

Exploiting Hybrid Semantics of Relation Paths for Multi-hop Question Answering Over Knowledge Graphs

Zile Qiao¹, Wei Ye^{2,†}, Tong Zhang^{1,3}, Tong Mo¹
Weiping Li¹, Shikun Zhang²

¹ School of Software and Microelectronics, Peking University

² National Engineering Research Center for Software Engineering, Peking University

³ TencentMT Oteam, China

{zileq, wye, zhangtong17, zhangsk}@pku.edu.cn

{motong, wpli}@ss.pku.edu.cn

Abstract

Answering natural language questions on knowledge graphs (KGQA) remains a great challenge in terms of understanding complex questions via multi-hop reasoning. Previous efforts usually exploit large-scale entity-related text corpora or knowledge graph (KG) embeddings as auxiliary information to facilitate answer selection. However, the rich semantics implied in off-the-shelf relation paths between entities is far from well explored. This paper proposes improving multi-hop KGQA by exploiting relation paths' hybrid semantics. Specifically, we integrate explicit textual information and implicit KG structural features of relation paths based on a novel rotate-and-scale entity link prediction framework. Extensive experiments on three existing KGQA datasets demonstrate the superiority of our method, especially in multi-hop scenarios. Further investigation confirms our method's systematical coordination between questions and relation paths to identify answer entities.

1 Introduction

Answering natural language questions on knowledge graphs (KGQA) is a challenging task (Bollacker et al., 2008; Tanon et al., 2016). Recent works mainly pay attention to a complex scenario, namely multi-hop KGQA (Sun et al., 2018; Hu et al., 2018; Saxena et al., 2020; Atzeni et al., 2021), where sophisticated reasoning over multiple edges (or relations) is required to infer the correct answer in the KG (Chen et al., 2019).

The main challenge of multi-hop KGQA is to understand complicated questions and reason under incomplete KG, usually without supervision signals at the intermediate reasoning steps (Lan et al., 2021). One common strategy to alleviate this dilemma is to exploit auxiliary information to enrich knowledge representation. For example,

researchers have exploited entity-related textual corpus (e.g., from Wikipedia) as additional nodes in graph-based neural models (Sun et al., 2018, 2019a), or directly encoded them into enhanced entity representations (Han et al., 2020). A more recent effort, namely EmbedKGQA (Saxena et al., 2020), leverages implicit yet rich information in KG embeddings to answer complex questions over sparse KG. Unfortunately, the relation paths, which may contain beneficial supplementary information to characterize a candidate target entity for a topic entity in a question, are commonly underutilized.

To the best of our knowledge, Yan et al. (2021) is the only effort involving exploiting the off-the-shelf relation path information. They use relation paths as simple coarse-grained input features by concatenating their text descriptions. From a more fine-grained and systematic perspective, given a question as a semantic view for the implied relational fact of a <topic entity, target entity> pair, a relation path can serve as another highly-related yet complementary one.

Therefore, we propose coordinating the question view and the relation path view to identify target entities more accurately. To make this idea work, we face two main challenges: 1) how to accurately represent relation paths and 2) how to fuse a relation path representation with the question representation.

For the first problem, we propose exploiting hybrid features of relation paths by integrating both explicit textual semantics and implicit KG embedding features. Firstly, previous works have shown the merits of introducing entity-related texts (Sun et al., 2018; Xiong et al., 2019; Sun et al., 2019a; Han et al., 2020), while we conjecture that relation-related texts (e.g., relation names or descriptions) can potentially offer helpful clues to answer a question. Meanwhile, relation-related texts are naturally available and on a much smaller scale compared with entity-related texts. Therefore, we utilize the

[†]Corresponding authors.

explicit text description of a relation path (a relation set) as an extra feature of KGQA models to facilitate target entity selection. Secondly, in addition to a text description, a relation also has a KG-based representation (relation embedding) that implicitly contains rich KG structural semantics. Therefore, we introduce RotatE, a KG embedding model that can well support relation composition by entity rotation in the complex vector space (Sun et al., 2019b). With RotatE, we can synthesize relation path representation by performing simple element-wise multiplication of individual relation embeddings. Finally, we characterize beneficial knowledge in candidate relation paths by fusing their structural and textual representations in a question-aware manner, which also facilitates filtering appropriate relation paths semantically consistent with the question among numerous noisy candidates.

For the second problem, inspired by (Saxena et al., 2020), we project the well-designed question-aware mixed representations of relation paths, as well as the question representation, into a rotating entity link prediction framework. However, our pilot experiments showed that the rotating-based link prediction did not yield the robust performance we expected. Further investigation revealed that the modulus of entity embeddings by RotatE mattered. For example, compared with 1-hop answer entities, 2-hop answer entities more significantly differ from their topic entities on modulus. This observation inspires us to match the integrated semantics of questions and relation paths by both entity rotation and entity modulus scaling. After introducing an entity modulus scaling mechanism, we achieve a promising rotate-and-scale prediction framework, which better coordinate knowledge of questions and relation paths for KGQA.

Extensive experiments on three existing KGQA datasets (WebQuestionSP (tau Yih et al., 2016), ComplexWebQuestions (Talmor and Berant, 2018), and MetaQA (Zhang et al., 2018)) verify the superiority of our method, especially in multi-hop scenarios. Our contributions are as follows:

- We propose a KGQA method from a novel perspective of exploiting hybrid features of the off-the-shelf relation paths.
- By systematically fusing explicit Textual information and implicit KG Embedding features of candidate Relation Paths based on a novel rotate-and-scale KG link prediction

framework, our method (**TERP**) achieves competitive performance on three KGQA datasets, especially in the multi-hop scenario.

- We reveal that questions and relation paths, as two facets of their corresponding relations between a topic entity and a target entity, are highly-relevant yet complementary information for question answering.

2 Problem Statement

KGQA is the task of factoid question answering over a knowledge graph. A knowledge graph is denoted as $\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, where \mathcal{E} is the set of all entities in the KG and \mathcal{R} is the set of all relations. A triple can be formally described as (h, r, t) , where $h, t \in \mathcal{E}$ and $r \in \mathcal{R}$ is the relation between them. Given a natural language question $Q = (w_1, \dots, w_{|q|})$ and a topic entity $h \in \mathcal{E}$, which should be present in the question, the task of KGQA is to extract answer entities $c_a \in \mathcal{E}$ that answer the question Q correctly from \mathcal{G} . In practice, we perform entity linking on the question Q , producing a set of topic entities \mathcal{H} .

3 Methodology

3.1 Overview

In this section, we mathematically present our TERP method in detail. Our main idea is to coordinate the question view and the relation path view to identify target entities more accurately. Figure 1 shows an overview of our approach. Firstly, we obtain the representations of the entities and relations in KG via a **KG embedding** module and the representations of questions via a **Question Encoder**. Then, we use a **Path Encoder** to encode the relation path by integrating explicit textual semantics and implicit KG embedding features of relation paths. An attention mechanism is employed to choose the appropriate relation paths semantically consistent with the question among numerous noisy candidate paths. A **Rotate-and-Scale** module projects the representations of the question and the chosen relation paths into the complex space of KG embeddings. Finally, an **Entity Predictor** scores all candidate entities in a link prediction manner.

3.2 KG Embedding

We first obtain the representations of entities and relations via a KG Embedding module. To mine the

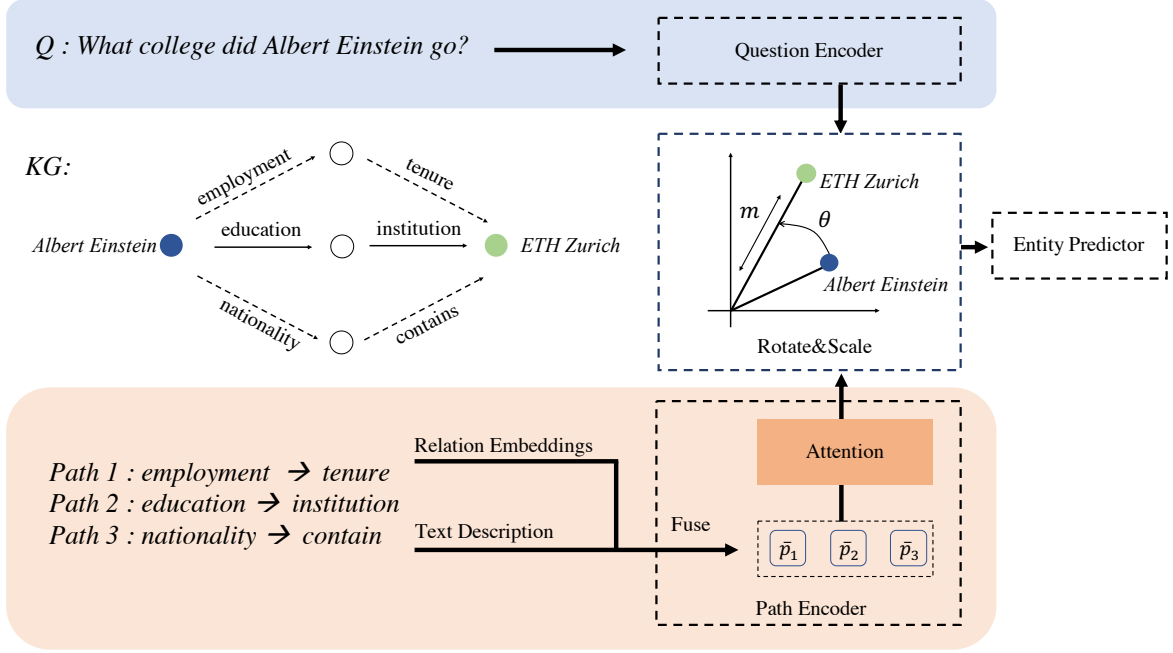


Figure 1: An overview of our TERP model. A question encoder is employed to extract the features from textual question. A path encoder is adopted to capture the consistent yet complementary information in implicit relation embeddings (i.e., structural features) and explicit text description (i.e., textual features) of potential relation paths for the question. Then, the Rotate and Scale mechanism projects the features from question and relation paths to a rotation angle θ and a scaling factor m , respectively. Finally, an entity predictor scores all candidate entities.

implicit KG structural semantics of relation paths, we adopt RotatE (Sun et al., 2019b) to model the composition of relations.

RotatE represents entities as complex vectors and relations as rotations in complex vector space. Given $h, t \in \mathcal{E}$ and $r \in \mathcal{R}$, RotatE generates $e_h, e_r, e_t \in \mathbb{C}^d$ and defines a scoring function:

$$s_r(h, t) = \phi(h, r, t) = -\|e_h \circ e_r - e_t\|, \quad (1)$$

where $|e_{r_i}| = 1$, and \circ denotes the Hadamard (or element-wise) product. RotatE can model the composition patterns. A relation $e_{r_3} = \exp(i\theta_3)$ is a combination of other two relations $e_{r_1} = \exp(i\theta_1)$ and $e_{r_2} = \exp(i\theta_2)$ if and only if:

$$e_{r_3} = e_{r_1} \circ e_{r_2}. \quad (2)$$

3.3 Question Encoder

Following previous work (Saxena et al., 2020; Shi et al., 2021; Atzeni et al., 2021), the question encoding model aims to embed a natural language question Q to a fixed dimension vector $q \in \mathbb{R}^d$ with a pre-trained language model.

$$q = \text{Encoder}_{\text{avg}}(Q), \quad (3)$$

where avg denotes the average pooling strategy.

3.4 Path Encoder

In addition to the question, we further model the explicit and implicit semantics in relation path between a topic entity and a candidate entity with a path encoding module. Our intuition behind this is that the textual semantics in the relation path, as well as the corresponding composed pre-trained relation embeddings produced by RotatE, are supplementary information to the question. Considering that the same relation path may have different semantics in different query contexts, we additionally add the question text before the textual description of the relation path. Formally, given a natural language question Q , a topic entity h and a candidate entity c , we can obtain the shortest paths between h and c . For a single path $r_p^1, r_p^2, \dots, r_p^k$, we generate a textual relation path P_t , in which every text description of relation in the path is surrounded by special tokens $\langle r \rangle, \langle /r \rangle$. We concatenate the textual question Q with the textual relation path P_t and feed them into the encoder to extract the textual features of the explicit relation paths. Meanwhile, we can obtain the implicit semantics from the embeddings of the relations $e_{r_p^1}, e_{r_p^2}, \dots, e_{r_p^k}$, which carry structural features learned by RotatE. The hybrid representation \bar{p} of the relation path is pro-

duced via fusing both explicit textual representation p_t and implicit KG embedding representation p_l with an FFN:

$$\begin{aligned}\bar{p}_t &= \text{Encoder}_{\text{avg}}(P_t : Q), \\ \bar{p}_l &= e_{r_p}^1 \circ e_{r_p}^2 \circ \dots \circ e_{r_p}^k, \\ \bar{p} &= \text{FFN}(\bar{p}_t, \bar{p}_l).\end{aligned}\quad (4)$$

One underlying challenge is that there could be multiple shortest paths between a topic entity and an answer entity. In TERP, we use a scaled dot-product attention mechanism to select the appropriate relation paths that are semantically consistent with the question.

$$p = \text{Attention}(q\mathbf{W}_1, \bar{p}_t\mathbf{W}_2, \bar{p}), \quad (5)$$

where \bar{p}_t and \bar{p} denote two version of representations of all the candidate paths for a given (h, c) pair. $\mathbf{W}_1, \mathbf{W}_2$ are learnable matrices.

3.5 Rotate-and-Scale

However, in our preliminary exploration, directly using RotatE as our KG embedding module do not yield a robust performance. Further investigation revealed that the underlying reason is that the modulus of entities varies in RotatE. If there is an edge between two entities, they will have similar modulus because the modulus of relation representations are fixed to be 1 in RotatE. However, in a multi-hop KBQA scenario, the multi-hop relation path could amplify the difference between the topic and answer entities, hence it can be challenging to match the answer only by rotation transformation.

To this end, we propose a rotate-and-scale framework to model the two views of implied relational fact of a <topic entity, target entity> pair as a rotation transformation and a scaling transformation in the complex space. For a natural language question Q , two independent feedforward networks (FFN) are used to generate the rotation transformation $\theta_q \in \mathbb{R}^d$ and the scaling transformation $m_q \in \mathbb{R}^d$:

$$\begin{aligned}\theta_q &= \text{FFN}(q), \\ m_q &= \text{FFN}(q).\end{aligned}\quad (6)$$

Then, we combine the two transformations into the final representation r_q for the question, which contains a real part Re and an imaginary part Im in the complex space.

$$\begin{aligned}\text{Re}(r_q) &= m_q \circ \cos(\theta_q), \\ \text{Im}(r_q) &= m_q \circ \sin(\theta_q).\end{aligned}\quad (7)$$

In this way, our rotate-and-scale framework can serve as the bridge between entity representations of RotatE and representations of textual questions. We handle relation paths in a same way.

$$\begin{aligned}\theta_p &= \text{FFN}(p), \\ m_p &= \text{FFN}(p), \\ \text{Re}(r_p) &= m_p \circ \cos(\theta_p), \\ \text{Im}(r_p) &= m_p \circ \sin(\theta_p),\end{aligned}\quad (8)$$

where $\theta_p \in \mathbb{R}^d$ and $m_p \in \mathbb{R}^d$ are rotation and scaling in the complex space respectively. As illustrated in Section 5.2, the rotate-and-scale mechanism improves the performance of KGQA with a large margin.

3.6 Entity Predictor

With the representations produced above, an entity predictor is used to score all candidate entities. Given a question Q , the candidate paths P , a topic entity $h \in \mathcal{E}$ and the candidate entity $c \in \mathcal{E}$, the score function is calculated as:

$$\begin{aligned}s_q(h, c) &= -\|e_h \circ r_q - e_c\|, \\ s_p(h, c) &= -\|e_h \circ r_p - e_c\|,\end{aligned}\quad (9)$$

where s_q and s_p denote the scores from the question view and the relation path view, respectively. The final score is $s = (1-\lambda)s_q(h, c) + \lambda s_p(h, c)$, where λ is a hyper-parameter. In training, the score s is calculated among N candidate entities sampled from the KG, where N is a hyper-parameter.

The overall training objective combines the Cross-entropy (CE) loss \mathcal{L}_{ques} and \mathcal{L}_{path} for the s_q and s_p , respectively.

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{ques} + \mathcal{L}_{path} \\ &= \text{CE}(s_q, targets) + \text{CE}(s_p, targets),\end{aligned}\quad (10)$$

where $targets$ denotes the ground truth label.

3.7 Inference

To address the challenge that huge numbers of paths may exist, we propose a two-stage inference strategy to reduce the computational cost. At stage 1, given a question Q , a topic entity h and all the entities in the question-specified subgraph $\mathcal{C} \subseteq \mathcal{E}$, we first compute $s_q(h, c)$ for each $c \in \mathcal{C}$. Then we select top-k candidate entities among them according to $s_q(h, c)$. At stage 2, we compute $s_p(h, c)$ only for the entities recalled in stage 1 and calculate the final score s from them. For questions with more

than one topic entity, we simply average the corresponding s_q and s_p calculated by different topic entities for each candidate c .

This two-stage answer acquisition strategy can empirically deliver a 15-40 \times inference speed-up on different datasets without sacrificing performance.

4 Experimental Settings

4.1 Datasets

We evaluate our model on three widely-used KGQA datasets, MetaQA (Zhang et al., 2018), WebQuestionsSP (tau Yih et al., 2016) datasets, and Complex WebQuestions (Talmor and Berant, 2018).

MetaQA is a multi-hop KGQA dataset with more than 400k questions, providing a KG with 135k triples, 43k entities and 9 kinds of relations.

WebQuestionSP(WebQSP) is a large scale multi-hop KGQA dataset with 4,737 questions. Following Sun et al. (2018, 2019a), we restrict the KG to be a subset of Freebase which contains all facts that are within 2-hops of any entity mentioned in the questions of WebQSP. Then we use the same PPR algorithm as in Sun et al. (2018) to retrieve a subgraph for each question. We further split the testset on WebQSP into 1- and 2-hop sets based on the inferential chain annotation (tau Yih et al., 2016) in the dataset. Note that this split is just for statistics convenience on testsets. During inference, we do not know whether a question is 1-hop or 2-hop, which is different from the MetaQA settings. Following Sun et al. (2018), we remove half of the triples in the KG to simulate an incomplete KG. We call this setting **WebQSP-50**. We use the same train/dev/test split as Sun et al. (2018).

ComplexWebQuestions(CompWebQ) is created by expanding the question entities or adding constraints to the answers in WebQuestionsSP. The questions require up to 4-hops of reasoning on the KG (He et al., 2021). We handle CompWebQ in the same way as WebQSP except that we limit each subgraph to a maximum of 2000 entities in CompWebQ. On average, there are 1349 entities in each subgraph and the recall of answers is 78.6% .

4.2 Implementation Details

We use the open source implementation of LibKGE (Broscheit et al., 2020) to train the KG embeddings. Following Saxena et al. (2020), the pre-trained KG embeddings are frozen for WebQSP

and CompWebQ in training, while tuneable for MetaQA. We use a pre-trained RoBERTa (Liu et al., 2019) as the text encoder. The size of a mini-batch is set to 10. The learning rate is 3e-5 and we adopt the Adam optimizer with $\beta_2 = 0.998$. The number of candidate entities for WebQSP and CompWebQ is 20000. For MetaQA, we use all entities in KG as candidate entities. Other hyper-parameters are the same as the default RoBERTa-base configuration. The weight λ for the entity predictor is 0.6. The number of candidate entities retrieved in stage 1 during inference is empirically set to be 15 for WebQSP and MetaQA, and 30 for CompWebQ.

4.3 Baselines

PullNet (Sun et al., 2019a) iteratively retrieves a subgraph from KG to create a question-specific sub-graph and rank the entities by a variant of graph CNN (Kipf and Welling, 2017); **Embed-KGQA** (Saxena et al., 2020) leverages KG embeddings to perform multi-hop KGQA. It adopts ComplEx (Trouillon et al., 2016) KG embeddings to score the entities; **EMQL** (Sun et al., 2020) leverages query embedding method and uses these embeddings to obtain the answers; **TransferNet** (Shi et al., 2021) leverages free texts retrieved from the textual corpus and pre-defined constrained predicates to perform multi-hop reasoning; **BERT-KGQA** (Yan et al., 2021) leverages textual information carried by the nodes and edges to perform KGQA. We choose the original version without additional annotated data for a fair comparison; **SQUALER** (Atzeni et al., 2021) addresses KGQA by first performing multi-hop reasoning on the KG and then refining the result with logical reasoning.

5 Experiment Results

5.1 Main Results

Table 1 shows the performance of the baseline models and our methods on three datasets under different settings. We achieve the best performance on four of six tasks. Here we mainly compare our TERP with two lines of works: embedding-based methods (e.g., EmbedKGQA) and path searching methods (e.g., SQUALER and TransferNet).

Comparison with embedding-based methods. Except for the similar performance on the MetaQA 1-hop task, TERP significantly outperforms EmbedKGQA on the other tasks. The results verify the effectiveness of incorporating relation path information into the link prediction framework.

Models	MetaQA			WebQSP	WebQSP-50	CompWebQ
	1-hop	2-hop	3-hop			
PullNet*	97.0	99.9	91.4	68.1	51.9	47.2
EmbedKGQA	97.5	98.8	94.8	66.6	54.3†	44.7†
EMQL	97.2	98.6	99.1	75.5	-	-
TransferNet*	97.5	100	100	71.4	-	48.6
BERT-KGQA	-	-	-	71.2	56.7	-
SQALER*	-	99.9	99.9	76.1	55.2	-
TERP(ours)	97.5	99.4	98.9	76.8	57.0	49.2

Table 1: Main results on MetaQA, WebQSP, WebQSP-50 and CompWebQ. The numbers reported in the table are hits@1. “†” denotes the result of our re-implementation. Methods that use external corpora are annotated with “*”.

Comparison with path searching methods. Generally, TERP performs better on WebQSP and ComWebQ, while SQALER and TransferNet are more competitive on MetaQA. The possible reason is that the link prediction framework relies on high-quality KG embeddings, consequently being more effective for knowledge graphs of a larger scale. Note that the scale of WebQSP’s knowledge graph is far more extensive than that of MetaQA’s (1.8 M v.s. 43 K of entity number, and 6101 v.s. 9 of relation type number). In addition to better trained KG embeddings, the 6101 relation types of WebQSP mean 6101 relation representations, introducing much richer semantics of relations, compared with the 9 relation representations of MetaQA. The difference between these knowledge graphs roughly explains the comparison results. Furthermore, the results on more challenging tasks (WebQSP and CompWebQ) verify the effectiveness of integrating explicit textual information and implicit KG structural information in KGQA.

Note that several baselines (EMQL, BERT-KGQA and PullNet) can not accurately fall into the above two categories. Compared with them, TERP also achieves competitive results, e.g., the superior hits@1 scores in 4 of the 6 test sets.

5.2 Effectiveness of Roate-and-Scale Mechanism

To reveal how the rotate-and-scale mechanism helps answer reasoning, we replace it with general ComplEx-based matching and RotatE-based (without scaling) matching, achieving two model variants named **w/ ComplEx** and **w/ RotatE** respectively. Table 2 contains two groups of results corresponding to whether or not hybrid features of relation paths is introduced. The observations on the two groups are generally similar, and here we

mainly analyze the results in the first group.

First, **w/ RotatE** demonstrates a notable performance degradation compared with **w/ ComplEx**, suggesting that simply replacing ComplEx with RotatE in the link prediction-based KGQA framework, can not satisfy our initial desire to exert relation composition capabilities of RotatE. Second, by incorporating the scaling module into RotatE, **w/ RotatE&Scale** surpasses **w/ ComplEx** with a significant margin. This observation verifies that modulus scaling is necessary to capture relation semantics under the hypothesis of using complex vector rotating to match complex multi-hop questions. Third, the superiority of **w/ RotatE&Scale** over **w/ ComplEx** is more visible on 2-hop questions than that on 1-hop ones, proving that **w/ RotatE&Scale** more accurately distinguish relation path semantics.

5.3 Overall Impacts of Relation Paths’ Hybrid Features

Another characteristic of TERP is using hybrid features of relation paths. In Table 2, the models of the second group are ones with relation path features. By comparing the results of the first group and second group in Table 2, we find 1) incorporating relation path information can consistently improve answering questions of different hops under both complete and incomplete KGs, and 2) the improvements on 2-hop questions surpass that on 1-hop ones by a large margin, verifying the potential of relation path information for multi-hop reasoning.

5.4 Ablation Study of Relation Paths’ Hybrid Features

We then perform an ablation study on the hybrid features. Table 3 shows two groups of results corresponding to using only textual representations and only structural representations of relation paths,

Models	WebQSP			WebQSP-50		
	All	1-hop	2-hop	All	1-hop	2-hop
w/o path						
w/ ComplEx	72.1	83.6	52.3	53.6	63.8	36.0
w/ RotatE	67.5	79.7	46.5	49.8	60.5	31.4
w/ RotatE&Scale	74.6	84.5	57.5	55.5	64.6	39.8
w/ path						
w/ ComplEx	73.6(+1.5)	84.6(+0.9)	54.6(+2.3)	54.6(+1.0)	64.2(+0.4)	38.0(+2.0)
w/ RotatE	69.3(+1.8)	80.5(+0.8)	50.4(+3.9)	51.0(+0.4)	60.6(+0.1)	34.2(+2.8)
w/ RotatE&Scale	76.8(+2.2)	84.9(+0.4)	62.6(+5.1)	57.0(+1.5)	65.1(+0.5)	43.1(+3.3)

Table 2: Hits@1 on WebQSP datasets in full KG settings (WebQSP) and incomplete KG settings (WebQSP-50). “All”, “1-hop”, and “2-hop” denote the statistics of 1&2-, 1-, and 2-hop questions of the same task. **w/ path** and **w/o path** denote whether the model is equipped with the path encoder. “w/ ComplEx” and “w/ RotatE” denote the models use ComplEx and RotatE, respectively. “w/ RotatE&Scale” denotes the TERP model with RotatE and the scaling strategy. Numbers in the parentheses denote the hit@1 improvements of **w/ path** over **w/o path**.

respectively. ComplEx does not support relation composition, so we only experiment on RotatE and RotatE&Scale. We have three observations here.

First, both textual and structural features improve model performance, indicating that the two kinds of relation path information benefit answer selection. Second, textual information brings more significant enhancements than structural information. The reasons are two-fold. On the one hand, structural information mainly involves multiplication of relation embedding, which overlaps more with implicit semantics in the link prediction process. On the other hand, textual information provides more complementary knowledge for link prediction, from another modality in a sense. Finally, combining them delivers further improvement, verifying the efficacy of the question-aware fusing process to capture the hybrid semantics.

5.5 Collaboration between Questions and Relation Paths

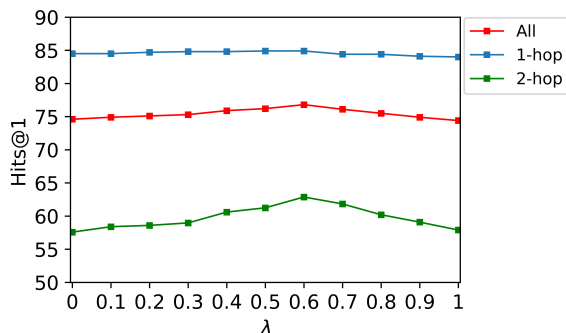


Figure 2: Hits@1 scores for different λ on WebQSP. The blue, green, and red lines denotes the testsets with 1-, 2-, and 1&2-hop questions, respectively.

Considering that exploiting relation paths also introduces many spurious ones, the collaboration of their hybrid features and questions is critical to balance the positive and negative effects. Therefore, we first analyze the impact of the hyper-parameter λ , which denotes the weighting strategy between predicting scores of questions and relation paths.

In Figure 2, the blue, green, and red poly-lines show the Hits@1 scores of all 1-hop, 2-hop, and 1&2-hops questions on WebQSP, respectively. Looking into these three poly-lines’ trends, we find that our model is best-performed when λ is 0.6, indicating the textual information can not either be ignored or overly dependent. In other words, we need to distinguish necessary features under tolerable noises introduced by a set of off-the-shelf relation paths. Another interesting observation is that the upward trend before the peak of the green line (2-hop questions) is more evident than that of the blue line (1-hop questions), though their downward trends after the peak are similar. The reason is that relation path information is more critical for multi-hop reasoning, and our method well characterizes them, hence delivering robust improvements.

To further investigate how relation paths and questions collaborate, we calculate the cosine similarity between relation path text and question text representations for WebQSP. Since there may be multiple candidate relation paths, the relation path with maximum similarity is selected. We then equally divide data samples in the test set into five groups, based on the cosine similarity scores. The performance of two compared models (**w/ path** and **w/o path**) for each group is shown in Figure 3, from which we observe two interesting trends.

Models	WebQSP			WebQSP-50		
	All	1-hop	2-hop	All	1-hop	2-hop
w/ only textual part						
w/ RotatE	68.4	80.2	48.2	50.4	60.4	33.2
w/ RotatE&Scale	76.1	84.6	61.6	56.1	64.7	41.4
w/ only structural part						
w/ RotatE	67.8	79.7	47.3	50.2	60.4	32.7
w/ RotatE&Scale	75.3	84.7	58.9	53.4	61.8	37.9
w/ both						
w/ RotatE	69.3	80.5	50.4	51.0	60.6	34.2
w/ RotatE&Scale	76.8	84.9	62.6	57.0	65.1	43.1

Table 3: Ablation results. “w/ only textual part” denotes the models that only use the textual features. “w/ only structural part” denotes the models that only use the structural features. “w/ both” denotes the models that use both the structural features and the textual features.

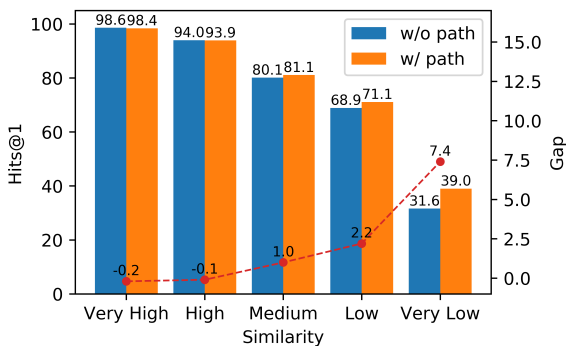


Figure 3: Average Hits@1 scores of **TERP w/o path** and **TERP w/ path** on sub-testsets of WebQSP with different similarities between question and relation paths. The red line denotes the performance gap between the two compared models.

First, model performance degrades as cosine similarity decreases. For example, the hits@1 for “Very High” and “Very Low” differ enormously (e.g., 98.4 v.s. 39.0 with our full TERP). Intuitively, a question is relatively easy to answer if it is similar to a potential relation path. Otherwise, it is more challenging to find the answer. In other words, the question may not provide enough clues, making question understanding more difficult. Second, the relation path information provides more significant improvement for more difficult questions. Incorporating relation path information may even hinder model performance for the groups of “Very High” and “High” (e.g., -0.2 and -0.1 hits@1). This is because many relation paths will bring noises but no extra valuable clues. On the contrary, the hits@1 improvements on “Medium”, “Low”, and “Very Low” are +1.0, +2.2 and +7.3, respectively. These results clearly demonstrate that relation paths

provide complementary information for hard questions, and our method effectively extracts and synthesizes essential features of relation paths. That is where the superiority of our method mainly comes.

6 Related Work

There are two categories of KGQA methods commonly known as semantic parsing-based methods and information retrieval-based methods (Lan et al., 2021). We mainly focus on the second one. Miller et al. (2016) proposes to use Memory Networks to learn dense embeddings of the facts present in the KG to perform QA. Sun et al. (2018, 2019a) create a question-specific subgraph with entities and sentences from the external text corpora and then use a variant of graph CNN to rank the candidate entities. Recently, He et al. (2021) and Shi et al. (2021) utilize path searching methods to perform KGQA. However, they ignore the information in complete relation path. Yan et al. (2021) leverages relation paths to identify answers, but they only explore the textual form of relation. In another line of work, Li et al. (2018) uses TransE (Bordes et al., 2013) to answer the question, but it cannot work in the scenario of KGQA. EmbedKGQA (Saxena et al., 2020) leverages KG embeddings and projects the question into a link prediction framework.

7 Conclusion

We have presented our method for KGQA, which offers a novel perspective of exploiting hybrid features of the off-the-shelf relation paths. We distill essential relation path features by fusing explicit textual information and implicit structural features via a question-aware manner. By project-

ing a natural language question as well as the obtained hybrid features of candidate relation paths into a novel rotate-and-scale entity link prediction framework, we effectively coordinate question and relation paths to select the answer entity. We reveal that questions and relation paths can be seen as two relevant yet complementary facets of their corresponding relations between a topic entity and a target entity.

Acknowledgements

We thank anonymous reviewers for their valuable comments. This research was supported by the National Key Research and Development Program of China (No. 2018YFB1403002).

References

- Mattia Atzeni, Jasmina Bogojeska, and Andreas Loukas. 2021. Sqaler: Scaling question answering by decoupling multi-hop and logical reasoning. In *NeurIPS*.
- Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD Conference*.
- Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*.
- Samuel Broscheit, Daniel Ruffinelli, Adrian Kochsiek, Patrick Betz, and Rainer Gemulla. 2020. LibKGE - A knowledge graph embedding library for reproducible research. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 165–174.
- Zi-Yuan Chen, Chih-Hung Chang, Yi-Pei Chen, Jijnasa Nayak, and Lun-Wei Ku. 2019. UHop: An Unrestricted-Hop Relation Extraction Framework for Knowledge-Based Question Answering. *arXiv:1904.01246 [cs]*. ArXiv: 1904.01246.
- Jiale Han, Bo Cheng, and Xu Wang. 2020. Open domain question answering based on text enhanced knowledge graph with hyperedge infusion. In *FINDINGS*.
- Gaole He, Yunshi Lan, Jing Jiang, Wayne Xin Zhao, and Ji-Rong Wen. 2021. Improving Multi-hop Knowledge Base Question Answering by Learning Intermediate Supervision Signals. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 553–561. ArXiv: 2101.03737.
- Sen Hu, Lei Zou, and Xinbo Zhang. 2018. A state-transition framework to answer complex questions over knowledge base. In *EMNLP*.
- Thomas Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. *ArXiv*, abs/1609.02907.
- Yunshi Lan, Gaole He, Jinhao Jiang, Jing Jiang, Wayne Xin Zhao, and Ji-Rong Wen. 2021. A survey on complex knowledge base question answering: Methods, challenges and solutions. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4483–4491. ijcai.org.
- Dingcheng Li, Jingyuan Zhang, and Ping Li. 2018. Representation learning for question classification via topic sparse autoencoder and entity embedding. *2018 IEEE International Conference on Big Data (Big Data)*, pages 126–133.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.
- Alexander H. Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-value memory networks for directly reading documents. In *EMNLP*.
- Apoorv Saxena, Aditay Tripathi, and Partha Talukdar. 2020. Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4498–4507, Online. Association for Computational Linguistics.
- Jiaxin Shi, Shulin Cao, Lei Hou, Juan-Zi Li, and Hanwang Zhang. 2021. Transfernet: An effective and transparent framework for multi-hop question answering over relation graph. In *EMNLP*.
- Haitian Sun, Andrew O. Arnold, Tania Bedrax-Weiss, Fernando Pereira, and William W. Cohen. 2020. Faithful embeddings for knowledge base queries. *arXiv: Learning*.
- Haitian Sun, Tania Bedrax-Weiss, and William W. Cohen. 2019a. PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text. *arXiv:1904.09537 [cs]*. ArXiv: 1904.09537.
- Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Kathryn Mazaitis, Ruslan Salakhutdinov, and William W. Cohen. 2018. Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text. *arXiv:1809.00782 [cs]*. ArXiv: 1809.00782.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019b. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. *arXiv:1902.10197 [cs, stat]*. ArXiv: 1902.10197.

- Alon Talmor and Jonathan Berant. 2018. The web as a knowledge-base for answering complex questions. In *NAACL*.
- Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. 2016. From freebase to wikidata: The great migration. *Proceedings of the 25th International Conference on World Wide Web*.
- Wen tau Yih, Matthew Richardson, Christopher Meek, Ming-Wei Chang, and Jina Suh. 2016. The value of semantic parse labeling for knowledge base question answering. In *ACL*.
- Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. [Complex Embeddings for Simple Link Prediction](#). *arXiv:1606.06357 [cs, stat]*. ArXiv: 1606.06357.
- Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. 2019. Improving question answering over incomplete kbs with knowledge-aware reader. In *ACL*.
- Yuanmeng Yan, Rumei Li, Sirui Wang, Hongzhi Zhang, Zan Daoguang, Fuzheng Zhang, Wei Wu, and Weiran Xu. 2021. Large-scale relation learning for question answering over knowledge bases with pre-trained language models. In *EMNLP*.
- Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alex Smola, and Le Song. 2018. Variational reasoning for question answering with knowledge graph. In *AAAI*.