Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases

Chuxu Zhang[1]**, Lu Yu**[3]**, Mandana Saebi**[2]**, Meng Jiang**[2]**, Nitesh V. Chawla**[2]

[1]Brandeis University, [2]University of Notre Dame, [3]King Abdullah University of Science and Technology

[1]chuxuzhang@brandeis.edu, [2]{msaebi, mjiang2, nchawla}@nd.edu, [3]lu.yu@kaust.edu.sa

Abstract

Multi-hop relation reasoning over knowledge base is to generate effective and interpretable relation prediction through reasoning paths. The current methods usually require sufficient training data (fact triples) for each query relation, impairing their performances over few-shot relations (with limited triples) which are common in knowledge base. To this end, we propose FIRE, a novel few-shot multi-hop relation learning model. FIRE applies reinforcement learning to model the sequential steps of multi-hop reasoning, besides performs heterogeneous structure encoding and knowledge-aware search space pruning. The meta-learning technique is employed to optimize model parameters that could quickly adapt to few-shot relations. Empirical study on two datasets demonstrate that FIRE outperforms state-of-the-art methods.

1 Introduction

Nowadays large scale knowledge bases (KB), e.g., NELL (Mitchell et al., 2018) or Freebase (Bollacker et al., 2008), are serving as useful resources for many natural language processing applications such as semantic search or question answering. Due to the nature of incompleteness (Bordes et al., 2013), it is essential to automate the KB completion. One typical problem is fact (triple) prediction. For example, given a query *"What is the nationality of Barack Obama?"* denoted as *(Barack Obama, Nationality, ?)*, the task is to infer *USA* as the answer. There have been a lot of work for solving this problem by embedding learning approaches (Bordes et al., 2013; Socher et al., 2013; Yang et al., 2015) or deep learning models (Dettmers et al., 2018; Schlichtkrull et al., 2018).

The fact prediction ignores the compositional relations in KB and results answer that lacks of interpretation. Accordingly, an alternative problem, multi-hop relation reasoning, was presented. The task is to infer facts using multi-hop reasoning paths, e.g., *(Barack Obama, BornIn, Hawaii)* $\wedge$ *(Hawaii, LocateIn, USA)* $\rightarrow$ *(Barack Obama, Nationality, USA)*. A number of recent models (Xiong et al., 2017; Das et al., 2018; Lv et al., 2019) formulate the problem as sequential decision process and leverage reinforcement learning to achieve considerable performance.

Most current multi-hop relation reasoning models require a good amount of training data (fact triples) for each query relation. However, the relation frequency distribution in KB is usually long-tail (Xiong et al., 2018), showing that a large portion of relations only have few-shot fact triples for model training. Despite that some few-shot relation learning methods (Chen et al., 2019; Lv et al., 2019; Zhang et al., 2020) have been proposed recently, they target at fact prediction only or their performance is suboptimal due to deficiency in capturing heterogeneous structural information and pruning search space in KB.

In this work, we aim at addressing the few-shot challenge and improving relation reasoning performance. In particular, we propose a novel model called FIRE for few-shot multi-hop relation learning over KB. FIRE utilizes on-policy reinforcement learning to model the sequential steps of multi-hop reasoning, encodes entity embedding using heterogeneous structural information, and prunes the reasoning search space using knowledge graph embedding. The meta-learning based optimization procedure is further employed to learn model parameters that could be fast adapted for few-shot relations. To summarize, our main contributions are: (1) we study the problem of few-shot multi-hop relation reasoning over KB, which is new and important; (2) we propose a novel model called FIRE to solve the problem by exploring several beneficial components; (3) we conduct experiments on two datasets and the evaluation results demonstrate

Findings of the Association for Computational Linguistics: EMNLP 2020, pages 580–585

the superior performance of FIRE over state-of-the-art methods.

2 Approach

In this section, we first define the problem of few-shot multi-hop relation reasoning in knowledge bases, then present the FIRE model to solve it.

2.1 Problem Definition

A knowledge base is represented as a knowledge graph (KG) $G = \{\mathcal{E}, \mathcal{R}, \mathcal{T}\}$, where $\mathcal{E}$ and $\mathcal{R}$ denote the entity set and relation set, respectively. $\mathcal{T}$ is the collection of fact triples $(e_s, r_q, e_o) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ in KG. We divide all relations into two groups: few-shot and normal. If the number of triples containing r is smaller than a given threshold K, it is a few-shot relation, otherwise it is a normal relation. The relation reasoning task is to either predict the target entity e_o given the source entity e_s and the query relation r_q: $(e_s, r_q, ?)$, or predict unseen relation r between source entity and target entity: $(e_s, ?, e_o)$. In this work, we will focus on the former one as we want to predict the unseen facts of a given relation. Formally, the problem is defined as follows.

Given a query $(e_s, r_q, ?)$, where e_s is the source entity and r_q is the query few-shot relation, the goal is to perform a multi-hop search over KG and reach the target entity e_o for this query.

2.2 Reinforcement Learning Framework

The problem of multi-hop relation reasoning aims at generating a sequential path from e_s to e_o in KG to interpretate the whole reasoning process. We build the model based on the on-policy reinforcement learning framework (RL) proposed in (Lin et al., 2018). To be more specific, the reasoning process is viewed as a Markov Decision Process (MDP): given the query relation r_q, the agent starts from source entity e_s, then sequentially traverses through a number of relations and entities until it arrives at target entity e_o. In particular, the MDP includes the following modules.

- **State** Each state is represented as $s_t = \{e_t, (e_s, r_q)\} \in \mathcal{S}$, where e_t is the entity visited at step t. Besides, (e_s, r_q) denotes the (source entity, query relation) shared by all states as global context.

- **Action** The action space $\mathcal{A}_t$ for s_t includes all outgoing relations and entities of e_t, i.e., $\mathcal{A}_t = \{(r_{t+1}, e_{t+1}) | (e_t, r_{t+1}, e_{t+1}) \in G\}$. The self-loop edge is added to $\mathcal{A}_t$ for terminating search in a fixed number of steps T.

- **Transition** The transition function is formulated as $\tau(s_t, \mathcal{A}_t) = \{e_t, (e_s, r_q), \mathcal{A}_t\}$. That is, the agent at s_t selects an action $(r_{t+1}, e_{t+1}) \in \mathcal{A}_t$ and changes to $s_{t+1} = \{e_{t+1}, (e_s, r_q)\}$.

- **Reward** The agent will receive a terminal reward $\mathcal{R}(s_T) = 1$ if it finally arrives at the correct target entity, i.e., $e_T = e_o$, otherwise, it will get a reward $\mathcal{R}(s_T) = g((e_s, r_q), e_T)$, where g is a reward shaping function (Lin et al., 2018) using pre-trained knowledge graph embeddings.

To solve the above MDP problem, we apply the policy network to determine action at each state. Specifically, each entity and relation in G is assigned with an embedding vector $\mathbf{e} \in \mathbb{R}^d$ and $\mathbf{r} \in \mathbb{R}^d$. The action $a_t = (r_{t+1}, e_{t+1})$ is denoted as $\mathbf{a}_t = [\mathbf{r}_{t+1} \oplus \mathbf{e}_{t+1}]$, where $\oplus$ is concatenation operator. The search history before step t is encoded with LSTM (Hochreiter and Schmidhuber, 1997):

$$\begin{aligned} \mathbf{h}_0 &= \mathrm{LSTM}(0, [\mathbf{r}_0 \oplus \mathbf{e}_s]) \\ \mathbf{h}_t &= \mathrm{LSTM}(\mathbf{h}_{t-1}, \mathbf{a}_{t-1}), t > 0 \end{aligned} \tag{1}$$

where r_0 is a special start relation introduced to form a start action with e_s, $\mathbf{h}_t$ is the encoded state at step t. The action space is represented by stacking all actions in $\mathcal{A}_t$, i.e., $\mathbf{A}_t \in \mathbb{R}^{|\mathcal{A}_t| \times 2d}$. The corresponding policy network is formulated as:

$$\varphi_\theta(a_t | s_t) = \sigma\{\mathbf{A}_t(\mathbf{W}_2 \mathrm{ReLU}(\mathbf{W}_1 [\mathbf{e}_t \oplus \mathbf{h}_t \oplus \mathbf{r}_q]))\} \tag{2}$$

where σ is the *Softmax* function, θ denotes the set of all model parameters. Let $\mathcal{D}$ be the set of fact triples of query relation r, the objective of policy network is to maximize the expected reward over all triples:

$$\mathcal{J}_r^{\mathcal{D}}(\theta) = \mathbb{E}_{(e_s, r, e_o \in \mathcal{D})}\{\mathbb{E}_{a_1, \cdots a_T \sim \varphi_\theta}[\mathcal{R}(s_T | e_s, r)]\} \tag{3}$$

2.3 Heterogeneous Structure Encoding

RL encodes each entity with an embedding vector. This way, however, is not able to utilize the heterogeneous graph structure information which has been demonstrated to benefit relation learning in graphs (Zhang et al., 2020, 2019; Saebi et al., 2020). Thus we are motivated to design a neural network aggregator (Fig. 1(a)) to enhance the

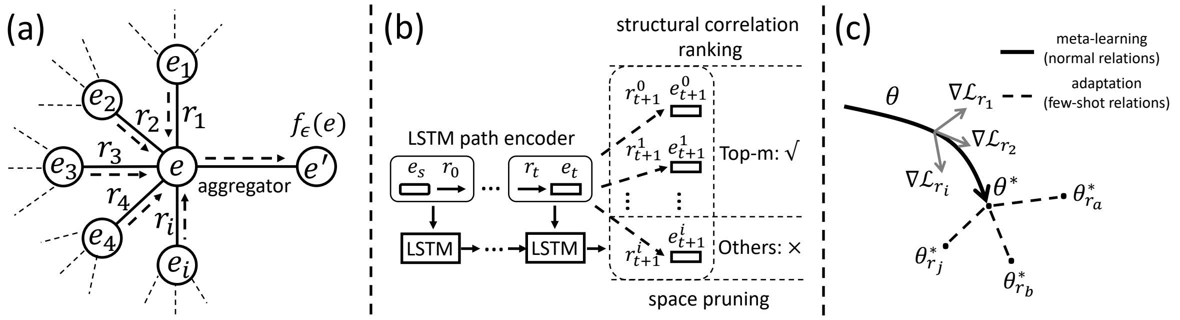

Figure 1: Illustrations of (a) heterogeneous structure encoding; (b) knowledge-aware search space pruning; (c) fast adaption with meta-learning.

entity embedding using heterogeneous neighbors information, which is formulated as follows:

$$f_\epsilon(e) = \delta \left\{ \frac{1}{|\mathcal{N}(e)|} \sum_{(r_i, e_i) \in \mathcal{N}(e)} \mathbf{W}[\mathbf{r}_i \oplus \mathbf{e}_i] \right\} \quad (4)$$

where $\mathcal{N}(e)$ denotes the neighbors set of e, δ is the *tanh* function, and ϵ = $\mathbf{W}$ is learnable parameter. We replace the entity embedding $\mathbf{e}$ in policy network with $f_\epsilon(e)$ such that the model is able to capture heterogeneous structural information for better relation reasoning over KB.

2.4 Knowledge-Aware Search Space Pruning

Some entities in KG have large degrees, making the action search space enormous or even redundant in specific steps. Unlike the previous work (Das et al., 2018; Lin et al., 2018) that cut outgoing edges via centrality score, e.g., PageRank, we assume that structural correlation is important in helping guide the action search, and introduce a knowledge-aware search space pruning strategy (Fig. 1(b)). Specifically, at each state s_t, we first compute structural correlation $C(e_t, e_{t+1})$ between e_t and e_{t+1} using off-the-shelf knowledge graph embedding pre-trained by the existing algorithms such as TransE (Bordes et al., 2013). Then we prune the search space by only considering the m most correlated entities as potential next step.

2.5 Fast Adaptation with Meta-Learning

We employ MAML (Finn et al., 2017) (Fig. 1(c)) to initialize and adapt the policy network parameters. The main idea is to use triples data of normal relations to learn well initialized parameters θ^* which is further adapted to few-shot relations. Formally, we take each relation r as a task $\mathcal{T}_r$. Let $\mathcal{D}_s$ and $\mathcal{D}_q$ denote the support set and query set randomly sampled from the triples of $\mathcal{T}_r$. The relation specific

Algorithm 1: Meta-learning Procedure

```
Require: Distribution of tasks (relations) p(R)
Require: Randomly initialized policy network
    parameters θ
while not done do
    Sample a batch of tasks T_meta from p(R)
    for r ∈ T_meta do
        Sample support set D_s and query set D_q
          from triples of task T_r
        Compute ∇_θ J_r^{D_s}(θ) of Eq. (3)
        Compute adapted parameters θ'_r by Eq. (5):
          θ'_r = θ − α∇_θ J_r^{D_s}(θ)
    end
    Update policy network parameters θ by Eq. (6):
      θ = θ − β∇_θ Σ_{T_r} J_r^{D_q}(θ'_r)
end
```

θ'_r of $\mathcal{T}_r$ is computed using a number of gradient descent updates as follows:

$$\theta'_r = \theta - \alpha \nabla_\theta \mathcal{J}_r^{\mathcal{D}_s}(\theta) \quad (5)$$

Then we evaluate the objective function with relation specific parameters θ'_r on $\mathcal{D}_q$ and go over a number of tasks to update the policy network parameters θ as follows:

$$\theta = \theta - \beta \nabla_\theta \sum_{\mathcal{T}_r} \mathcal{J}_r^{\mathcal{D}_q}(\theta'_r) \quad (6)$$

After sufficient training over normal relations, the well initialized parameters θ^* could further fast adapt to θ_r^* for reasoning for each few-shot relation r. Algorithm 1 shows the meta-learning procedure of the proposed model.

3 Experiments

In this section, we conduct experiments on different datasets to show model performance and related analytic study.

3.1 Datasets

We utilize two datasets NELL-995 (Xiong et al., 2017) and FB15K-237 (Toutanova et al., 2015) for experiment. By following the data processing in (Lv et al., 2019), we obtain normal and few-shot relations (tasks) for model training and adaptation & evaluation. Statistics of normal relations and few-shot relations of two datasets are reported in Table 1.

Dataset	#Ent	#Rel	#Triples
NELL (normal)	63,524	170	115,454
NELL (few-shot)	2,951	30	2,680
FB15K (normal)	14,448	200	268,039
FB15K (few-shot)	3,078	37	4,076

Table 1: Statistics of datasets used in this work.

3.2 Baseline Methods

We consider five recent multi-hop relation reasoning models for performance comparison, including (1) NeuralLP (Yang et al., 2018b); (2) NTP-λ (Rocktäschel and Riedel, 2017); (3) MINERVA (Das et al., 2018); (4) MultiHop (Lin et al., 2018); (5) MetaKGR (Lv et al., 2019).

3.3 Evaluation Metrics

For each query $(e_s, r_q, ?)$ in test data, the model generates a ranking list of possible target entities. We use two popular ranking metrics for performance evaluation: (1) the mean reciprocal rank of correct entities (**MRR**); (2) the proportion of correct entities that rank in the top-k list (**Hit@k**). In this study, k is set to 1.

3.4 Reproducibility

We perform grid search to select hyper-parameters of FIRE. The learning rate is set to 0.0001. The relation/entity embedding dimension and the reasoning step number in reinforcement learning are set to 100 and 3. We use three-layer LSTM for path encoding and the hidden dimension is set to 100 (same as the embedding dimension). The maximum neighbor size in heterogeneous structure encoding is set to 10. The threshold value m in search space pruning is set to 64 and 128 for NELL and FB. We use Pytorch for model implementation and run it on a GPU machine.

3.5 Performance Comparison

The overall performances of all methods are reported in Table 2, where the best results are highlighted in bold and the best baseline scores are indicated by underline. Overall, FIRE achieves the best performances in all cases, demonstrating its strong capability in learning and inferring few-shot multi-hop relations. Additionally, the improvement in NELL is larger than that in FB, showing the advantage of FIRE in sparse data (FB is denser than NELL). Moreover and unsurprisingly, MetaKGR is the best baseline as it involves adaptation for few-shot relations.

Model	NELL-995		FB15K-237	
	MRR	Hit@1	MRR	Hit@1
NeuralLP	17.9	4.8	10.2	7.0
NTP-λ	15.5	10.2	21.0	17.4
MINERVA	20.1	16.2	30.5	28.4
MultiHop	23.1	17.8	42.7	36.7
MetaKGR	<u>25.3</u>	<u>19.7</u>	<u>46.9</u>	<u>41.2</u>
FIRE	**27.3**	**22.5**	**47.8**	**42.3**

Table 2: Performance comparison of all methods. All scores are multiplied by 100.

3.6 Ablation Study

The RL framework of FIRE is augmented with several components. To study the contribution of each component, we perform ablation study by separately removing: (a) heterogeneous structure encoding (– HSE); (b) knowledge-aware space searching (– KAS) from FIRE. Then we compare the performances of these model variants with the whole model. The performance of each model is reported in Table 3. According to this table, removing each component results performance drop, indicating their effectiveness in relation reasoning. In addition, removing HSE impacts significantly, showing the large benefit of using heterogeneous structural information.

Model	NELL-995		FB15K-237	
	MRR	Hit@1	MRR	Hit@1
– HSE	25.1	20.9	47.0	41.4
– KAS	26.8	21.6	47.4	42.0
FIRE	**27.3**	**22.5**	**47.8**	**42.3**

Table 3: Results of model variants.

3.7 Robustness Analysis

As described in the problem definition, we use threshold K to select few-shot relations. Different settings of K represent different train/test data splits. Here we conduct experiment to study the impact of K on model performance. Some triples will be removed to make each few-shot relation only has K triples. The results of three best models on different K using FB data are shown in Figure 2, where K = max denotes the data split used in the original experiment (Table 2). It is easy to find that FIRE consistently outperforms baseline methods, showing its robustness in relation reasoning.

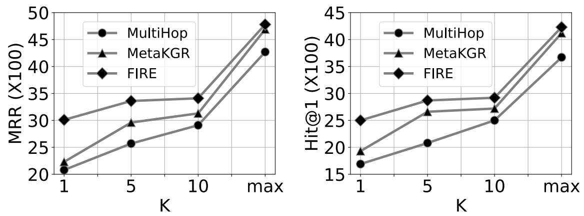

Figure 2: Impact of few-shot threshold K.

4 Related Work

This work is closely related to relation reasoning in knowledge bases and few-shot learning.

Relation Reasoning in Knowledge Bases There have been a lot of work modeling and reasoning relations over knowledge bases. A group of them aim at fact inference by embedding based methods (Bordes et al., 2013; Socher et al., 2013) or deep learning models (Dettmers et al., 2018; Schlichtkrull et al., 2018). For example, Bordes *et al.* (Bordes et al., 2013) proposed TransE that interprets relationships as translation operating on the low-dimensional embeddings of entities. Besides, some targets at generating interpretable multi-hop reasoning paths between entities through reinforcement learning (Xiong et al., 2017; Das et al., 2018; Lv et al., 2019). Recently, a number of work have been proposed (Xiong et al., 2017; Chen et al., 2019; Lv et al., 2019; Zhang et al., 2020) for either fact prediction or multi-hop relation reasoning in few-shot scenario. For instance, Xiong *et al.* (Xiong et al., 2018) presented GMatching model for one-shot relation learning in knowledge bases using matching network and meta-learning. In this paper, we are motivated to explore more potentiality of few-shot relation learning in knowledge bases and move the topic forward.

Few-Shot Learning Few-shot learning (or meta-learning) is to learn from prior experiences to form transferable knowledge for new tasks with few labeled data. Notable approaches have three categories. The first category is metric based methods (Vinyals et al., 2016; Snell et al., 2017) which learn effective similarity space for few-shot instances. For instance, Prototypical Network (Snell et al., 2017) classifies each data sample by computing the distance to prototype representation of each class. The second category is gradient based methods (Finn et al., 2017; Lee and Choi, 2018; Yao et al., 2019) that aim to quickly optimize the model parameters given the gradients on few-shot data instances. For example, MAML (Finn et al., 2017) effectively initializes model parameters via a small number of gradient updates and it can quickly adapt to new few-shot tasks. The last category is memory models (Santoro et al., 2016) which learn to store prior experience (from seen tasks) and generalizes them to unseen tasks. Unlike previous studies that focus on computer vision (Yang et al., 2018a), imitation learning (Duan et al., 2017), graph mining (Yao et al., 2020), we study few-shot relation learning over knowledge bases in this work.

5 Conclusions

In this paper, we studied the problem of multi-hop relation reasoning over knowledge bases in few-shot scenario, and proposed a novel model called FIRE to solve it. FIRE was built on on-policy reinforcement learning and additionally augmented with heterogeneous structure encoding and knowledge-aware search space pruning. It learned and adapted the model parameters for few-shot relations through meta-learning. Experiments on two datasets demonstrated the superior performance of FIRE over state-of-the-art methods. Future work might consider incorporating entity type information to refine entity embeddings and improve relation reasoning performance.

Acknowledgements

This work was supported in part by National Science Foundation grants CCI-1925607 and IIS-1849816. We also thank the anonymous reviewers for their valuable comments and helpful suggestions.

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD*.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NeurIPS*.

Mingyang Chen, Wen Zhang, Wei Zhang, Qiang Chen, and Huajun Chen. 2019. Meta relational learning for few-shot link prediction in knowledge graphs. In *EMNLP-IJCNLP*.

Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, and Andrew McCallum. 2018. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *ICLR*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *AAAI*.

Yan Duan, Marcin Andrychowicz, Bradly Stadie, OpenAI Jonathan Ho, Jonas Schneider, Ilya Sutskever, Pieter Abbeel, and Wojciech Zaremba. 2017. One-shot imitation learning. In *NeurIPS*.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Yoonho Lee and Seungjin Choi. 2018. Gradient-based meta-learning with learned layerwise metric and subspace. In *ICML*.

Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2018. Multi-hop knowledge graph reasoning with reward shaping. In *EMNLP*.

Xin Lv, Yuxian Gu, Xu Han, Lei Hou, Juanzi Li, and Zhiyuan Liu. 2019. Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations. In *EMNLP-IJCNLP*.

Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, Bo Yang, Justin Betteridge, Andrew Carlson, B Dalvi, Matt Gardner, Bryan Kisiel, et al. 2018. Never-ending learning. *Communications of the ACM*, 61(5):103–115.

Tim Rocktäschel and Sebastian Riedel. 2017. End-to-end differentiable proving. In *NeurIPS*.

Mandana Saebi, Steven Krieg, Chuxu Zhang, Meng Jiang, and Nitesh Chawla. 2020. Heterogeneous relational reasoning in knowledge graphs with reinforcement learning. *arXiv preprint arXiv:2003.06050*.

Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. 2016. Meta-learning with memory-augmented neural networks. In *ICML*.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC*.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *NeurIPS*.

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NeurIPS*.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases. In *EMNLP*.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *NeurIPS*.

Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. Deeppath: A reinforcement learning method for knowledge graph reasoning. In *EMNLP*.

Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. 2018. One-shot relational learning for knowledge graphs. In *EMNLP*.

Bishan Yang, Wentau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.

Flood Sung Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. 2018a. Learning to compare: Relation network for few-shot learning. In *CVPR*.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D Manning. 2018b. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *EMNLP*.

Huaxiu Yao, Ying Wei, Junzhou Huang, and Zhenhui Li. 2019. Hierarchically structured meta-learning. In *ICML*.

Huaxiu Yao, Chuxu Zhang, Ying Wei, Meng Jiang, Suhang Wang, Junzhou Huang, Nitesh V Chawla, and Zhenhui Li. 2020. Graph few-shot learning via knowledge transfer. In *AAAI*.

Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. 2019. Heterogeneous graph neural network. In *KDD*.

Chuxu Zhang, Huaxiu Yao, Chao Huang, Meng Jiang, Zhenhui Li, and Nitesh V Chawla. 2020. Few-shot knowledge graph completion. In *AAAI*.