrix for the instances r_i, s_k , the variable n represents the n -th row.

$$\alpha_r^n = \text{softmax}(\text{sum}(r_i(s_k^n)^T)) \quad (7)$$

The discriminative features are defined by the set of K features.

$$X_r = 1/K \sum_{k=1}^K R_k^i \quad (8)$$

3.4 Multi-task Joint Training

We concatenate the common and discriminative features to form the hybrid prototype representation by relation-meta R_t , entity-pair rectification E_t , and instance representation X_t .

$$C_j = [R_t^j; E_t^j; X_t^j] \quad (9)$$

where t denotes the task, R_t^j represents the j -th relations by the relation-meta learner, E_t^j represents the updated j -th relations by the entity-pair rectification learner, and X_t^j represents the contextualized j -th relations by the instance learner at the task-level learning. The model would calculate the classification probability of the query Q based on the given query and prototype representation of N relations.

$$P(y = j|Q, S) = \exp(d(C_j, Q)) / \sum_{k=1}^N \exp(d(C_k, Q)) \quad (10)$$

where variable N represents the number of classes, and the function $d(.,.)$ refers to the Euclidean distance. And the training objective in this scenario is to minimize the cross-entropy loss by employing the negative log-likelihood estimation probability with labeled instances:

$$L_c(Q, S) = -\log P(y = t|Q, S) \quad (11)$$

where t denotes the relation label ground truth. There are only a few number of instances for each task and the tasks vary differently in difficulty. Training a model using only cross-entropy can be challenging to achieve satisfactory results. To tackle this issue, we employ a multi-task joint strategy to enhance the model’s training process. Following previous work [4], we use mutual information loss and relax the constraints because of the complexity.

Next, to fully exploit the interaction between the relation-meta and entity pairs, the rectification loss is introduced. Furthermore, we take L_f with focal loss [17] instead of L_c to balance the different tasks. The focal loss can be expressed in the following manner:

$$L_f(Q, S) = -(1 - P(y = t|Q, S))^\gamma \log P(y = t|Q, S) \quad (12)$$

where factor $\gamma \geq 0$ is utilized to regulate the rate at which easy examples are down-weighted. Finally, the training loss is constructed as follows:

$$L = L_f + \alpha L_m + \beta L_s \quad (13)$$

where hyperparameters α , β control the relative importance of L_m , L_s , respectively, e.g., $\beta=0$ means no rectified relations. Following previous similar work [30], we simply set $\alpha=0.7$ and $\beta=0.3$, and have obtained better performance. And other rates can be explored in the future.

4 Experiments

4.1 Dataset and Evaluation Metrics

Datasets. Our AdaProto model is assessed on the FewRel 1.0 [13] and FewRel 2.0 [9] datasets, which are publicly available and considered as large-scale datasets for FSRE task. These datasets contain 100 relations, each consisting of 700 labeled instances derived from Wikipedia. To ensure fairness, We divide the corpus of 100 relations into three distinct subsets, with a ratio of 64:16:20, correspondingly dedicated to the tasks of training, validation and testing, following the official benchmarks. The relation set is typically provided in the auxiliary dataset. While the label data for the test set is not made public, we can still assess our model’s performance by submitting our prediction results and using the official test script to obtain test scores.

Evaluation. The distribution of FewRel dataset in various scenarios is frequently simulated using the N -way- K -shot (N -w- K -s) approach. Under the typical N -w- K -s setting, each evaluation episode involves sampling N relations, each of which contains K labeled instances, along with additional query instances. The objective is for the models to accurately classify the query instances into the sampled N relations, based on the provided $N \times K$ labeled data. Accuracy is employed as the performance metric in the N -w- K -s scenario. This metric measures the proportion of correctly predicted instances out of the total number of instances in the dataset. Building upon previous baselines, we have selected N values of 5 and 10, and K values of 1 and 5, resulting in four distinct scenarios.

4.2 Implementation Details

To effectively capture the contextual information of each entity and learn a robust representation for instances, leveraging the methodology introduced by Han et al. [11], we utilize the uncased $BERT_{base}$ as the encoder, which is a 12-layers transformer and has a 768 hidden size. To generate the statements for each instance, we merge the hidden states of the beginning tokens of both entity mentions together. To enhance the representation of each entity pair, we further incorporate their word embedding, entity id embedding, and entity

type embedding. In addition, we utilize the label name and its corresponding description of each relation to generate a template. To optimize our model, we utilize the AdamW optimizer [20] and set the learning rate to 0.00002. Our model is implemented using PyTorch framework and deployed on a server equipped with two NVIDIA RTX A6000 GPUs.

4.3 Comparison to Baselines

We assess the performance of our model by comparing it to a set of strong baseline models:

Proto: Snell et al. [29] proposed the original prototype network algorithm.

GNN: Garcia et al. [10] introduced a meta-learning method based on graph neural networks.

MLMAN: Ye and Ling [36] introduced a modified version of the prototypical network that incorporates multi-level matching and aggregation.

REGRAB: Qu et al. [24] proposed a method for bayesian meta-relational learning that incorporates global relational descriptions.

BERT-PAIR: Gao et al. [9] proposed a novel approach to measure the similarity between pairs of sentences.

ConceptFERE: Yang et al. [35] put forward a methodology that incorporates the inherent concepts of entities, introducing additional cues derived from the entities involved, to enhance relation prediction and elevate the overall effectiveness of relation classification.

DRK: Wang et al. [31] proposed an innovative method for knowledge extraction based on discriminative rules.

HCRP: Han et al. [11] designed a hybrid prototype representation learning method based on contrastive learning, considering task difficulty.

Most existing models use BERT as an encoder, and we follow what is known a priori.

4.4 Main Results

Overall Results. In our evaluation, We compare our proposed AdaProto model against a set of strong baselines on the FewRel dataset. And the results of this comparison are presented in Table 1 and 2. The model’s performance is demonstrated in Table 1, our model effectively improves prediction accuracy and outperforms the strong models in each setting on FewRel 1.0, demonstrating better generalization ability. Besides, the performances gains from the 5-shot settings over the second best method (i.e., HCRP) are larger than those of 1-shot scenario. This observation may suggest that the availability of only one instance per relation class restricts the extraction of discriminative features. And the semantic features of a single instance are more prone to deviating from the common features shared by its relational class. By observational analysis, the performance is mainly due to three reasons. 1) Relation-meta learning allows obtaining common features while adapting to different tasks. 2) Entity pair rectification and

instance representation learning empowers prototype network with discriminative features. 3) Joint training also plays an important role.

Table 1. Main results from the validation and testing of the FewRel 1.0. The evaluation metric used in this study is accuracy(%).

Model	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s
Proto	82.92/80.68	91.32/89.60	73.24/71.48	83.68/82.89
GNN	-/75.66	-/89.06	-/70.08	-/76.93
MLMAN(CNN)	79.01/82.98	88.86/92.66	67.37/75.59	80.07/87.29
REGTAB	87.95/90.30	92.54/94.25	80.26/84.09	81.76/87.02
BERT-PAIR	85.66/88.32	89.48/93.22	76.84/80.63	81.76/87.02
ConceptFERE	-/84.28	-/90.34	-/74.00	-/81.82
DRK	-/89.94	-/92.42	-/81.94	-/85.23
HCRP	90.90/93.76	93.22/95.66	84.11/89.95	87.79/92.10
AdaProto(ours)	91.19/94.26	93.87/96.19	85.91/90.61	89.59/93.38

Table 2. The performance on the domain adaptation test set of FewRel 2.0.

Model	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s
Proto	40.12	51.50	26.45	36.39
BERT-PAIR	67.41	78.57	54.89	66.85
HCRP	76.34	83.03	63.77	72.94
AdaProto(ours)	77.13	84.29	65.17	73.85

Performance on FewRel 2.0 Dataset. In order to assess the adaptability of our proposed model, we performed cross-domain experiments using the FewRel 2.0 dataset. Table 2 clearly illustrates the impressive domain adaptability of our model, as evidenced by the results. In particular, our model obtain better results on 5-shot, probably because the local features of the task play a dominant role compared to the common features.

Performance on Hard Tasks in Few-Shot Scenarios. To showcase the adaptability and efficacy of our model in handling difficult tasks, we use Han’s setup [11] to assess the overall capabilities of the models on the validation set of FewRel 1.0 dataset. We considered three distinct 3-way-1-shot scenarios and display the main results in Table 3. *Easy* denotes tasks with easily distinguishable

Table 3. The performance on random, easy, and hard task.

Model	Random	Easy	Hard
Proto	87.37	98.51	35.63
BERT-PAIR	91.14	99.76	38.21
HCRP	93.86	99.93	62.40
AdaProto(ours)	94.75	99.91	65.37

relations, *Random* encompass a set of 10,000 tasks randomly sampled from the validation relations, and *Hard* denotes tasks with similar relations. We choose *Mother*, *Child* and *Spouse* as the three difficult relations because they not only have similar relation descriptions but also have the same entity types. The performance degrades sharply for the difficult tasks, as shown in the Table 3, which indicates that the hard task still faces a great challenge and also illustrates the significant advantage of our AdaProto model.

4.5 Ablation Study

To assess the efficacy of the various modules in our AdaProto model, we conducted an ablation study where we disabled each module individually and assessed the impact on the model’s overall performance. Table 4 illustrates the outcomes of the ablation study, showcasing the performance decline observed for each module when disabled.

Table 4. The Ablation study performance of AdaProto model on FewRel 2.0.

Model	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s
AdaProto	77.13	84.29	65.17	73.85
-Relation-Meta	71.16	78.65	58.43	67.92
-Entity-Pair Rectification	74.98	81.17	63.23	72.19
-Instance Representation	73.11	79.86	61.34	69.25

Experiments were performed on the validation set using the FewRel 2.0 dataset for domain adaptation. Table 4 presents the detailed experimental results. The second part of the Table 4 indicates the results after removing certain modules. It is observed that every module affects performance, especially for the relation-meta learner. In the common features, we eliminate the relation-meta learner and degrade the model to the discriminative features without relation descriptions, and the performance has a large degradation, indicating that common features contribute more to the FSRE task. In the discriminate features, we remove the entity pair rectification and instance representation, respectively, and the results demonstrate that textual context is still the main source of the

discriminative features. In addition, removing entity pair rectification also shows a more significant decrease in performance.

5 Conclusion

In this study, we propose AdaProto, an adaptive prototype network representation that incorporates relation-meta features and various instance features to address the challenges in FSRE tasks. The adaptability of the task heavily relies on the effectiveness of the embedding space. To enhance the model’s generalization, we introduce relation-meta learning to learn common features and better capture the class-centered prototype representation. Additionally, we propose entity pair rectification and instance representation learning to identify discriminative features that differentiate similar relations. Our approach also utilizes a multi-task training strategy to ensure the development of a high-quality model. Furthermore, the model can adapt to different tasks by leveraging multiple feature extractors. Through comparative studies on FewRel in various few-shot settings, we demonstrate that AdaProto, along with each of its components, boosts the classification performance of FSRE tasks and effectively enhances the overall effectiveness and robustness.

Acknowledgements. The authors would like to thank the Associate Editor and anonymous reviewers for their valuable comments and suggestions. This work is funded in part by the National Natural Science Foundation of China under Grants No.62176029. This work also is supported in part by the Chongqing Technology Innovation and Application Development Special under Grants CSTB2022TIAD-KPX0206. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

References

1. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: *Advances in Neural Information Processing Systems*, vol. 26 (2013)
2. Brody, S., Wu, S., Benton, A.: Towards realistic few-shot relation extraction. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 5338–5345 (2021)
3. Chen, M., Zhang, W., Zhang, W., Chen, Q., Chen, H.: Meta relational learning for few-shot link prediction in knowledge graphs. 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)*, pp. 4217–4226 (2019)
4. Cheng, P., et al.: Improving disentangled text representation learning with information-theoretic guidance. arXiv preprint: [arXiv:2006.00693](https://arxiv.org/abs/2006.00693) (2020)
5. Dong, B., et al.: Meta-information guided meta-learning for few-shot relation classification. In: *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 1594–1605 (2020)

6. Dong, M., Pan, C., Luo, Z.: MapRE: an effective semantic mapping approach for low-resource relation extraction. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 2694–2704 (2021)
7. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning, pp. 1126–1135. PMLR (2017)
8. Gao, T., Han, X., Liu, Z., Sun, M.: Hybrid attention-based prototypical networks for noisy few-shot relation classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 6407–6414 (2019)
9. Gao, T., et al.: FewRel 2.0: towards more challenging few-shot relation classification. arXiv preprint: [arXiv:1910.07124](https://arxiv.org/abs/1910.07124) (2019)
10. Garcia, V., Bruna, J.: Few-shot learning with graph neural networks. arXiv preprint: [arXiv:1711.04043](https://arxiv.org/abs/1711.04043) (2017)
11. 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, pp. 2605–2616 (2021)
12. Han, M., et al.: Not all instances contribute equally: Instance-adaptive class representation learning for few-shot visual recognition. *IEEE Trans. Neural Netw. Learn. Syst.* (2022)
13. Han, X., et al.: FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. arXiv preprint: [arXiv:1810.10147](https://arxiv.org/abs/1810.10147) (2018)
14. Lai, N., Kan, M., Han, C., Song, X., Shan, S.: Learning to learn adaptive classifier-predictor for few-shot learning. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(8), 3458–3470 (2020)
15. Lee, W.Y., Wang, J.Y., Wang, Y.C.F.: Domain-agnostic meta-learning for cross-domain few-shot classification. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1715–1719. IEEE (2022)
16. Li, W.H., Liu, X., Bilen, H.: Cross-domain few-shot learning with task-specific adapters. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7161–7170 (2022)
17. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
18. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
19. 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 the Association for Computational Linguistics: NAACL 2022, pp. 1822–1831 (2022)
20. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint: [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017)
21. Nasution, M.K.: Social network mining (SNM): a definition of relation between the resources and SNA. arXiv preprint: [arXiv:2207.06234](https://arxiv.org/abs/2207.06234) (2022)
22. Peng, H., et al.: Learning from context or names? An empirical study on neural relation extraction. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 3661–3672 (2020)
23. Popovic, N., Färber, M.: Few-shot document-level relation extraction. arXiv preprint: [arXiv:2205.02048](https://arxiv.org/abs/2205.02048) (2022)

24. Qu, M., Gao, T., Xhonneux, L.P., Tang, J.: Few-shot relation extraction via Bayesian meta-learning on relation graphs. In: International Conference on Machine Learning, pp. 7867–7876. PMLR (2020)
25. Quan, C., Wang, M., Ren, F.: An unsupervised text mining method for relation extraction from biomedical literature. *PLoS ONE* **9**(7), e102039 (2014)
26. Ren, H., Cai, Y., Chen, X., Wang, G., Li, Q.: A two-phase prototypical network model for incremental few-shot relation classification. In: Proceedings of the 28th International Conference on Computational Linguistics, pp. 1618–1629 (2020)
27. 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.* **35**(5), 4852–4866 (2022)
28. Simon, C., Koniusz, P., Nock, R., Harandi, M.: Adaptive subspaces for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4136–4145 (2020)
29. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
30. Tran, V.H., Ouchi, H., Watanabe, T., Matsumoto, Y.: Improving discriminative learning for zero-shot relation extraction. In: Proceedings of the 1st Workshop on Semiparametric Methods in NLP: Decoupling Logic from Knowledge, pp. 1–6. Association for Computational Linguistics, Dublin, Ireland and Online (2022). <https://doi.org/10.18653/v1/2022.spanlp-1.1>, <https://aclanthology.org/2022.spanlp-1.1>
31. 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, pp. 2129–2140 (2022)
32. Wang, Y., Salamon, J., Bryan, N.J., Bello, J.P.: Few-shot sound event detection. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 81–85. IEEE (2020)
33. 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. 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, pp. 2273–2276 (2020)
35. Yang, S., Zhang, Y., Niu, G., Zhao, Q., Pu, S.: Entity concept-enhanced few-shot relation extraction. arXiv preprint: [arXiv:2106.02401](https://arxiv.org/abs/2106.02401) (2021)
36. Ye, Z.X., Ling, Z.H.: Multi-level matching and aggregation network for few-shot relation classification. arXiv preprint: [arXiv:1906.06678](https://arxiv.org/abs/1906.06678) (2019)