



# Article Introducing External Knowledge to Answer Questions with Implicit Temporal Constraints over Knowledge Base

# Wenqing Wu, Zhenfang Zhu \*, Qiang Lu<sup>D</sup>, Dianyuan Zhang and Qiangqiang Guo

School of Information Science and Electrical Engineering, Shandong Jiao tong University, Jinan 250357, China; winchywwq@163.com (W.W.); lqsdjtu@163.com (Q.L.); csdzdy@163.com (D.Z.); gqqq777@163.com (Q.G.) \* Correspondence: zhuzf@sdjtu.edu.cn; Tel.: +86-6137-9310-0702

Received: 27 January 2020; Accepted: 4 March 2020; Published: 5 March 2020



Abstract: Knowledge base question answering (KBQA) aims to analyze the semantics of natural language questions and return accurate answers from the knowledge base (KB). More and more studies have applied knowledge bases to question answering systems, and when using a KB to answer a natural language question, there are some words that imply the tense (e.g., original and previous) and play a limiting role in questions. However, most existing methods for KBQA cannot model a question with implicit temporal constraints. In this work, we propose a model based on a bidirectional attentive memory network, which obtains the temporal information in the question through attention mechanisms and external knowledge. Specifically, we encode the external knowledge as vectors, and use additive attention between the question and external knowledge to obtain the temporal information, then further enhance the question vector to increase the accuracy. On the WebQuestions benchmark, our method not only performs better with the overall data, but also has excellent performance regarding questions with implicit temporal constraints, which are separate from the overall data. As we use attention mechanisms, our method also offers better interpretability.

Keywords: knowledge base question answering; attention mechanism; external knowledge

# 1. Introduction

A knowledge base (KB) [1] stores a lot of information, which is complex and structured; they describe things (or entities) and their relationships. KB offers a more readable knowledge network for a machine and provides a more natural way to obtain abundant underlying knowledge. Freebase [2] is one such knowledge base that describes and organizes more than 3 billion facts in a consistent ontology. In fact, KB is usually represented as triples [3], such as, (subject, relation, object), where the subject and object represent entities, and the relation describes the semantic relations between subject and object. These triples are often referred to as facts and can be used for answering questions. For example, the triple (Donald Trump, President, America) can be used to answer the question "Who is the president of America". KB is increasingly used for building question answering systems [4,5].

Knowledge base question answering (KBQA) aims to analyze the semantics of natural language questions and return accurate answer from the knowledge base. At present, the methods proposed to tackle the KBQA task can be roughly categorized into two groups: semantic parsing (SP) methods and information retrieval (IR) methods. SP-based methods [6] aim to transform natural language problems into logical expressions through semantic analysis, then they are transformed into a query language such as SPARQL to retrieve the knowledge base and obtain the answer [7]. Although many SP-based methods can achieve good results in the limited domain, many important components in these works, such as vocabularies and rulesets in Combinatory categorial grammar (CCG) [8], are written manually.

Traditional semantic parses [9] require labeled training data and are limited to narrow domains with a small number of logical predicates, but manually labeling data is time-consuming and laborious. Recent studies handle these limitations through the construction of hand-crafted rules or features [6,10], schema matching [11], and using weak supervision from external resources [12].

SP-based methods are still based on symbolic logic and lack flexibility. When analyzing question semantics, it will be affected by the semantic differences between symbols. IR-based methods directly retrieve answers from the KB in light of the information conveyed in the questions. These IR-based methods can adapt better to large and complex KBs, as they do not need hand-made rules. In recent years, with the rapid development of deep learning technology, deep learning is used more and more in KBQA. On the basis of IR-based approaches, many embedding-based methods [13,14] have been proposed and have shown promising results. Compared with the traditional KBQA methods based on symbols, the KBQA method based on representation learning [14] is more robust, and it has gradually exceeded the traditional method in effect. These methods adopt various ways to encode questions and KB subgraphs into a common embedding space, then directly match them in that space, finally typically trained in an end-to-end manner.

Although the above methods have shown good results, they are not satisfactory in some specific problems, such as questions with implicit temporal constraints. In order to solve this problem, we introduce external knowledge on the basis of Bidirectional Attentive Memory Networks (BAMnet) [15], called Temporal Attention Networks (TAnet), that captures the implicit temporal information in question. We use a novel bidirectional attentive mechanism to obtain the temporal information in question in the light of external knowledge. In the experiments, we prove that our method not only shows better results in the original datasets, but also in the data with implicit temporal constraints.

We summarize the contributions of this paper as follows: (1) we introduce external knowledge to solve the questions with implicit temporal constraints; (2) due to the attention mechanism, it offers good interpretability; (3) on the WebQuestions benchmark, our method performs better, and, on the questions with implicit temporal constraints, performs excellently.

The rest of this paper is organized as follows. After introducing related works in Section 2, we describe our proposed methods in Section 3, and then we show our experimental results in Section 4. Finally, we summarize our work and future work in Section 5.

#### 2. Related Work

Generally, the solutions of KBQA can be divided into IR-based methods and SP-based methods. SP-based methods aim to transform natural language problems into logical expressions through semantic analysis, such as simple  $\lambda - DCS$  [16], query graphs [17], or executable queries, such as SPARQL [5]. Then the logical forms are executed by the corresponding technique and find the answers from the knowledge base. More recently, neural sequence-to-sequence models have been applied to semantic parsing with promising results [18,19], these methods eschew the need for extensive feature engineering.

Some studies have focused on approaches based on weak supervision from either external resources [20], schema matching [11], or using hand-crafted rules and features [6]. A series of studies has been explored to generate semantic query graphs from nature language questions, such as searching partial logical forms via an agenda-based strategy [21], exploiting rich syntactic information in nature language questions [22], using coarse alignment between phrases and predicates [23], or pushing down the disambiguation step into the query evaluation stage [24]. Notably, some SP-based methods try to exploit IR-based techniques [25] by computing the similarity between two sequences as features, utilizing a neural network-based answer type prediction model, or training end-to-end neural symbolic machine via REINFORCE [26]. However, most SP-based methods more or less rely on handcrafted rules or features, which limit their flexibility.

The general process of IR-based methods directly retrieves answers from the KB in light of the information conveyed in the questions [4,27]. Their main difference is how to select the correct answers

from the candidate set. Yao and Van Durme [28] used rules to extract question features from the dependency parse of questions and used relations and properties in the retrieved topic graph as knowledge base features. Then, the production of these two kinds of features was fed into a logistic regression model to classify the question's candidate answers into correct/wrong.

In contrast, we do not use rules, dependency parse results, or hand-crafted features for question understanding. Recently, embedding-based methods for KBQA are becoming more and more popular, Bordes et al. [29] first applied an embedding-based approach for KBQA, afterwards Bordes et al. [30] proposed the idea of subgraph embedding, which encodes more information (e.g., answer path and context) about the candidate answers. In a follow-up work [30], memory networks [31] were used to store candidate answers and could be accessed iteratively to mimic multi-hop reasoning. Different from the above methods that mainly use a bag-of-words (BOW) representation to encode questions and KB resources, Dong et al. [32] and Hao et al. [14] applied more advanced network modules (e.g., convolutional neural networks (CNNs) and long short-term memory networks) to encode questions. Das et al. [33] proposed Hybrid methods, which achieve improved results by leveraging additional knowledge sources, such as free text.

With the development of the attention mechanism [34], bidirectional attention was first proposed applied in machine reading comprehension [35,36] and was then then applied to KBQA. Most embedding-based approaches encode questions and answers independently. Hao et al. [15] proposed a cross-attention mechanism to encode questions according to various candidate answer aspects. Chen, Y et al. [17] goes one step further by modeling the bidirectional interactions between questions and a KB. This work not only modeled the interactions between questions and a KB but also introduced external knowledge to handle questions with implicit temporal constraints through an attention mechanism. As these previous works cannot handle questions with implicit temporal constraints without rules, dependency parse results, or hand-crafted features, therefore we focus on capturing the interactions between external knowledge and questions. We use deep learning (attention mechanism) to handle questions with implicit temporal constraints without rules, dependency parse results, or hand-crafted features, dependency parse results, or hand-crafted features.

Another line of related work is applying deep learning techniques for the question answering task. Grefenstette et al. [37] proposed a deep architecture to learn a semantic parser from annotated logic forms of questions. Iyyer et al. [38] introduced dependency-tree recursive neural networks for the quiz bowl game, which asked players to answer an entity for a given paragraph. Yu et al. [39] proposed a bigram model based on convolutional neural networks to select answer sentences from text data. The model learned a similarity function between questions and answer sentences. Yih et al. [40] used convolutional neural networks to answer single-relation questions on REVERB [41]. However, the system worked on relation-entity triples instead of more structured knowledge bases. We can utilize richer information (such as entity types) in structured knowledge bases.

Based on these works [42,43], we introduce external knowledge using attention mechanism to acquire temporal information on knowledge and questions, then further enhance the question vector through information to increase the accuracy of the question answer.

## 3. Model

On the basis of BAMnet, we propose a method to solve the implicit temporal constraints in the natural language question. The model is shown in Figure 1.



Figure 1. Model overview.

## 3.1. Input Module

Formally, an input question  $P = p_1, p_2, ..., p_m$  was denoted as a word embedding  $P^m$  by using a word embedding layer. Then, we encoded the question as  $L^Q$  with a bidirectional LSTM [44] (long short-term memory), where  $L^Q$  is the sequence of hidden states (i.e., the concatenation of forward and backward hidden states) generated from the BiLSTM.

## 3.2. Memory Module

We used a key-value memory network to store all candidate answers  $\{A_j\}_{j=1}^{|A|}$  (the closest to main entity and h-hops entity), which were encoded as answer types (entity type in KB), path (sequence of relations from a candidate answer to a topic entity in KB) and context (surrounding entities of a candidate in KB). Using  $[M^{kt}; M^{vt}]$ ,  $[M^{kp}; M^{vp}]$ ,  $[M^{kc}; M^{vc}]$  to represent the key-value pair of answer type, path, and context, respectively.

## 3.3. Temporal Attention Module

The temporal information implied in a question is very important for answering the question. In order to solve questions with temporal (e.g., tense) constraints, we proposed a temporal attention module to focus on temporal constraints of a question in Figure 2, which used an attention mechanism to obtain temporal information related to the external knowledge and question, and the external knowledge mentioned earlier are common tense words (such as original, previous, and former). We first used a bidirectional LSTM to encode the pre-processed tense related words  $q^T$  as  $q^t$ :

$$\vec{h}_{i}^{t} = LSTM\left(\vec{h}_{i-1}^{t}, q^{T}\right)$$

$$\vec{h}_{i}^{t} = LSTM\left(\vec{h}_{i+1}^{t}, q^{T}\right)$$

$$q^{t} = \left[\vec{h}_{i}^{t}; \vec{h}_{i}^{t}\right]$$
(1)

where the parameters of the LSTM are from those of input layer LSTM and  $q^t$ , which is fa orward and backward hidden state vector combination of LSTM.



Figure 2. Temporal attention module.

Then, we used an additive attention to gain the most relevant temporal information about the question. Putting  $q^t$  and  $L^Q$  as inputs into the additive attention to obtained implicit temporal information, and in questions  $a^t$ :

$$a^{T} = \tanh(W_{t}L^{Q} + W_{T}q^{t} + b^{T})$$
  

$$a^{t} = softmax(\sigma(W_{t}a^{T} + b^{t}))$$
(2)

where  $\sigma$  is the element-wise sigmoid function;  $W_T$  and  $W_{T'}$  are the weight matrices corresponding to the question vector  $L^Q$  and tense vector  $q^t$ ;  $W_t$  is the weight matrix corresponding to their non-linear combination;  $b^T$  and  $b^t$  are the bias vectors.

We have gained temporal information  $a^t$  in view of the question. Then, we integrated the temporal information into the question vector:

$$L^{Qt} = L^Q + a^t. aga{3}$$

By now, we have obtained temporal-aware question vector  $L^{Qt}$ , as it contains implicit temporal in light of the question.

### 3.4. Bidirectional Attention Module

The bidirectional attention module aims at catching the connection between the question and knowledge base. As not all components in a question are useful, we focused on the important parts of a question in light of the KB in Figure 3.

We used self-attention for whole new question vector about temporal  $L^{Qt}$  to obtain a question vector  $q^s$ :

$$q^{s} = BiLSTM(\left[L^{Qt}(a^{q})^{T}, L^{Qt}\right]$$

$$a^{q} = softmax(L^{q})$$

$$L^{q} = \left(L^{Qt}\right)^{T}L^{Qt}$$
(4)

where softmax is used in last dimension of  $L^q$ . Then, we put  $M^{kt}$ ,  $M^{kp}$ ,  $M^{kc}$ , and  $q^s$  as inputs into additive attention to obtain KB summary:

$$m^{y} = a^{y} \cdot M^{v_{y}}$$

$$a^{y} = softmax(\tanh(W_{1}q^{s} + W_{2}M^{k_{y}})W_{3})$$
(5)

where  $y \in \{t, p, c\}$ ,  $W_1$ ,  $W_2$  and  $W_3$  are trainable weights, and  $M^v$  is the value of the memory network.



Figure 3. The question in light of the knowledge base (KB). Mul: Vector multiplication.

So far, we already have a KB summary in light of the question. Then, we concatenated  $m^y$  as m. We multiply  $L^{Qt}$  and m, in order to obtain an attention matrix  $A^{QK}$ , which is a connection between the question and KB. We used max pooling on  $A^{QK}$  to obtain the best connection  $a^{QK}$  between the question and KB. Finally, we applied a softmax  $a^{QK}$  over to obtain  $\tilde{a}$ , which is the importance of the question about the KB.

After finding the vector, which is the question regarding the KB, we continued to obtain a vector that is the KB for the question. First, we obtained a dot-product between  $L^{Qt}$  and  $M^k$  (concatenation of  $M^{kt}$ ,  $M^{kp}$ , and  $M^{kc}$ ) to find a connection of the question and the KB attention matrix  $A^{qm}$ . We used max pooling on the last dimension of  $A^{qm}$  and normalized it to obtain the attention matrix  $A^x$ , which is the importance of the answer aspect for the candidate answer. Then, we continued to compute the question-aware KB representations  $\widetilde{M}^k$  and  $\widetilde{M}^v$  as follows:

$$\widetilde{M}^{v} = \sum_{i=1}^{3} M_{i}^{v} 
\widetilde{M}^{k} = A^{x} M^{k} 
A^{x} = softmax \left( \max_{j} \left\{ A_{j}^{qm} \right\}_{j=1} \right) \cdot$$

$$A^{qm} L^{Qt} \left( M^{K} \right)^{T}$$
(6)

Then, we enhanced the question and KB representation. We used max pooling on the last dimension of  $A^{qm}$  and normalized it to obtain an attention matrix  $A^y$ , which is the question for the KB attention matrix. Finally, we used  $A^y$ , the importance of the question about KB  $\tilde{a}$ , the KB summary, and  $M^v$ , the tense vector  $a^t$  that we obtained before to find an enhanced question representation  $\tilde{q}$ :

$$\widetilde{q} = \widetilde{aL}^{Qt} 
\widetilde{L}^{Qt} = L^{Qt} + \widetilde{a} \cdot \left(A^y \widetilde{M}^v\right) + a^t 
A^y = softmax(\max_i \left\{A_i^{qm}\right\}_{i=-1})$$
(7)

Similarly, we enhanced the KB representation  $\hat{M}^k$  that included the question information:

$$\hat{M}^{k} = \widetilde{M}^{k} + \hat{a}^{m} \cdot \left( \left( \widetilde{A}^{qm} \right)^{l} \widetilde{L}^{qt} \right) 
\hat{a}^{m} = A^{y} \widetilde{a} \qquad .$$

$$\tilde{A}^{qm} = softmax(A^{qm})$$
(8)

#### 3.5. Generalization Module

We used a one-hop attention process before answering. First, we used an attention mechanism to obtain the most relevant information from the memory. Then we renewed the question vector via a GRU (gate recurrent unit) [45]. Finally, we used a residual layer and batch normalization (BN), which can help the model performance in practice. Thus, we have:

$$\hat{q} = BN(\tilde{q} + \bar{q}) 
\bar{q} = GRU(\tilde{q}, \bar{m}) 
\bar{m} = a \cdot \tilde{M}^{v} 
a = Att_{add}^{GRU}(\tilde{q}, \hat{M}^{k})$$
(9)

#### 3.6. Answer Module

Given the question representation  $\hat{q}$  and candidate answer representation  $\{A_j\}_{j=1}^{|A|}$ , which is  $\{\hat{M}_j^k\}_{j=1}^{|A|}$ , we computed the matching score  $S(\hat{q}, \hat{M}_j^k)$  between every pair  $(Q, A_j)$  as  $S(q, a) = q^T \cdot a$ , and ranked their scores to obtain the candidate answers.

#### 4. Results

#### 4.1. Experimental Datasets

Our experiments were based on the WebQuestions dataset [46], which contains 3,778 training examples and 2,032 test examples. We further split the training data into a training set and validation set, where the training set contained 2298 examples and the validation set contained 755 examples. The validation data is randomly selected from the initial sample. The knowledge base is Freebase KB, which consists of general facts organized as subject–property–object triples. In order to prove the validity of the temporal attention module, we extracted part questions (about 11% of the WebQuestion dataset) and included implicit temporal (e.g., what did James K. Polk do before he was president?) to test what we call t-data. According to some tense related words in the question, we extracted these questions from the test data.

Following Berant et al. [23], macro F1 scores are reported on the WebQuestions test set, where macro F1 scores mean calculate F1 scores through the accuracy and recall rate of each question, and then calculate the average value of F1 scores over all questions. The reason for doing this is that the training and test sets are processed in batches.

#### 4.2. Experimental Parameters

When answering the question, not all entities and relations in Freebase will be used, so we only extracted entities and relations that existed in the dataset. The vocabulary size of words is v = 100, 797. There are 1712 entity types and 4996 relation types in the dataset. In particular, the entity may have multiple representations in Freebase, so we only used the entity that is in question. If the entity is boolean values or numbers, we used "bool" and "nums" as their types.

During the training time, we extracted a 2-hop entity, which is close to the topic entity as candidate answers. The memory network size is  $M_{\text{max}}$ =96. We used a pre-trained Glove vector to initialize word embedding with a size  $w_v$ =300. The relation embedding size  $r^e$  and hidden size h were 128. The word

8 of 13

embedding layer, question encoder side, tense words encoder and the answer encoder side in dropout rates were 0.3, 0.3, 0.3, and 0.2. The batch size was 32. In the training process, we used the Adam optimizer [47] to train the model. Initially we set the learning rate as 0.001, then reduced ten times if the performance of the model was not improved in the consecutive epoch. We stopped training if there was no promotion for 20 consecutive times on the verification set.

## 4.3. Results and Analysis

## 4.3.1. Results

We show the main results of different KBQA method in Table 1.

Methods (Baseline)	Macro F1
SP-based	
Yavuz et al. [20]	0.516
Bao et al. [48]	0.524
Yih et al. [19]	0.525
IR-based	
Hao et al. [15]	0.429
Xu et al. [4]	0.471
BAMnet and our method	
BAMnet	0.557
Our method	0.563

Table 1. Results on the WebQuestions test set.

Here, the topic entity is known. Compared to the previous KBQA method (SP-based and IR-based), our method achieved better results with an F1 score of 0.563. We can see that our method is superior to previous state-of-the-art IR-based methods and still remains competitive with SP-based methods, with the effectiveness of bidirectional interaction between question and KB.

It is important to note that compared with the state-of-the-art SP-based methods [19,48], after the introduction of external knowledge, the performance of the method is better than that of SP-based methods and beyond BAMnet. We selected the methods that have performed better in recent years for comparison. For example, Yih et al. [17] used a lot of manual rules to deal with questions with constraints and aggregations, and Bao et al. [48] directly added detected constraint nodes to query graphs to deal with questions with constraints. Yavuz et al. [30] and Bao et al. [48] trained their models on external question answering (Q&A) datasets to obtain extra supervision.

For a fairer comparison, we only show their results without training on external datasets. Although our method also introduces external knowledge, our method is to use deep learning to let the model learn autonomously instead of adding artificial rules. Additionally, the knowledge we introduce is simply words related to tenses, not data sets. Our method uses a deep learning method to handle information and has better interpretability for questions with implicit temporal constraints. Compared to IR-based methods, our method has better performance in WebQuestions. By comparison, our method is sequence to sequence, does not use any rules, and compared to the same sequence to sequence BAMnet, our model can better model questions with implicit temporal constraints. Not only does our model perform better in WebQuestions, but also in the t-data. This fully proves that it is effective to acquire the interactive information between questions and external knowledge through an attention mechanism. The results demonstrate that our method is valid.

#### 4.3.2. Ablation Study

To study the effect of temporal attention module, we conducted ablation analysis under a known topic entity. As shown in Table 2, we can see that temporal attention module is essential to the

performance. Not only does it contribute to the overall model performance but also it performs better on the temporal data we split, suggesting that the introduction of external knowledge is valid.

Tuble 2. Ablation results on the WebQuestions lest set.				
Model	Macro F1			
Full Model	0.563			
w/o Temporal attention module	0.556			

Table 2. Ablation results on the WebQuestions test set.

#### 4.3.3. Qualitative Analysis

We visualized the attention matrix  $a^t$  and checked it whether obtains temporal information in questions. Figure 4 reveals the attention heatmap generated from a test question "who is the current Ohio state senator?". We can see that the attention matrix successfully obtained the temporal information (current) from the question, so we can further strengthen the question vector through the extracted information. In other words, in the question "who is the current ohio state senator?", the word "current" implied temporal information. In the attention matrix, value of the word "current" is the highest, thus there is temporal information in the question, so we can strengthen the question vector in this way.



Figure 4. Attention heatmap generated by the temporal attention module.

In order to further prove that the introduction of external knowledge is effective, we reveal the predicted answer of our method and BAMnet from the t-data in Table 3. We divided the predicted answers into two categories, which are answer right and answer rank up. Answer right is where the predicted answers are the correct answer but not predicted in other methods, answer rank up is where the predicted answer becomes first place but includes other wrong answers. In the first type, without the temporal attention module, the model cannot capture the information of before, last, and now in the question, so generates the wrong answer, and our method obtains the right answer as the model finds the temporal information through the temporal attention module, it is important that there is no other wrong answer. In the second type, although it includes the correct answer, the model is without the temporal attention module, resulting in the generation of candidate answers, thus, the score of the correct answer is higher than other answers and ranks first. As we can see, compared with other methods without temporal attention module, our method predicts more valid answers and has better performance accuracy.

Category	Questions	Model w/o T-att	Full Model	Correct Answer
answer right	What did James K Polk do before he was president?	Governor of Tennessee,  United States Representative	Lawyer	Lawyer
	Who did Cliff Lee play for last year?	Cleveland Indians	Philadelphia Phillies	Philadelphia Phillies
	Where does Michelle Pfeiffer live now?	Santa Ana	Orange County	Orange County
answer rank up	Who was the original voice of Meg Griffin on family guy?	Mila Kunis, Lacey Chabert	Lacey Chabert, Mila Kunis	Lacey Chabert
	Where was the first microsoft headquarters located?	Washington, Albuquerque, Redmond	Redmond, Albuquerque	Redmond
	Who are the senators of New Jersey now?	Frank Lautenberg,  Bob Menendez	Bob Menendez,  John Rutherfurd	Bob Menendez

**Table 3.** Predicted answers of the full model and model w/o the temporal attention module. WhereT-att is the temporal attention module.

# 5. Conclusions

In this paper, we present a novel method that obtains temporal information from questions through introducing external knowledge for the purpose of KBQA. Specifically, we encoded external knowledge into the embedding space, obtained temporal information between the question and external knowledge through an attention mechanism, then strengthened the question vector to improve the accuracy. The results show that our method successfully captured the temporal information, and significantly outperformed previous IR-based methods, while remaining competitive with SP-based methods and BAMnet. Qualitative analysis shows that our idea of introducing external knowledge is effective. Although our method works for some questions with implicit temporal constraints, there are some limitations, that is, too many answers will be generated for some complex questions, because complex questions may contain some unknown information. Attention mechanisms have the defect of over learning. In future work, we will explore more effective ways of modeling question with implicit temporal constraints, at the same time, we will address the defects of the attention mechanism to reduce the generation of wrong answers.

**Author Contributions:** Writing—original draft, W.W.; Writing—review and editing, Z.Z.; Data curation, W.W.; Methodology, W.W.; Software, Q.G.; Funding acquisition, W.W.; Project administration, W.W.; Investigation, D.Z.; Validation, Q.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported National Social Science Foundation (19BYY076); Science Foundation of Ministry of Education of China(14YJC860042), Shandong Provincial Social Science Planning Project (19BJCJ51, 18CXWJ01, 18BJYJ04).

**Conflicts of Interest:** The authors declare no conflict of interest.

# References

- Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A Nucleus for a Web of Open Data. The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, Busan, Korea, 11–15 November 2007; pp. 722–735.
- 2. Google. Freebase Data Dumps. Available online: https://developers.google.com/freebase (accessed on 20 December 2019).
- 3. Sintek, M.; Decker, S. TRIPLE—A Query, Inference, and Transformation Language for the Semantic Web. In Proceedings of the International Semantic Web Conference, Sardinia, Italy, 9–12 June 2002; pp. 364–378.
- 4. Xu, K.; Reddy, S.; Feng, Y.; Huang, S.; Zhao, D.; Erk, K.; Smith, N.A. Question Answering on Freebase via Relation Extraction and Textual Evidence. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 2326–2336.

- Bast, H.; Haussmann, E.; Bailey, J.; Moffat, A.; Aggarwal, C.C.; De Rijke, M.; Kumar, R.; Murdock, V.; Sellis, T.; Yu, J.X. More Accurate Question Answering on Freebase. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management-CIKM '15, Melbourne, VIC, Australia, 19–23 October 2015; pp. 1431–1440.
- Abujabal, A.; Yahya, M.; Riedewald, M.; Weikum, G. Automated Template Generation for Question Answering over Knowledge Graphs. In Proceedings of the 26th International Conference on World Wide Web Companion-WWW '17 Companion, Perth, Australia, 3–7 April 2017; pp. 1191–1200. [CrossRef]
- Berant, J.; Liang, P. Semantic Parsing via Paraphrasing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 1415–1425.
- Kwiatkowski, T.; Zettlemoyer, L.; Goldwater, S.; Steedman, M. Lexical generalization in ccg grammar induction for semantic parsing. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, 27–31 July 2011; pp. 1512–1523.
- Wong, Y.W.; Mooney, R. Learning synchronous grammars for semantic parsing with lambda calculus. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 23–30 June 2007; pp. 960–967.
- Hu, S.; Zou, L.; Yu, J.X.; Wang, H.; Zhao, D. Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs (Extended Abstract). In Proceedings of the 2018 IEEE 34th International Conference on Data Engineering (ICDE), Paris, France, 16–20 April 2018; pp. 1815–1816.
- Cai, Q.; Yates, A. Large-scale semantic parsing via schema matching and lexicon extension. In Proceedings
  of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August
  2013; pp. 423–433.
- Jayant Krishnamurthy, J.; Mitchell, M.T. Weakly supervised training of semantic parsers. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, 12–14 July 2012; pp. 754–765.
- Bordes, A.; Weston, J.; Usunier, N. Open Question Answering with Weakly Supervised Embedding Models. In Formal Aspects of Component Software; Springer-Verlag: New York, NY, USA, 2014; Volume 8724, pp. 165–180.
- Hao, Y.; Zhang, Y.; Liu, K.; He, S.; Liu, Z.; Wu, H.; Zhao, J.; Barzilay, R.; Kan, M.-Y. An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, 30 July–4 August 2017; Volume 1, pp. 221–231.
- 15. Chen, Y.; Wu, L.; Zaki, M.J. Bidirectional Attentive Memory Networks for Question Answering over Knowledge Bases. *arXiv* **2019**, arXiv:1903.02188. Available online: https://arxiv.org/abs/1903.02188 (accessed on 20 December 2019).
- Liang, P.; Jordan, M.; Klein, D. Learning dependency- based compositional semantics. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 590–599.
- Yih, W.-T.; Chang, M.-W.; He, X.; Gao, J. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 1321–1331.
- Dong, L.; Lapata, M.; Erk, K.; Smith, N.A. Language to Logical Form with Neural Attention. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Association for Computational Linguistics (ACL), Berlin, Germany, 7–12 August 2016; pp. 33–43.
- 19. Jia, R.; Liang, P.; Erk, K.; Smith, N.A. Data Recombination for Neural Semantic Parsing. In Proceedings of the Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 12–22.
- Yavuz, S.; Gur, I.; Su, Y.; Srivatsa, M.; Yan, X. Improving Semantic Parsing via Answer Type Inference. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 149–159. [CrossRef]
- 21. Berant, J.; Liang, P. Imitation Learning of Agenda-based Semantic Parsers. *Trans. Assoc. Comput. Linguistics* 2015, *3*, 545–558. [CrossRef]

- Xu, K.; Wu, L.; Wang, Z.; Yu, M.; Chen, L.; Sheinin, V. Exploiting Rich Syntactic Information for Semantic Parsing with Graph-to-Sequence Model. *arXiv* 2018, arXiv:1808.07624. Available online: https://arxiv.org/ pdf/1808.07624.pdf (accessed on 20 December 2019).
- Berant, J.; Chou, A.; Frostig, R.; Percy Liang, P. Semantic parsing on freebase from question-answer pairs. In Proceedings of the EMNLP; Conference on Empirical Methods in Natural Language Processing, Chicago, IL, USA, 18–21 October 2013; pp. 1533–1544.
- 24. Hu, S.; Zou, L.; Yu, J.X.; Wang, H.; Zhao, D. Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs (Extended Abstract). *TKDE* **2018**, *30*, 1815–1816.
- 25. Liang, C.; Berant, J.; Le, Q.; Forbus, K.D.; Lao, N.; Barzilay, R.; Kan, M.-Y. Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision. *arXiv* **2016**, arXiv:1611.00020.
- 26. Williams, R.J. Simple statistical gradient following algorithms for connectionist reinforcement learning. *Mach. Learn.* **1992**, *8*, 229–256. [CrossRef]
- 27. Ben Veyseh, A.P.; Chakraborty, T.; Riedl, M.; Vydiswaran, V. Cross-Lingual Question Answering Using Common Semantic Space. In Proceedings of the TextGraphs-10: the Workshop on Graph-based Methods for Natural Language Processing, San Diego, CA, USA, 17 June 2016; pp. 15–19. [CrossRef]
- 28. Yao, X.; Berant, J.; Van Durme, B. Freebase QA: Information Extraction or Semantic Parsing? In Proceedings of the ACL 2014 Workshop on Semantic Parsing, Baltimore, MD, USA, 26 June 2014; pp. 82–86. [CrossRef]
- Bordes, A.; Chopra, S.; Weston, J. Question Answering with Subgraph Embeddings. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 615–620.
- Jain, S.; Andreas, J.; Choi, E.; Lazaridou, A. Question Answering over Knowledge Base using Factual Memory Networks. In Proceedings of the NAACL Student Research Workshop, San Diego, CA, USA, 12–17 June 2016; pp. 109–115. [CrossRef]
- 31. Weston, J.; Chopra, S.; Antoine Bordes, A. Memory networks. *arXiv* **2014**, arXiv:1410.3916. Available online: https://arxiv.org/abs/1410.3916 (accessed on 20 December 2019).
- 32. Dong, L.; Wei, F.; Zhou, M.; Xu, K. Question Answering over Freebase with Multi-Column Convolutional Neural Networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 260–269. [CrossRef]
- Das, R.; Zaheer, M.; Reddy, S.; McCallum, A.; Barzilay, R.; Kan, M.-Y. Question Answering on Knowledge Bases and Text using Universal Schema and Memory Networks. *arXiv* 2017, arXiv:1704.08384. Available online: https://arxiv.org/abs/1704.08384 (accessed on 20 December 2019).
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N. SIG. In Proceedings of the conference on empirical methods in natural language processing, Copenhagen, Denmark, September 2017, Copenhagen, Denmark, September 2017. [CrossRef]
- Seo, M.; Kembhavi, A.; Farhadi, A.; Hajishirzi, H. Bidirectional attention flow for machine comprehension. arXiv 2016, arXiv:1611.01603. Available online: https://arxiv.org/abs/1611.01603 (accessed on 20 December 2019).
- 36. Xiong, C.; Zhong, V.; Socher, S. Dynamic coattention networks for question answering. *arXiv* 2016, arXiv:1611.01604. Available online: https://arxiv.org/abs/1611.01604 (accessed on 20 December 2019).
- Grefenstette, E.; Blunsom, P.; de Freitas, N.; Hermann, M.K. A Deep Architecture for Semantic Parsing, Baltimore. In Proceedings of the ACL 2014 Workshop on Semantic Parsing, chapter A Deep Architecture for Semantic Parsing, Baltimore, MD, USA, 26 June 2014; pp. 22–27. [CrossRef]
- Iyyer, M.; Boyd-Graber, J.; Claudino, L.; Socher, R.; Hal Daume III, H. A neural ´ network for factoid question answering over paragraphs. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 633–644.
- 39. Yu, L.; Hermann, M.K.; Blunsom, P.; Pulman, S. Deep Learning for Answer Sentence Selection. *aXiv* 2014, arXiv:1412.1632. Available online: https://arxiv.org/abs/1412.1632 (accessed on 20 December 2019).
- Yih, W.-T.; He, X.; Meek, C. Semantic Parsing for Single-Relation Question Answering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, MD, USA, 22–27 June 2014; pp. 643–648.

- Fader, A.; Soderland, S.; Etzioni, O. Identifying relations for open information extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11, Edinburgh, UK, 30–31 July 2011; pp. 1535–1545.
- 42. Qian, Q.; Huang, M.; Lei, J.; Zhu, X.; Barzilay, R.; Kan, M.-Y. Linguistically Regularized LSTM for Sentiment Classification. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, 30 July–4 August 2017; pp. 1679–1689.
- 43. Lu, Q.; Zhu, Z.; Xu, F.; Guo, Q. Chinese Sentiment Classification Method with Bi-LSTM and Grammar Rules. *Data Anal. Knowl. Discov.* **2019**, *3*, 99–107. [CrossRef]
- 44. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
- 45. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555. Available online: https://arxiv.org/abs/1412.3555 (accessed on 20 December 2019).
- 46. Baudis, P.; Jan Pichl, J. Dataset Factoid Webquestions. Available online: https://github.com/brmson/datasetfactoid-webquestions (accessed on 20 December 2019).
- 47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: https://arxiv.org/abs/1412.6980 (accessed on 20 December 2019).
- Bao, J.; Duan, N.; Yan, Z.; Zhou, M.; Zhao, T. Constraint-based question answering with knowledge graph. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–17 December 2016; pp. 2503–2514.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).