Improving Question Answering over Knowledge Base with External Linguistic Knowledge

Peyun Wu1, Xiaowang Zhang1,2,
1School of Computer Science and Technology, Tianjin University, Tianjin 300350, China
2Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300350, China
xiaowangzhang@tju.edu.com

Introduction
Most of the existing semantic parsing methods in KBQA focus on capturing the semantics of questions and relations in query graphs. The label of the relation in the first edge of the query graph is taken as core relations for measuring similarity together with some manual features [1, 2]. Another approach to detecting individual relations with word-level is presented in [3] to improve the performance of matching from questions to query graphs where each word of a relation is independently represented, and relation is directly represented by integrating its words’ representations. [3] extends [4] by improving the representation of questions and processing more complex questions. However, existing work that encode the relation semantics in a query graph mainly focus on torus aspects: relation-level word and word-level (we denote their combination as fine-grained relation), the latent semantic information behind the meaning of relations is ignored. For relation-level, each relation is considered as a whole and individual token, which pays more attention to the overall information of the relation. The problem with this aspect is that it suffers from data sparsity in the training process. For word-level, the relation is treated as a sequence of words with the lack of global information of the original relations. In this aspect, word meanings are diversified without considering the word-level relation that might bring noises. For example, for relation “be subject to”, it means “to be made to undergo an unpleasant experience”, the word-level word “subject” might lose that meaning due to its polysemy.

In this paper, to cope with the above limitations, we propose a fine-grained relation with leveraging WordNet to enhance relations representation. WordNet as a lexical dictionary for English, contains synset-like set of synonyms for each word, which can instruct word sense representation learning and word disambiguation. We both consider the original global relation and its word-level relations with word disambiguation.

Query graphs generation
The generation process of query graphs is shown in Figure 1.

![Figure 1: Query graph generation process](image)

Our approach to generating query graphs is an iterative process. Inspired by query graphs generation of [1], we consider 5 kinds of semantic constraints: entity type, temporal (explicit and implicit), order, and compare to process more complex queries. Given a natural language question, we split it into a set of tokens $T$.

Firstly, we extract all possible entities from $T$ with the help of the existing entity linking tool [7]. We can obtain the (mention, entity) pairs to get the entity linking list, which will be filtered with the scores evaluated by the entity linking tool. The set of candidate entities $E$ will be finally returned.

And then, we construct new candidate graphs with a set of predefined operations set $OP$: Take entity, Add Type, Add Temporal (explicit and implicit), Add Compare and Add Order. Fig. 1 shows a flow path of our candidate semantic graph generation process.

Our Approach

Question Representation Encoding: We encode the syntactic dependency tree of a question to a vector as the representation of the given question via GAT. The module is used to provide the condition of calculating the similarity of a question with candidate query graphs.

Figure 2: Query graph generation process

Relation Representation: We obtain the semantic representation of relations in two steps, namely, sense-based attention and hidden relation-based attention (see Fig. 2). We obtain the representation of each fixed-grained relation occurring in a query graph with leveraging WordNet, and we apply hidden relation in questions to assign weights of all relations to query graphs via attention mechanisms.

Experiments and Evaluations

Experiments Setup
Due to Freebase no longer up-to-date, including unavailability of APIs and new changes, we use the full Wikidata dump as our KB and host it with Vincenzo engine. Besides, we adopt Spacy as a syntactic dependency parser and use Deep Graph Library (DGL) to transmute query graphs into DGL graph objects. We conduct our experiments on the following two popular data sets, namely, WebQuestionsSP and QALD-7.

<table>
<thead>
<tr>
<th>Table 1: WebQSP and QALD-7 results over Wikidata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>STAGG (2015)</td>
</tr>
<tr>
<td>Predict Edges</td>
</tr>
<tr>
<td>Pre-trained Entity Linking</td>
</tr>
<tr>
<td>Luo et al. (2018)</td>
</tr>
<tr>
<td>Our w/o ELK</td>
</tr>
<tr>
<td>Our</td>
</tr>
</tbody>
</table>

Table 2: Ablation Results on Questions Representation

We explore the contributions of various components in our model. For the ablation study, we use F1-score as our metrics, and the results are shown in Table 2 and Table 3.

Conclusions
In this paper, we utilize ELK to improve relation representation in KBQA. The knowledge is useful to extract word-level relation semantics with disambiguation. Our approach pays more attention to relational semantics of query graphs, which is an important semantic part of questions in KBQA. In this sense, we provide a new method of exploring external knowledge in KBQA.

References

Acknowledgements
This work is supported by the National Key Research and Development Program of China (2017YFC0808401) and the National Natural Science Foundation of China (61972455). Xiaowang Zhang is supported by the Poyang Young Scholars in Tianjin University (2018004XK-0032).