Efficient Information Extraction in Few-Shot Relation Classification through Contrastive Representation Learning

Anonymous ACL submission

Abstract

Differentiating relationships between entity pairs with limited labeled instances poses a significant challenge in few-shot relation classification. Representations of textual data extract rich information spanning the domain, entities, and relations. In this paper, we introduce a novel approach to enhance information extraction combining multiple sentence representations and contrastive learning. While representations in relation classification are commonly extracted using entity marker tokens, we argue that substantial information within the internal model representations remains untapped. To address this, we propose aligning multiple sentence representations, such as the [CLS] token, the [MASK] token used in prompting, and entity marker tokens. Our method employs contrastive learning to extract complementary discriminative information from these individual representations. This is particularly relevant in low-resource settings where information is scarce. Leveraging multiple sentence representations is especially effective in distilling discriminative information for relation classification when additional information, like relation descriptions, are not available. We validate the adaptability of our approach, maintaining robust performance in scenarios that include relation descriptions, and showcasing its flexibility to adapt to different resource constraints[1].

1 Introduction

Relation classification (RC) is an important subtask in the relation extraction, focusing on identifying the types of relations between pairs of entities within a given textual context. Extracting relevant information is central to this task. To achieve this, RC models must distill rich information from sentences, including contextual cues, entity attributes, and relation characteristics. While language models are essential for extracting text representations, their use of vector space in sentence representations is suboptimal (Ethayarajh, 2019). To improve this, recent advances have enhanced sentence representations through various techniques like flow-based methods (Li et al., 2020), whitening operations (Huang et al., 2021), prompting (Jiang et al., 2022), and contrastive learning (Gao et al., 2021; Kim et al., 2021; Zhou et al., 2022).

Relation extraction faces challenges with limited data for many relation types and disproportionate data acquisition costs (Yang et al., 2021). To address this challenge, few-shot RC trains models to quickly adapt to new relation types using only few labeled examples. Common strategies include meta-learning and prototypical networks, leveraging representation similarity to match unseen query instances with few labeled support instances (Snell et al., 2017). Recent studies have enhanced model representations using supplementary data. Yang et al. (2021) and Qu et al. (2020) incorporate information from external knowledge bases, augmenting entity-related knowledge. Wang et al. (2020b) and Yu et al. (2022) utilize linguistic dependencies to integrate structural sentence information in models. Textual relation descriptions also improve prototypical network performance by offering additional insights into relation types (Han et al., 2021; Dong et al., 2021; Liu et al., 2022).

Given the inherent complexity of distinguishing between various relation types, RC applications commonly combine representations of entity marker tokens as sentence representations (Baldini Soares et al., 2019; Dong et al., 2021). Recent work employs contrastive learning in few-shot RC for more discriminative representations (Han et al., 2021; Zhang and Lu, 2022; Dong et al., 2021). Additionally, studies show that representing sentences with the [MASK] token through prompting improves sentence representations (Jiang et al., 2022).

In this study, our goal is to enhance sentence representations for few-shot relation classification

[1] Our model is available at https://anonymous.4open.science/r/MultiRep-6E39.

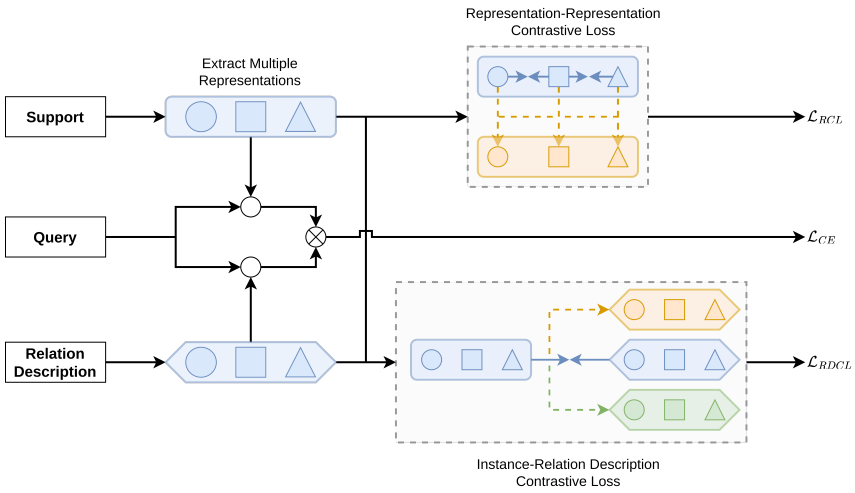

Figure 1: Overview of the MultiRep model, which integrates relation description information. The $\circ$ represents the vector dot product between the instance or relation description and the query samples, while the addition operation is denoted by $\otimes$. Attracting and repelling forces in contrastive learning are represented by $\rightarrow\leftarrow$ and $\dashrightarrow$, respectively.

by aligning multiple representations, such as entity markers and the [MASK] token used in prompting. We recognize that individual model representations, being condensed summaries of the model's internal workings, often carry information that is not directly relevant to relation classification. To address this, we combine various representations to create more comprehensive sentence embeddings, thereby improving the model's discriminative power in low resource settings. While multiple representations offer diverse perspectives, their informational content tends to overlap significantly. To align this information, we introduce contrastive learning objectives providing an effective solution for extracting complementary and discriminative information. A key advantage in our approach is the efficient utilization of resources, since all representations are derived from a single forward pass. Our method is adaptable to different resource constraints by varying the number of representations used in the sentence embeddings. This adaptability allows for a balance between performance and resource usage. Additionally, we extend our approach to include additional information sources, such as relation descriptions. In summary, our contributions are:

- We introduce a novel method for aligning multiple representations in few-shot relation classification using contrastive learning.
- Our approach adapts to various resource constraints and extends to additional information sources, including relation descriptions.
- We emphasize the resource efficiency of our method, improving performance in low-resource settings.

2 Approach

This section provides a detailed overview of our approach, as depicted in Figure 1.

2.1 Sentence Representations

In line with related work, we utilize the BERT-Base model (Devlin et al., 2019) to encode textual inputs (Han et al., 2021; Liu et al., 2022). We describe below the techniques used to generate various sentence representations from the BERT encoder.

Average Pooling computes sentence representations by averaging token representations. Concurrently, the [CLS] token is used as a sentence representation during BERT-Base encoder's pre-training, capturing information of the entire input sequence. The **entity marker** technique augments the input sentence x with markers indicating entities in the text (Baldini Soares et al., 2019). This augments the input into $\bar{x} = [x_0, ..., \texttt{[E1S]}, x_i, \texttt{[E1E]}, ..., x_n]$. The sentence representation is constructed by concatenating the entity start marker representations [E1S] and [E2S] (Baldini Soares et al., 2019). In the **prompting** approach, the RC task is reformulated as a masked language modeling problem. With a template $\mathcal{T}$, each input is transformed into $x_{\text{prompt}} = \mathcal{T}(x)$ containing at least one **[MASK]** token. This masked token represents the relation label and is predicted from the context, e.g., $\bar{x} = \texttt{[MASK]}: x$. (Schick and Schütze, 2021). Gao et al. (2021) introduce using **dropout** masks for generating augmented sentence representations. Since entity marker representations are not suitable for relation descriptions, we employ prompting and [CLS] representations with varying dropout masks.

2.2 Contrastive Representation Learning

The objective of our **representation-representation contrastive loss** term is aligning individual representations to extract discriminatory information for the relation classification task. A key difference to contrastive learning objectives in related work lies in our method of constructing positive instance pairs. In a single forward pass, we derive M different representations from each sentence, and consider these representations as positive pairs. Consequently, representations from other sentences in the training set serve as negative instance pairs. For a given representation r_i^m (where $m \in M$, $i \in N \times K$), we define positive instances r_i^+ and negative instances r_i^- as follows:

$$r_i^+ = \{r_i^{k \neq m} \,|\, k \in M\}$$
$$r_i^- = \{r_{j \neq i}^m \,|\, j \in N \times K\}$$

This aims to maximize the similarity between different representations of the same sentence and minimize the similarity to representations obtained from other sentences (van den Oord et al., 2019; Gao et al., 2021). It ensures that the differentiating factors encoded in the embeddings primarily reflect the underlying sentences, regardless of how these representations are derived from the internal model representations. The representation-representation contrastive loss is computed as follows:

$$\mathcal{L}_{RCL} = \sum_{i=1}^{N \times K} \sum_{m=1}^{M} -log \frac{exp\left(\phi(r_i^m, r_i^+)/\tau\right)}{exp\left(\phi(r_i^m, r_j^-)/\tau\right)},$$

where τ is a temperature scaling parameter, and $\phi(r_i^m, r_i^+)$ represents the element-wise cosine similarity $\sum_{k=1}^{M-1} r_i^m \cdot r_i^k / \|r_i^m\| \|r_i^k\|$ between representation r_i^m and each representation in r_i^+.

In the **instance-relation description contrastive loss**, we leverage the relation descriptions to maximize the similarity between instance representations and corresponding relation description representations. To construct the instance representations R_i and the relation description representations D_i, we concatenate all representations extracted from the encoder $R_i = [r_i^1; r_i^2; ...; r_i^M]$ and $D_i = [d_i^1; d_i^2; ...; d_i^M]$. For instance representation R_i, we select the corresponding relation description D^+ based on the label information in the support set. Non-corresponding relation descriptions D^- form negative pairs. The instance-relation description contrastive loss is computed as follows:

$$\mathcal{L}_{RDCL} = \sum_{i=1}^{N \times K} -log \frac{exp\left(\phi(R_i, D^+)/\tau\right)}{exp\left(\phi(R_i, D^-)/\tau\right)}$$

2.3 Relation Classification

We obtain N class prototypes by averaging the K instance representations in the support set. We compute the similarity between query instances and support prototypes using the vector dot product and selecting the most similar class prototype. For the relation description, we compute the similarity between query instances and relation description representations D. We add the similarities obtained from the relation descriptions with the similarities obtained from the class prototypes and select the most similar prototype and relation description. This is in line with Liu et al. (2022), who instead directly add the prototype and relation description representations. We compute the cross-entropy loss $\mathcal{L}_{CE} = -log\,(z_y)$, where z_y is the probability for class y. The total loss is defined as the sum of the individual loss terms $\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{RCL} + \mathcal{L}_{RDCL}$.

3 Experiments

3.1 Task Definition

In the N-way K-shot evaluation setting, episodes are randomly sampled from the training set. An episode consists of $N \times K$ input sentences x in the support set $S = \{(x_i, rel_i)\}_{i=1}^{N \times K}$ and $N \times K$ inputs from the query set $Q = \{x_i\}_{i=1}^{N \times K}$. The relations are randomly sampled from the relation types included in the training dataset. Importantly, the relation types in the training set are not overlapping with the test set (and validation set) $rel_{\text{train}} \cap rel_{\text{test}} = \varnothing$ (Gao et al., 2019).

3.2 Dataset and Evaluation

Our experiments were conducted using the FewRel dataset, which consists of 700 instances for each of the 100 relation types (Han et al., 2018). Originating from Wikipedia, the dataset is segmented into training, validation, and test sets, containing 64, 16, and 20 relation types, respectively. Following the training and evaluation procedure outlined in Gao et al. (2019), we train MultiRep on randomly sampled episodes from the training set and subsequently evaluate the model on previously unseen data from the validation and test sets. We report classification accuracy, with validation accuracy metrics being averaged over three random seeds. The test set accuracies are referenced from the public FewRel leaderboard[2]. Our MultiRep model is trained for 30,000 iterations on the FewRel training

[2]Our evaluation results on the FewRel test set are available on the public leaderboard, listed under the "multirep" alias.

Model	Relation Descriptions	5-1	5-5	10-1	10-5	Avg.
Proto	-	- / 80.68	- / 89.60	- / 71.48	- / 82.89	- / 81.16
BERT-Pair	-	85.66 / 88.32	89.48 / 93.22	76.84 / 80.63	81.76 / 87.02	83.44 / 87.30
CTEG	-	84.72 / 88.11	92.52 / **95.25**	76.01 / 81.29	84.89 / 91.33	84.54 / 89.00
DAPL	-	- / 85.94	- / 94.28	- / 77.59	- / 89.26	- / 86.77
SimpleFSRE	-	84.77 / **89.33**	89.54 / 94.13	76.85 / 83.41	83.42 / 90.25	83.64 / 89.28
MultiRep (Ours)	-	**87.13** / 89.20	**92.93** / 95.09	**78.42 / 84.18**	**87.29 / 91.65**	**86.44 / 90.03**
TD-Proto	✓	- / 84.76	- / 92.38	- / 74.32	- / 85.92	- / 84.34
HCRP	✓	90.90 / 93.76	93.22 / 95.66	84.11 / 89.95	87.79 / 92.10	89.01 / 92.87
SimpleFSRE	✓	91.29 / **94.42**	**94.05 / 96.37**	86.09 / 90.73	**89.68 / 93.47**	90.28 / **93.75**
MultiRep (Ours)	✓	**92.73** / 94.18	93.79 / 96.29	**86.12 / 91.07**	88.80 / 91.98	**90.36** / 93.38

Table 1: Accuracy on the FewRel validation / test set.

Model	5-1	10-1
MultiRep	92.73	86.12
w/o $\mathcal{L}_{RCL}$	92.08	85.95
w/o $\mathcal{L}_{RDCL}$	90.14	84.16
w/o Avg. Pooling	92.26	85.82
w/o Entity Marker	91.90	84.83
w/o [CLS]	91.35	85.51
w/o [MASK]	91.87	85.80
w/ prototype addition	91.75	85.82

Table 2: Model variants with (w/) or without (w/o) indicated representations and architectural changes evaluated on the FewRel validation set.

set with a batch size of 4 and a learning rate of 2e-5, ensuring our experiments are consistent with related work (Han et al., 2021; Liu et al., 2022).

3.3 Results

We present the results of our MultiRep approach and compare them to relevant benchmark models designed for few-shot RC, some of which incorporate relation descriptions as additional information. For a consistent and fair model comparison, all benchmarked models exclusively utilize BERT-Base (Devlin et al., 2019) as the sentence encoder. This approach aligns with recent findings suggesting that LLMs like GPT-3.5 do not yield superior performance in relation extraction tasks (Meng et al., 2023). The benchmarked models include Proto (Gao et al., 2019), BERT-Pair (Gao et al., 2019), TD-Proto (Yang et al., 2020), CTEG (Wang et al., 2020a), DAPL (Yu et al., 2022), HCRP (Han et al., 2021), and SimpleFSRE (Liu et al., 2022).

Our model evaluation results are summarized in Table 1. We analyze these results for two distinct scenarios: (i) models that do not incorporate additional information, and (ii) models that incorporate relation description information. We observe that MultiRep outperforms existing models, particularly in settings where information is limited. Specifically, this includes scenarios where relation description information is unavailable, as well as 1-Shot settings in the presence of relation description information. To validate the importance of individual components in the MultiRep model, we conducted ablation studies and present the results in Table 2. These results are based on the MultiRep model that incorporates relation description information, evaluated on the FewRel validation set in the 5-Way 1-Shot and 10-Way 1-Shot settings. Our findings indicate that removing the contrastive learning loss terms, $\mathcal{L}_{RCL}$ and $\mathcal{L}_{RDCL}$, substantially reduces model performance. Furthermore, removing individual representations from the MultiRep model has a negative impact on performance, and there are no specific representations that disproportionately affect the model's performance. Additionally, we validate our approach of computing separate instance prototypes and relation description prototypes, as compared to the direct prototype addition method introduced by Liu et al. (2022). Evaluating the performance across varying numbers of representations (see Figure 2), we observe that the average performance improves with an increasing number of representations. Our results demonstrate that our approach yields the best model performance for MultiRep.

4 Conclusion

In this study, we propose aligning multiple sentence representations in few-shot RC through contrastive learning. This method efficiently distills discriminative information from these representations. Our results emphasize its adaptability and effectiveness in diverse scenarios, particularly in low-resource and 1-Shot environments, particularly when relation descriptions are unavailable. A key advantage of our approach lies in its resource efficiency, deriving all sentence representations from a single forward pass.

5 Limitations

Although our approach efficiently utilizes multiple sentence representations within a single forward pass, it is important to note that this involves combining these representations into larger vectors. This aggregation process may require additional memory and computational resources. Moreover, the application of contrastive learning comes with additional computational requirements. Our method is specifically designed for few-shot RC tasks, and its performance might vary when applied to different types of NLP tasks.

References

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.

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

Manqing Dong, Chunguang Pan, and Zhipeng Luo. 2021. MapRE: An effective semantic mapping approach for low-resource relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2694–2704, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Kawin Ethayarajh. 2019. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. 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)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.

Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019. 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)*, pages 6250–6255, Hong Kong, China. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiale Han, Bo Cheng, and Wei Lu. 2021. Exploring task difficulty for few-shot relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2605–2616, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809, Brussels, Belgium. Association for Computational Linguistics.

Junjie Huang, Duyu Tang, Wanjun Zhong, Shuai Lu, Linjun Shou, Ming Gong, Daxin Jiang, and Nan Duan. 2021. WhiteningBERT: An easy unsupervised sentence embedding approach. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 238–244, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ting Jiang, Jian Jiao, Shaohan Huang, Zihan Zhang, Deqing Wang, Fuzhen Zhuang, Furu Wei, Haizhen Huang, Denvy Deng, and Qi Zhang. 2022. PromptBERT: Improving BERT sentence embeddings with prompts. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8826–8837, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Taeuk Kim, Kang Min Yoo, and Sang-goo Lee. 2021. Self-guided contrastive learning for BERT sentence representations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2528–2540, Online. Association for Computational Linguistics.

Bohan Li, Hao Zhou, Junxian He, Mingxuan Wang, Yiming Yang, and Lei Li. 2020. On the sentence embeddings from pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9119–9130, Online. Association for Computational Linguistics.

Yang Liu, Jinpeng Hu, Xiang Wan, and Tsung-Hui Chang. 2022. A simple yet effective relation information guided approach for few-shot relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 757–763, Dublin, Ireland. Association for Computational Linguistics.

Shiao Meng, Xuming Hu, Aiwei Liu, Shuang Li, Fukun Ma, Yawen Yang, and Lijie Wen. 2023. RAPL: A relation-aware prototype learning approach for few-shot document-level relation extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5208–5226, Singapore. Association for Computational Linguistics.

Meng Qu, Tianyu Gao, Louis-Pascal A. C. Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via bayesian meta-learning on relation graphs. *International Conference on Machine Learning (ICML)*, pages 7867–7876.

Timo Schick and Hinrich Schütze. 2021. Exploiting cloze-questions for few-shot text classification and natural language inference. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 255–269, Online. Association for Computational Linguistics.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical Networks for Few-shot Learning. *Advances in Neural Information Processing Systems*, page 11.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2019. Representation learning with contrastive predictive coding.

Yingyao Wang, Junwei Bao, Guangyi Liu, Youzheng Wu, Xiaodong He, Bowen Zhou, and Tiejun Zhao. 2020a. Learning to decouple relations: Few-shot relation classification with entity-guided attention and confusion-aware training. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5799–5809, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Yuxia Wang, Karin Verspoor, and Timothy Baldwin. 2020b. Learning from unlabelled data for clinical semantic textual similarity. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 227–233, Online. Association for Computational Linguistics.

Kaijia Yang, Nantao Zheng, Xinyu Dai, Liang He, Shujian Huang, and Jiajun Chen. 2020. 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, page 2273–2276, New York, NY, USA. Association for Computing Machinery.

Shan Yang, Yongfei Zhang, Guanglin Niu, Qinghua Zhao, and Shiliang Pu. 2021. 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)*, pages 987–991, Online. Association for Computational Linguistics.

Tianshu Yu, Min Yang, and Xiaoyan Zhao. 2022. Dependency-aware prototype learning for few-shot relation classification. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2339–2345, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Peiyuan Zhang and Wei Lu. 2022. Better few-shot relation extraction with label prompt dropout. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6996–7006, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Kun Zhou, Beichen Zhang, Xin Zhao, and Ji-Rong Wen. 2022. Debiased contrastive learning of unsupervised sentence representations. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6120–6130, Dublin, Ireland. Association for Computational Linguistics.

A Model Training

The MultiRep model consists of 109.48 million parameters and was trained on a single NVIDIA A6000 48GB GPU. The combined training and evaluation time for the 5-Way 1-Shot and 10-Way 5-Shot models, incorporating relation descriptions, was 16 hours and 25 hours, respectively.

B Input Templates

To obtain various representations for a given input sequence and relation description within a single forward pass, we augment the inputs with special tokens. For input sequences, we employ the template "`[CLS] [E1], [MASK], [E2] [SEP] [TEXT]`", where `[E1]` and `[E2]` denote the respective entities in the input sentence, and `[TEXT]` represents the augmented input sentence with entity markers. For the relation description, we utilize the template "`[CLS] [MASK]: [RELATION DESCRIPTION]`", where the relation descriptions included in the FewRel dataset are represented by `[RELATION DESCRIPTION]`. The FewRel dataset provides a relation type category and textual description of the relation type. We include both the relation type category and textual description, separated by a comma in the `[RELATION DESCRIPTION]`.

C Ablation Resource Adaptability

To examine the resource adaptability of our method, we explored how varying the number of representations affects relation classification performance.

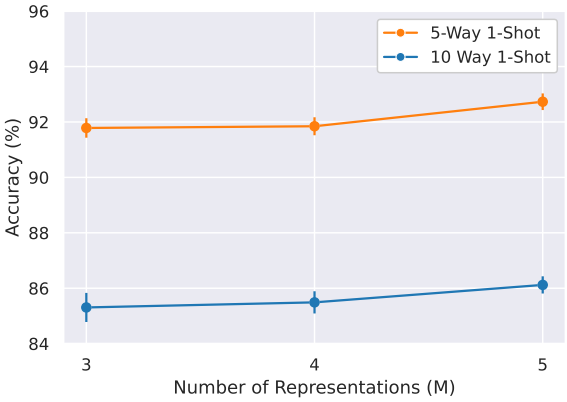

Figure 2: Average accuracy and standard deviation for varying number of representations M evaluated on the FewRel validation set.

We experimented with different combinations of individual representations as outlined in Section 2.1. Figure 2 shows the average accuracy achieved with different numbers of representations (M) on the FewRel validation set, using the MultiRep model including relation descriptions. Our results indicate that the inclusion of more representations generally leads to improved model performance and a reduction in performance variance. This finding implies that the number of representations can be adjusted flexibly to accommodate resource constraints.

D Case Study

This case study highlights the contributions of our MultiRep approach, which integrates both representation-representation and instance-relation description contrastive learning objectives. We demonstrate this by comparing the embeddings and predictions of the MultiRep model, trained with these contrastive learning objectives, against a variant trained without them (MultiRep w/o CL). In a 5-Way 1-Shot setting on the FewRel validation set, the MultiRep w/o CL model achieves an accuracy of 84.83%, substantially lower than the 87.13% achieved by the MultiRep model with contrastive learning objectives.

Both models combine five individual representations to form sentence representations, as detailed in Section 2.1. We display selected prediction results in Table 3. Notably, our MultiRep model demonstrates a tendency to misclassify more challenging cases, such as instances involving difficult relation types like "spouse", "mother", and "child". This observation aligns with the hard classification examples highlighted and addressed by Han et al. (2021). Notably, our MultiRep method outperforms their HCRP model. Conversely, the model variant without contrastive learning objectives is more prone to errors in simpler classification tasks, leading to an overall lower performance. To further demonstrate the benefit of our contrastive representation learning approach, we present a t-SNE projection of sentence representations from both the MultiRep and MultiRep w/o CL models in Figures 3 and 4. These figures allow for a direct comparison of sentence embeddings for relation types that were frequently confused, as listed in Table 3. The projections reveal that sentence embeddings for the same relation type cluster more closely together, while those from different types are more dispersed. This clustering pattern suggests that the sentence embeddings obtained from our MultiRep model effectively extract discriminative information for relation classification.

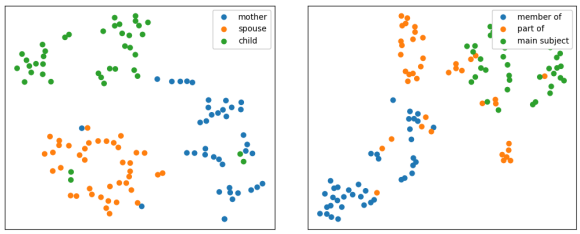

Figure 3: t-SNE visualization of 120 randomly sampled support embeddings for hard relation classification examples obtained from MultiRep.

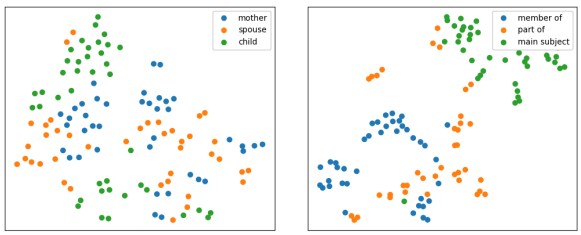

Figure 4: t-SNE visualization of 120 randomly sampled support embeddings for hard relation classification examples obtained from MultiRep w/o CL.

Label	Prediction MultiRep	Prediction MultiRep w/o CL	Sentence
sport	sport	military rank	Jürgen Hasler E1 (born 7 May 1973) is a Liechtenstein former alpine skier E2 who competed in the 1994 Winter Olympics , 1998 Winter Olympics and 2002 Winter Olympics.
main subject	main subject	constellation	Comet 252P / LINEAR E1 is a periodic comet and near - Earth object E2 discovered by the LINEAR survey on April 7 , 2000.
part of	part of	competition class	During the Caucasus Campaign E1 of World War I E2, the Russian Empire occupied most of the Armenian - populated regions of the Ottoman Empire.
spouse	spouse	mother	He was the father - in - law of Azerbaijan 's President Heydar Aliyev E2, who married his daughter Zarifa Aliyeva E1, and maternal grandfather of Azerbaijan 's current President Ilham Aliyev.
child	child	mother	He was the father of Frank Islacker E1 and the grandfather of Mandy Islacker E2.
child	mother	child	"Me and Liza" is about Wainwright's relationship with Liza Minnelli E2, who was reportedly upset by his 2006 tribute concerts to her mother , American actress and singer Judy Garland E1.
spouse	mother	spouse	His brother , Fridolin Weber , who died in 1785 , was the father of Mozart E2 's wife , Constanze E1.
member of	part of	military rank	Designers Domenico Dolce E1 and Stefano Gabbana (Dolce & Gabbana E2) spoke of working with Minogue for the costumes of the tour .
main subject	member of	part of	His grandmother Elisabeth was a member of the Ephrussi family E2, whose history he chronicled in " The Hare with Amber Eyes E1 ".
part of	constellation	constellation	The SS " Edward Y. Townsend " (official number 203449) was a American E1 Great Lakes freighter that served on the Great Lakes of North America E2.

Table 3: Comparative prediction examples: MultiRep with contrastive learning objectives (MultiRep) versus MultiRep without contrastive learning objectives (w/o CL). Examples are obtained from the FewRel validation set.