Answer-enhanced Path-aware Relation Detection over Knowledge Base

Daoyuan Chen[1], Min Yang[2], Hai-Tao Zheng[3], Yaliang Li[4], Ying Shen[1,*]
[1]School of Electronics and Computer Engineering, Peking University Shenzhen Graduate School
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3]Graduate School at Shenzhen, Tsinghua University [4]Alibaba Group, Bellevue, USA
chendaoyuan@pku.edu.cn,min.yang@siat.ac.cn,zheng.haitao@sz.tsinghua.edu.cn
yaliang.li@alibaba-inc.com,shenying@pkusz.edu.cn

ABSTRACT

Knowledge Based Question Answering (KBQA) is one of the most promising approaches to provide suitable answers for the queries posted by users. Relation detection that aims to take full advantage of the substantial knowledge contained in knowledge base (KB) becomes increasingly important. Significant progress has been made in performing relation detection over KB. However, recent deep neural networks that achieve the state of the art on KB-based relation detection task only consider the context information of question sentences rather than the relatedness between question and answer candidates, and exclusively extract the relation from KB triple rather than learn informative relational path. In this paper, we propose a Knowledge-driven Relation Detection network (KRD) to interactively learn answer-enhanced question representations and path-aware relation representations for relation detection. A Siamese LSTM is employed into a similarity matching process between the question representation and relation representation. Experimental results on the SimpleQuestions and WebQSP datasets demonstrate that KRD outperforms the state-of-the-art methods. In addition, a series of ablation test show the robust superiority of the proposed method.

CCS CONCEPTS

• **Information systems → Information retrieval**.

KEYWORDS

relation detection, knowledge base, representation learning, relational path inference

ACM Reference Format:

Daoyuan Chen, Min Yang, Hai-Tao Zheng, Yaliang Li, Ying Shen. 2019. Answer-enhanced Path-aware Relation Detection over Knowledge Base. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331328

ACM ISBN 978-1-4503-6172-9/19/07...$15.00
https://doi.org/10.1145/3331184.3331328

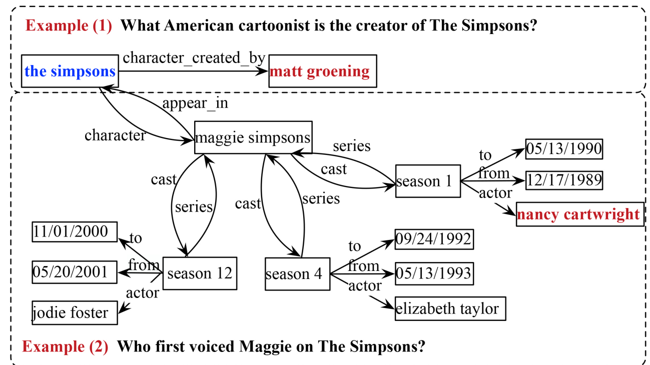

Figure 1: Freebase subgraph of The Simpsons.

1 INTRODUCTION

With the rapid growth of knowledge bases (KBs) on the web, Knowledge Base Question Answering (KBQA) system that returns answers from the KB given natural language questions becomes one of the most promising approaches for acquiring substantial knowledge [15]. To process a case, the KBQA system mainly performs two key tasks, i.e., entity linking that links n-grams in questions to KB entities [10] and relation detection that identifies the KB relation(s) a question refers to [6]. The main focus of this paper is to improve the relation detection on the KBQA task.

Although text-based relation detection has been extensively studied, its counterpart research in KBQA is still relatively new territory and under-explored [11]. Considering the example in Figure 1. **Example (1)** is a single relation example that refers to a single fact of the KB. Based on the detected entity and relation, a query is formed to search the KB for the correct answer "Matt Groening". **Example (2)** is a more complex question that needs multi-hop relation inference. After linking "The Simpsons" to TheSimpsons (the TV show) in the KB, the procedure needs only to examine the predicates that can be applied to TheSimpsons instead of all the predicates in the KB. It is clear that Maggie refers to MaggieSimpsons (the character in The Simpsons). An additional constraint detection takes the play time as a constraint, to filter the correct answer "Nancy Cartwright" from all candidates found by the topic entity and relation.

However, KB-based relation detection remains a challenge compared to text-based relation detection tasks, in that they

- Exclusively represent a question into a single vector using a simple bag-of-words (BOW) model [3] rather than consider its relatedness to the answer candidates. We argue that a question should be represented differently depending on the various aspects of answer candidates.

- Often become a zero-shot learning task, since some test instances may have unseen relations in the training data [8]. In this context, learning and predicting unknown knowledge plays a crucial role. However, limited performance caused by noisy path and abundant candidate relation on large-scale KBs with complex relational network. For example, the number of target relations in most text-based relation detection tasks is limited and is typically less than 100. But even a small KBQA dataset (e.g., Freebase2M [1]) contains more than 6,000 relationship types. In some KBQA datasets such as WebQuestions [2], it is necessary to learn informative relational path from noisy paths instead of extracting the relation directly from a triple, which increases the difficulty of KB relation detection and remains the issue of explicit reasoning to be settled.

To alleviate these limitations, we propose a knowledge-driven relation detection network (KRD) to interactively learn answer-enhanced question representations and path-aware relation representations for relation detection. In specific, we first propose a neural attentive model to represent the questions dynamically according to various candidate answer aspects. Then we explore the KB information to find the core relation path between given entities. Finally, a Siamese LSTM is used into a similarity matching process between the question representation and relation representation.

The main contributions of this paper can be summarized as follows: (1) We propose a novel knowledge-driven relation detection network tailored to the KBQA task, which considers the influence of the answer aspects for representing questions and leverages external knowledge from KB to learn relation representation; (2) We develop a question-relation similarity measure architecture that explores the similarity between question and relation representation learning; (3) The experimental results show that KRD consistently outperforms the state-of-the-art methods.

2 METHODOLOGY

Given a knowledge based question q, our model aims to detect the relations among entities in the question. Formally, our model performs binary classification to decide whether a relation path $[r_1, \ldots, r_n]$ is right answer for the question q, where n is the *hop* for entity pair (e_1, e_2) in a external KG, we call $n = 1$ relation as *single relation* and $n > 1$ relation as *multi-hop relation*.

Figure 2 illustrates the overall architecture of the proposed KRD. Concretely, we first employ an answer-enhanced attentive neural network to learn the representations of questions. Then we use different techniques to perform single relation detection and multi-hop relation detection between a pairwise entity. Afterwards, we introduce our proposed Siamese LSTM to measure the similarity between the representation learning of question and relation.

2.1 Answer-enhanced Question Representation

Given a question $q = [w_1, w_2, \ldots, w_n]$ where w_i denotes the i-th word and n is the length of the question, we adopt Freebase API that contains more than 3 billion facts stored in triples to identify the topic entity of the question. According to the work in [12], Freebase API method can resolve up to 86% questions if we use the top 1 topic entity matching result. We collect the one-hop and two-hop neighbors of topic entity to enrich knowledge for comprehensive reasoning. These entities constitute a candidate set E_c.

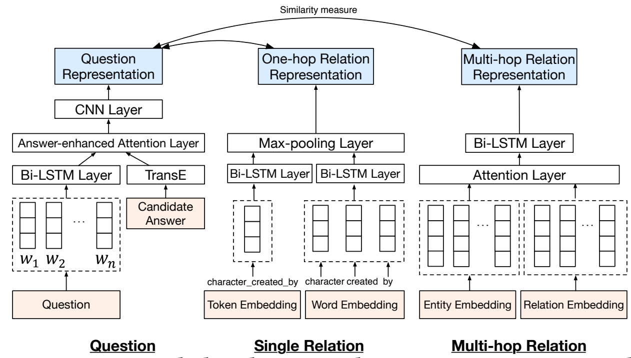

Figure 2: Knowledge-driven Relation Detection network (KRD). Left, middle and right panels denote question representations, single relation representations and multi-hop representations, separately.

As shown in Figure 2, we first look up a word embedding matrix $M_q \in \mathbb{R}^{d \times v_w}$ to get the word embeddings, where d is the dimension of the embeddings and v_w is the vocabulary size of words. Then, these embeddings are fed into a bidirectional long short-term memory (Bi-LSTM) network to capture the past and future hidden state sequences through the forward and backward layer respectively. The output h_t at time step t which concatenates the output of forward layer $\overrightarrow{h_t}$ and backward layer $\overleftarrow{h_t}$, can be calculated as:

$$h_t = [\overrightarrow{h_t} \bigoplus \overleftarrow{h_t}]. \tag{1}$$

Thus the initial question representation $Q_{init} \in \mathbb{R}^{n \times d_q}$ is:

$$Q_{init} = BiLSTM(q), \tag{2}$$

where d_q is the dimension of word embeddings in question.

To learn the answer representation, we adopt the embedding for two answer aspects, i.e., answer entity and answer relation. To be specific, we look up each answer candidate entity and relation through the pre-trained KB embedding matrix $M_e \in \mathbb{R}^{d \times v_E}$ and $M_r \in \mathbb{R}^{d \times v_R}$, where v_E and v_K denotes the KB vocabulary size of entities and relations respectively. A lot of KB embedding methods [9] can be employed for the pre-trained matrices. Then we concatenate the embeddings of these two answer aspects as A_{init}.

Afterwards, we present an answer-enhanced attention mechanism to adaptively pay due attention to the important information of question representations based on answer representations:

$$r_{(t)} = W_a A_{init} + W_q Q_{init}, \tag{3}$$

$$a(t) \propto exp(w_r^\top tanh(r_{(t)})), \tag{4}$$

$$\widetilde{Q_{init}} = Q_{init} a(t)^\top, \tag{5}$$

where W_a, W_q and w_r are attention parameters to be learned, $r_{(t)}$ are answer aspect-guided knowledge vectors, and $a(t)$ is the attention weight value that is applied over question representation. An CNN layer is then employed to learn the final question representation:

$$Q_{final} = CNN(\widetilde{Q_{init}}), \tag{6}$$

where $Q_{final} \in \mathbb{R}^{n \times d_f}$, d_f is the total filter sizes of CNN.

[1] https://developers.google.com/freebase/
[2] https://github.com/brmson/dataset-factoid-webquestions

2.2 Path-aware Relation Representation

Given an entity pair detected in the question q (the entity pair can be generated by NER and entity linking systems), we perform relation detection on both single relation level and multi-hop relation level.

2.2.1 Single Relation Detection. For the single relation detection on the simple question answering task, we adopt two type of embeddings, i.e., **token embeddings** that regards each relation name as a unique token to represent the global information from the original relation names, and **word embeddings** that is treated as a sequence of words from the tokenized relation name. Take a triple (the simpsons, character_created_by, matt groening) as an example, the relation token is "character_created_by" while the relation word includes "character", "created" and "by".

As a result, the input relation includes one token and according word sequence. We transform the input relations to their word embeddings then use two different Bi-LSTMs with shared parameters to get their hidden representations $[H^{token}_{1:M1} : H^{word}_{1:M2}]$, where each row vector H_i is the concatenation between forward/backward representations at i. We initialize the token sequence LSTMs with the final state representations of the word sequence, as a back-off for unseen relations. Next we apply a max-pooling on these two sets of vectors and get the final single relation representation h_s.

2.2.2 Multi-Hop Relation Detection. For the multi-hop relation detection, we attempt to evaluate the relevance among directly linked entities to emphasize paths with rich information and reduce the impact of noisy paths. Given two entities linked by relational path, we first pre-train the TransE model with Freebase to obtain the vector representation of KB entity and relationship. The structures of the entity and relation are encoded as $v_e \in \mathbb{R}^{dl}$ and $v_r \in \mathbb{R}^{dl}$ respectively, where dl is the vector dimension of TransE.

The path representation is learned by alternately merging the entity representation v_e and relation representation v_r as a multi-hop sequence $\{u_1, u_2, \dots, u_l\}$, where l is the number of path elements, u_{2i} and u_{2i+1} are the relation vector and entity vector respectively. We feed the multi-hop sequence to a LSTM layer and obtain the path hidden vectors $P = [p_1, p_2, \dots, p_l]$, $P \in R^{dp \times l}$, where d_p is dimension of LSTM hidden layer size. Next, we apply the self-attention mechanism to encode the path with different information weights as follows:

$$M_p = tanh(W_p P), \tag{7}$$

$$\alpha = softmax(w_p^\top M_p), \tag{8}$$

$$h_m = P\alpha^\top, \tag{9}$$

where $M_p \in R^{dp \times l}$ is nonlinear transformed states, $W_p \in R^{dp \times dp}$ is projection parameter to transform entity and relation embeddings into a same semantic space, and $w_p \in R^{dp}$ is attention parameter to be learned. As a result, we obtain the final multi-hop relation representation h_m, which is encoded from attentive path information within entities and relations.

2.3 Siamese LSTM Similarity Evaluation

Here we use Siamese LSTM to compute the bilinear similarity score between final question representation Q_{final} and single relation representation h_s as well as multi-hop relation representation h_m:

$$sim(Q_{final}, h_s) = LSTM(Q_{final})^\top \hat{W} LSTM(h_s), \tag{10}$$

$$sim(Q_{final}, h_m) = LSTM(Q_{final})^\top \hat{W} LSTM(h_m), \tag{11}$$

where $\hat{W} \in \mathbb{R}^{L \times L}$ is a similarity matrix to be learned. Note the LSTM layer parameters is shared by Q_{final}, h_s and h_m to enhance the answer-question interaction. Afterwards, the similarity scores are fed into a softmax layer and the overall end-to-end model is trained to minimize the cross-entropy loss function:

$$L = -\sum_{i=1}^{N} [y_i \log p_i + (1 - y_i) \log(1 - p_i)] + \gamma \, \|\theta\|_2^2, \tag{12}$$

where p is the output of softmax layer. y denotes the ground-truth label indicates whether the relation is right answer. θ contains all the parameters of the network and $\gamma \, \|\theta\|_2^2$ is the L2 regularization.

3 EXPERIMENT

3.1 Datasets and Implementation Details

To evaluate the single relation detection, we adopt the SimpleQuestions dataset [3] [4] which consists of 108,442 single-relation questions and their corresponding (topic entity, predicate, answer entity) triples associated to Freebase facts. For the multi-hop relation detection, we employ the WebQSP dataset[4] released by Yu et al. [15]. In the single relation detection and multi-hop relation detection tasks, we adopt Freebase2M and the entire Freebase KB for evaluation purpose, respectively.

Pre-trained GloVE embeddings[5] of 300 dimensions are adopted as word embeddings. For all the implemented models, we apply the same parameter settings. The Bi-LSTM hidden layer size and the final hidden layer size are both set to 230. In the implementation, we employ dropout on the output layer and adopt AdaGrad as optimizer. The learning rate and the dropout rate are set to 0.001 and 0.5 respectively. We train our models in batches with size of 32. All other parameters are randomly initialized from [-0.1, 0.1]. The model parameters are regularized with a L2 regularization strength of 0.0001. The maximum length of sentence is set to be 80. The width of the convolution filters is set to be 2 and 3, the number of convolutional feature maps and the attention sizes are set to be 200.

3.2 Experimental Results

Several state-of-the-art baselines are adopted for the comparison of single relation detection: (i) **MemNN** [4], a framework of Memory Networks for retrieving the evidence given a question on simple question answering task. (ii) **CED** [7], a character-level encoder-decoder framework for single-relation question answering. (iii) **CFO** [5], a conditional focused neural network-based approach to answering factoid questions with knowledge bases. (iv) **Attentive CNN** [14], an attention-based CNN combined with maxpooling for both char-CNN and word-CNN.

For the multi-hop relation detection, we employ the following baselines: (i) **SP1** [1], a semantic parser that scales up to Freebase for learning question-answer pairs. (ii) **SP2** [2], a pipeline model that combines an association model and a vector space model, and trains them jointly from question-answer pairs. (iii) **STAGG** [13] a semantic parsing framework that leverages the knowledge base

[3] http://fb.ai/babi

[4] https://github.com/Gorov/KBQA_RE_data

[5] http://nlp.stanford.edu/data/glove.6B.zip

Table 1: Accuracy of single relation detection task using SimpleQuestions dataset and multi-hop relation detection task using WebQSP dataset. The numbers of other systems are either from the original papers or derived from the evaluation script, when the output is available.

Model	Single Relation Detection	Multi-hop Relation Detection
MemNN [4]	88.3	n/a
CED [7]	89.6	n/a
CFO [5]	90.2	n/a
Attentive CNN [14]	91.3	n/a
SP1 [1]	90.0	56.21
SP2 [2]	88.7	55.45
STAGG [13]	87.6	77.42
HR-BiLSTM [15]	93.3	82.53
Our Method (KRD)	**93.5**	**85.72**

Table 2: Ablation Study Results.

Model	Single Relation Detection	Multi-hop Relation Detection
KRD	**93.5**	**85.72**
w/o answer aspect	93.1 (-0.4)	83.37 (-2.35)
w/o token embeddings	91.3 (-2.2)	83.19 (-2.53)
w/o path denoising	90.5 (-3.0)	82.65 (-3.07)

when forming the parse for an input question. (iv) **HR-BiLSTM** [15], a hierarchical recurrent neural network enhanced by residual learning which detects KB relations given an input question.

The experimental results on SimpleQuestions and WebQSP datasets are summarized in Table 1. There are several notable observations:

(1) Our proposed method (KRD) substantially and consistently outperforms the existing methods on both single-hop and multi-hop relation detection tasks, which demonstrates the superiority of taking into account both answer-question interactive information and KB structured knowledge on KB-based relation detection.

(2) The result suggests that our model have an obvious impact on multi-hop unseen relation detection. KRD improves more than 3% on accuracy over all baselines on the multi-hop relation detection task, which indicates that both entity information and structure information in KB are important and require simultaneous consideration especially in path denoising.

3.3 Ablation Analysis

In order to analyze the effectiveness of different factors of KRD, we also report the ablation test in terms of discarding answer aspect learning for the question representation (w/o answer aspect), token embeddings for the single relation detection (w/o token embeddings), and path attention/denoising for the multi-hop relation detection (w/o path denoising), respectively.

The results are summarized in Table 2. Generally, all factors contribute: (1) The results of the ablation test show that incorporating answer aspect can slightly improve the question representation. However, its performance is limited because the application of answer candidates may introduce noise to some extent. (2) A simple token embedding tremendously increases the accuracy of relation detection in simple QA that can be answered with a single fact. (3) It makes larger performance boosting to integrate path denoising on multi-hop relation detection task. The simultaneous consideration of entity and relation in multi-hop path and the attention mechanisms adopted are proven to significantly reduce noise in path searching and selection.

4 CONCLUSION

In the paper, we propose a knowledge-driven relation detection network for relation detection on KBQA task, which effectively considers the impacts of different answer aspects when learning the question representations, and interactively learns KB information for the path denoising and representation. Experimental results on two benchmark datasets demonstrate the superiority of our proposed method on relation detection task. In the future, we will explore the few-shot relation detection beyond the graph structure by leveraging external knowledge from a text corpus to enrich the representational learning of paths. In addition, we will utilize more attention mechanics such as multi-head attention to effectively assemble information from different interaction perspectives toward improving overall question and answer representation learning.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (No.61602013 and No. 61773229), and the Shenzhen Fundamental Research Project (No. JCYJ20170818091546869). Min Yang was sponsored by CCF-Tencent Open Research Fund.

REFERENCES

[1] Jonathan Berant, Andrew Chou, Roy Frostig, and Percy Liang. 2013. Semantic parsing on freebase from question-answer pairs. In *EMNLP*. 1533–1544.
[2] Jonathan Berant and Percy Liang. 2014. Semantic parsing via paraphrasing. In *ACL*, Vol. 1. 1415–1425.
[3] Antoine Bordes, Sumit Chopra, and Jason Weston. 2014. Question answering with subgraph embeddings. In *EMNLP*. 615–620.
[4] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *arXiv preprint arXiv:1506.02075*.
[5] Zihang Dai, Lei Li, and Wei Xu. 2016. Cfo: Conditional focused neural question answering with large-scale knowledge bases. In *ACL*. 800–810.
[6] Wenyu Du, Baocheng Li, Min Yang, Qiang Qu, and Ying Shen. 2019. A Multi-Task Learning Approach for Answer Selection: A Study and a Chinese Law Dataset. In *AAAI*.
[7] David Golub and Xiaodong He. 2016. Character-level question answering with attention. In *EMNLP*. 1598–1607.
[8] Arvind Neelakantan, Benjamin Roth, and Andrew Mc-Callum. 2015. Compositional vector space models for knowledge base inference. In *AAAI*. 156–166.
[9] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *TKDE* 29, 12, 2724–2743.
[10] Min Yang, Qiang Qu, Xiaojun Chen, Wenting Tu, Ying Shen, and Jia Zhu. 2019. Discovering author interest evolution in order-sensitive and Semantic-aware topic modeling. *Information Sciences* 486 (2019), 271–286.
[11] Min-Chul Yang, Nan Duan, Ming Zhou, and Hae-Chang Rim. 2014. Joint relational embeddings for knowledge-based question answering. In *EMNLP*. 645–650.
[12] Xuchen Yao and Benjamin Van Durme. 2014. Information extraction over structured data: Question answering with freebase. In *ACL*, Vol. 1. 956–966.
[13] Scott Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic parsing via staged query graph generation: Question answering with knowledge base. In *IJCNLP*. 1321–1331.
[14] Wenpeng Yin, Mo Yu, Bing Xiang, Bowen Zhou, and Hinrich Schütze. 2016. Simple question answering by attentive convolutional neural network. In *COLING*. 1746–1756.
[15] Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved neural relation detection for knowledge base question answering. In *ACL*, Vol. 1. 571–581.