Entity Concept-enhanced Few-shot Relation Extraction

ShanYang[1]**, Yongfei Zhang**[1,2,3][*]**, Guanglin Niu**[1]**, Qinhua Zhao**[4]**, Shiliang Pu**[5]

[1]Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering,
BeiHang University, Beijing 100191, China
[2]State Key Laboratory of Virtual Reality Technology and Systems,
BeiHang University,Beijing 100191, China
[3]Pengcheng Laboratory, Shenzhen 518055, China
[4]State Key Laboratory of Software Development Environment School of Computer Science
and Engineering, BeiHang University, Beijing 100191, China
[5]Hikvision Research Institute, Hangzhou 311500, China
{shanyang, yfzhang, beihangngl, zhaoqh}@buaa.edu.cn, pushiliang.hri@hikvision.com

Abstract

Few-shot relation extraction (FSRE) is of great importance in long-tail distribution problem, especially in special domain with low-resource data. Most existing FSRE algorithms fail to accurately classify the relations merely based on the information of the sentences together with the recognized entity pairs, due to limited samples and lack of knowledge. To address this problem, in this paper, we proposed a novel entity CONCEPT-enhanced FEw-shot Relation Extraction scheme (ConceptFERE), which introduces the inherent concepts of entities to provide clues for relation prediction and boost the relations classification performance. Firstly, a concept-sentence attention module is developed to select the most appropriate concept from multiple concepts of each entity by calculating the semantic similarity between sentences and concepts. Secondly, a self-attention based fusion module is presented to bridge the gap of concept embedding and sentence embedding from different semantic spaces. Extensive experiments on the FSRE benchmark dataset FewRel have demonstrated the effectiveness and the superiority of the proposed ConceptFERE scheme as compared to the state-of-the-art baselines. Code is available at *https://github.com/LittleGuoKe/ConceptFERE*.

1 Introduction

Relation extraction (RE) is a fundamental task for knowledge graph construction and inference, which however often encounters challenges of long-tail distribution and low-resource data, especially in practical applications including medical or public security fields. In this case, it is difficult for existing RE models to learn effective classifiers (Zhang et al., 2019; Han et al., 2020). Therefore, FSRE has become a hot topic in both academia and industry. Existing FSRE methods can be roughly divided into two categories according to the type of adopted training data. The models of the first category only uses the plain text data, without any external information. The representative Siamese (Koch et al., 2015) and Prototypical (Snell et al., 2017) network in metric learning are used in the FSRE task to learn representation and metric function. BERT-PAIR (Gao et al., 2019b) pairs up all supporting instances with each query instance, and predicts whether the pairs are of the same category, which can be regarded as a variant of the Prototypical network. Gao (Gao et al., 2019a) and Ye (Ye and Ling, 2019) add the attention mechanism to enhance the prototype network. In order to alleviate the problem of insufficient training data, MICK (Geng et al., 2020) learns general language rules and grammatical knowledge from cross-domain datasets. Wang (Wang et al., 2020) proposes the CTEG model to solve the relation confusion problem of FSRE. Cong (Cong et al., 2020) proposes an inductive clustering based framework, DaFeC, to solve the problem of domain adaptation in FSRE.

Since the information of the plain text is limited in FSRE scenarios, the performance gain is marginal. Thus, the algorithms in the second category introduce external information, to compensate the limited information in FSRE, so as to enhance the performance. In order to improve the model's generalization ability for new relations, Qu (Qu et al., 2020) studies the relationship between different relations by establishing a global relation graph. The relations in the global relation graph

---

[*]Corresponding Author

Relation	founder
Sentence	**Microsoft** was founded by **Bill Gates** and Paul Allen on April 4, 1975
Head entity concept	company, vendor, client
Tail entity concept	person, billionaire, entrepreneur

Table 1: The bold words in the sentence correspond to the head entity and the tail entity.

come from Wikidata.[1] TD-Proto (Yang et al., 2020) introduces text descriptions of entities and relations from Wikidata to enhance the prototype network and provide clues for relation prediction.

Although the introduction of knowledge of text description can provide external information for FSRE and achieve state-of-the-art performance, TD-proto only introduces one text description in Wikidata for each entity. However, this might suffer from the mismatching between entity and text description and leads to the degraded performance. Besides, since the text description for each entity is often relatively long, it is not a easy job to extract the most useful information within the long text description.

In contrast to the long text descriptions, the concept is an intuitive and concise description of an entity and can be readily obtained from concept databases, like YAGO3, ConceptNet and Concept Graph, etc. Besides, the concept is more abstract than the specific text description for each entity, which is an idea compensation to the limited information in FSRE scenarios.

As shown in Table 1, intuitively knowing that the concept of head entity is a *company* and the concept of tail entity is an *entrepreneur*, the relation corresponding to the entity pair in the sentence can be limited to a range: *ceo*, *founder*, *inauguration*. On the other hand, some relations should be wiped out, e.g., *educated at*, *presynaptic connection*, *statement describes*. The semantic information of concept can assist determining the relation: *founder* predicted by the model.

To address the above challenges, we propose a novel entity CONCEPT-enhanced FEw-shot Relation Extraction scheme (ConceptFERE), which introduces the entity concept to provide effective clues for relation prediction. *Firstly*, as shown in Table 1, one entity might have more than one concept from different aspects or hierarchical levels and only one of the concepts might be valuable for final relation classification. Therefore, we design a concept-sentence attention module to choose

the most suitable concept for each entity by comparing the semantic similarity of the sentence and each concept. *Secondly*, since the sentence embedding and pre-trained concept embedding are not learned in the same semantic space, we adopt the self-attention mechanism (Devlin et al., 2018) for word-level semantic fusion of the sentence and the selected concept for final relation classification. Experimental results on benchmark dataset show that our method achieves state-of-the-art FSRE performance.

2 Model

2.1 System Overview

Figure 1 shows the structure of our proposed ConceptFERE. The sentence representation module uses BERT to obtain the sentence embedding, the concept representation adopts the pre-trained concept embedding (Shalaby et al., 2019), which uses the skip-gram model to learn the representation of the concept on the Wikipedia text and the Concept Graph. Relation classifier can be implemented by the fully connected layer. The remaining modules of the model will be described in detail below.

2.2 Concept-Sentence Attention Module

Intuitively, one needs to pay more attention to the concept of high semantic correlation with the sentence, which can provide more effective clues for RE. Firstly, since the pre-trained concept embedding (v_c) and sentence embedding (v_s) are not learned in the same semantic space, we can not compare the semantic similarity directly. So the semantic transformation is performed by multiplying the v_c and v_s by the projection matrix P to get their representations $v_c P$ and $v_s P$ in the same semantic space, where P can be learned by fully connected networks. Secondly, by calculating the semantic similarity between sentence and each concept of entity, the similarity value is obtained the dot product of the concept embedding v_c and the sentence embedding v_s as similarity sim_{cs}. Finally, in order to select a suitable concept from the calculated similarity value, we design the 01-GATE.
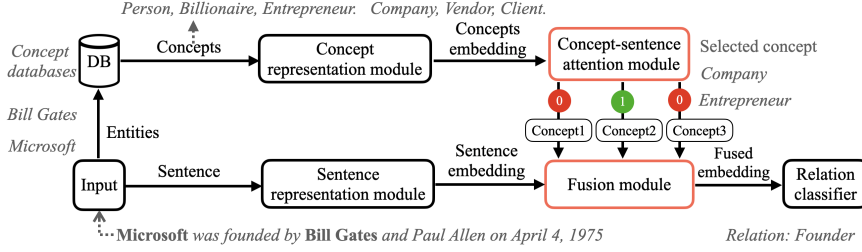
Figure 1: Structure diagram of ConceptFERE model.

The similarity value is normalized by the Softmax function. If sim_{cs} is less than the set threshold α, 01-GATE assigns 0 to the attention score of the corresponding concept, and this concept will be excluded in subsequent relation classification. We choose the suitable concept with the attention score of 1, which is used as a effective clue to participate in relation prediction.

2.3 Self-Attention based Fusion Module

Since concept embedding and the embedding of words in sentences are not learned in the same semantic space, we design a self-attention (Devlin et al., 2018) based fusion module to perform word-level semantic fusion of the concept and each word in the sentence. First, the embedding of all words in the sentence and the selected concept embedding are concatenated, and then fed to the self-attention module. As shown in Figure 2, the self-attention module calculates the similarity value between the concept and each word in the sentence. It multiplies the concept embedding and the similarity value, and then combine with its corresponding word embedding as follow:

$$fusion_{v_i} = \sum_{j=1}^{N} sim\left(q_i, k_j\right) v_j \qquad (1)$$

where $fusion_{v_i}$ represents the embedding of v_i after v_i performs the word-level semantic fusion. The q_i, k_j, and v_j are derived from self-attention, they represent the concept embedding or the word embedding.

3 Experiment

3.1 Dataset, Evaluation and Comparable Models

Dataset: In order to verify our proposed method, we use the most commonly used FSRE dataset FewRel (Han et al., 2018), which contains 100 relations and 70,000 instances extracted from
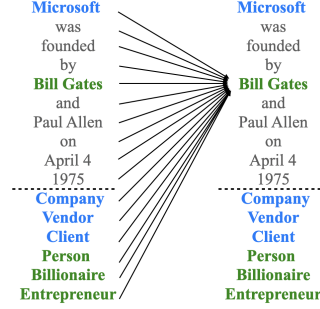

Figure 2: The word-level semantic fusion.

Wikipedia, with 20 relations in the unpublished test set. So we follow previous work (Yang et al., 2020) to re-split the published 80 relations into 50, 14 and 16 for training, validation and testing, respectively.

Evaluation: N-way-K-shot (N-w-K-s) is commonly used to simulate the distribution of FewRel in different situations, where N and K denote the number of classes and samples from each class, respectively. In N-w-K-s scenario, accuracy is used as the performance metric.

Comparable Models: We choose excellent baseline models, GNN (Garcia and Bruna, 2017), SNAIL (Mishra et al., 2017), Proto (Snell et al., 2017), HATT-Proto (Gao et al., 2019a), MLMAN (Ye and Ling, 2019) and TD-proto (Yang et al., 2020) for comparison, and their experimental results are derived from (Yang et al., 2020).

3.2 Model Training Details

The BERT parameters are initialized by bert-base-uncased, and the hidden size is 768. The threshold α is 0.7. Hyperparameters such as learning rate follow the settings in (Gao et al., 2019b). The entity concept is obtained from Concept Graph[2]. Concept Graph is a large-scale common sense conceptual knowledge graph developed by Microsoft, which contains concept of entities stored in triplets (Entity,

---

[2]https://concept.research.microsoft.com/Home/Download

Model	Encoder	5-w-1-s	5-w-5-s	10-w-1-s	10-w-5-s
GNN (Garcia and Bruna, 2017)	CNN	67.30	78.84	54.10	62.89
SNAIL (Mishra et al., 2017)	CNN	71.13	80.04	50.61	66.68
Proto (Snell et al., 2017)	CNN	74.29	85.18	61.15	74.41
HATT-Proto (Gao et al., 2019a)	CNN	74.84	85.81	62.05	75.25
MLMAN (Ye and Ling, 2019)	CNN	78.21	88.01	65.70	78.35
Bert-PAIR (Gao et al., 2019b)	BERT	82.57	88.47	73.37	81.10
TD-Proto (Yang et al., 2020)	BERT	84.76	92.38	74.32	85.92
ConceptFERE	BERT	**89.21**	–	**75.72**	–
ConceptFERE (Simple)	BERT	84.28	90.34	74.00	81.82

Table 2: Accuracies (%) of different models on test set.

IsA, Concept) and can provide concept knowledge for entities in ConceptFERE. The concept embedding adopts the pre-trained concept embedding[3] (Shalaby et al., 2019).

Our proposed scheme is implemented on top of BERT-PAIR, since the concept provided by ConceptFERE can be used as an effective clue.

3.3 Performance and Comparisons

Table 2 tabulates the performance of different comparable models on the test set, where the algorithms in the first group are those state-of-the-art schemes without using any external information, while the TD-Proto in the second group uses external information of text descriptions of entities, and finally our proposed scheme in the third group. It should be noted that, due to the insufficient computing power of our GPU, the performance of the proposed ConceptFERE scheme is tested only under 5way1shot and 10way1shot scenarios. It can be observed from Table 2 that the proposed ConceptFERE model achieves the best performance, as compared to all the comparable schemes. More specifically, ConceptFERE achieves respectively 4.45 and 1.4 gains over the latest TD-Proto using external entity descriptions. And a performance gain of 6.64 and 2.35 is registered as compared to Bert-PAIR, the best model in the first category, under the 5way1shot and 10way1shot scenarios, respectively. This might due to that the generalization ability of concepts is stronger than text description and it is more suitable for FSRE. In theory, 1-shot relation extraction is a more difficult task than 5-shot relation extraction. The experimental results of 1-shot relation extraction have illustrated the effectiveness and superiority of our approach. We believe that our ConceptFERE scheme would also

Model	5-w-1-s
ConceptFERE (bert)	89.21
w/o FUSION	83.11
w/o ATT	84.03
w/o ATT & FUSION	82.57

Table 3: Results of ablation study with ConceptFERE.

achieve the best performance under the other two scenarios.

3.4 Ablation Study

In this section, to verify the effectiveness of the proposed concept-sentence attention module and self-attention based fusion module, presented in 2.2 and 2.3, respectively. As shown in Table 3, without using the concept-sentence attention and fusion module, the model performance of ConceptFERE (simple) drops sharply. This proves that the proposed concept-sentence attention module (ATT) and fusion module (FUSION) can effectively select appropriate concepts and perform word-level semantic integration of concepts and sentences. On the other hand, we present a simplified version of the ConceptFERE model, denoted as ConceptFERE (Simple), in which both the concept selection and fusion module are removed and the concepts and sentences are concatenated and inputted into the relation classification model. Specifically, we can input the concatenated sentences and concepts into BERT-PAIR (Gao et al., 2019b). As shown in Table 2, ConceptFERE (simple) achieves much better performance as compared to Bert-PAIR, the best model in the first category, under all four scenarios. This further validates the effectiveness of introducing the concept in enhancing the RE performance. More importantly, it can be easily applied to other models. As mentioned above, we only need to in-

put the concatenated entity concepts and sentences into the model.

4 Conclusion

In this paper, we have studied the FSRE task and presented a novel entity concept-enhanced FSRE scheme (ConceptFERE). The concept-sentence attention module was designed to select the appropriate concept from multiple concepts corresponding to each entity, and the fusion module was designed to integrate the concept and sentence semantically at the word-level. The experimental results have demonstrated the effectiveness of our method against state-of-the-art algorithms. As a future work, the commonsense knowledge of the concepts as well as the possible relations between them will be explicitly considered to further enhance the FSRE performance.

References

Xin Cong, Bowen Yu, Tingwen Liu, Shiyao Cui, Hengzhu Tang, and Bin Wang. 2020. Inductive unsupervised domain adaptation for few-shot classification via clustering. *arXiv preprint arXiv:2006.12816*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. 2019a. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6407–6414.

Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019b. Fewrel 2.0: Towards more challenging few-shot relation classification. *arXiv preprint arXiv:1910.07124*.

Victor Garcia and Joan Bruna. 2017. Few-shot learning with graph neural networks. *arXiv preprint arXiv:1711.04043*.

Xiaoqing Geng, Xiwen Chen, Kenny Q Zhu, Libin Shen, and Yinggong Zhao. 2020. Mick: A meta-learning framework for few-shot relation classification with small training data. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 415–424.

Xu Han, Tianyu Gao, Yankai Lin, Hao Peng, Yaoliang Yang, Chaojun Xiao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2020. More data, more relations, more context and more openness: A review and outlook for relation extraction. *arXiv preprint arXiv:2004.03186*.

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. *arXiv preprint arXiv:1810.10147*.

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2. Lille.

Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. 2017. A simple neural attentive meta-learner. *arXiv preprint arXiv:1707.03141*.

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

Walid Shalaby, Wlodek Zadrozny, and Hongxia Jin. 2019. Beyond word embeddings: learning entity and concept representations from large scale knowledge bases. *Information Retrieval Journal*, 22(6):525–542.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in neural information processing systems*, pages 4077–4087.

Yingyao Wang, Junwei Bao, Guangyi Liu, Youzheng Wu, Xiaodong He, Bowen Zhou, and Tiejun Zhao. 2020. Learning to decouple relations: Few-shot relation classification with entity-guided attention and confusion-aware training. *arXiv preprint arXiv:2010.10894*.

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*, pages 2273–2276.

Zhi-Xiu Ye and Zhen-Hua Ling. 2019. Multi-level matching and aggregation network for few-shot relation classification. *arXiv preprint arXiv:1906.06678*.

Ningyu Zhang, Shumin Deng, Zhanlin Sun, Guanying Wang, Xi Chen, Wei Zhang, and Huajun Chen. 2019. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. *arXiv preprint arXiv:1903.01306*.