What Is the Longest River in the USA? Semantic Parsing for Aggregation Questions *

Kun Xu1, Sheng Zhang1, Yansong Feng1, Songfang Huang2, Dongyan Zhao1
1Institute of Computer Science & Technology, Peking University, Beijing, China
2IBM China Research Lab, Beijing, China
{xukun, evancheung, fengyansong, zhaodongyan} @pku.edu.cn, huangsf@cn.ibm.com

Abstract

Answering natural language questions against structured knowledge bases (KB) has been attracting increasing attention in both IR and NLP communities. The task involves two main challenges: recognizing the questions’ meanings, which are then grounded to a given KB. Targeting simple factoid questions, many existing open domain semantic parsers jointly solve these two subtasks, but are usually expensive in complexity and resources. In this paper, we propose a simple pipeline framework to efficiently answer more complicated questions, especially those implying aggregation operations, e.g., argmax, argmin. We first develop a transition-based parsing model to recognize the KB-independent meaning representation of the user’s intention inherent in the question. Secondly, we apply a probabilistic model to map the meaning representation, including those aggregation functions, to a structured query. The experimental results show that our method can better understand aggregation questions, outperforming the state-of-the-art methods on the Free917 dataset while still maintaining promising performance on a more challenging dataset, WebQuestions, without extra training.

Introduction

The open domain semantic parsing (Berant et al. 2013; Cai and Yates 2013; Berant and Liang 2014) involves two main challenges: understanding the meaning of the question and instantiating the meaning against a KB. The existing semantic parsers try to model the two aspects in a nice uniform model, but usually have difficulties when: (i) simultaneously learning the meaning representations and the mappings against KB items will lead to a huge search space, thus it is often inefficient to train such a parser in open domain, e.g., taking several days to train over 3000 sentences, and difficult to adapt to other KBs, let alone retrieving multiple KBs within one query (some questions in the QALD task are answered via querying over both DBpedia and Yago (Cimiano et al. 2013)). (ii) current joint models may be appropriate for simple factoid questions, and mainly rely on hand-crafted rules to answer more complex but real questions, e.g., count

Figure 1: The meaning representation of a natural language question

in aggregation questions. However, it is really hard to enumerate all sort of rules or patterns for those challenging but common questions, e.g., argmax, argmin. A more tough case is that, as shown in Figure 1, the model may have to resolve the phrase the longest to function argmax and KB predicate RiverLength at the same time.

As indicated in Kwiatkowski et al., we find that recognizing the meaning representation of user’s intention in a question is naturally KB-independent, while the mapping phase is indeed KB-related. We thus propose a pipeline paradigm involving two steps: recognizing the KB-independent meaning representations inherent in the questions, and then converting the meaning representations into KB-related structured queries. Firstly, we use the basic simple Lambda Dependency-Based Compositional Semantics (λ-DCS) as the KB-independent meaning representation language, except that each predicate is a natural language phrase, as shown in Figure 1. We build an efficient transition-based semantic parser to perform the structure prediction, which recognizes and introduces functional phrases into the meaning representation. Secondly, we propose a probabilistic model to determine the probabilities of mapping between natural language phrases and KB items as well as aggregation functions. Specifically, we propose a distant supervision method to solve the mappings between natural language phrases and KB predicates with their preferred aggregation functions by mining the context of Wikipedia pages and WordNet. The experimental results showed our approach outperforms the state-of-the-art parsers by 3% on Free917, and can still maintain a competitive result of 42.6% on WebQuestions, without any extra training.

The Transition-based Semantic Parser

Our transition-based parser will take a natural language sentence as input and output a tree-like meaning representation of the sentence, where phrases of different types and the structures among them are recognized. Following the convention in the transition-based dependency parsing, our parser

*This work was supported by the National High Technology R&D Program of China (Grant No. 2012AA011101, 2014AA015102), National Natural Science Foundation of China (Grant No. 61272344, 61202233, 61370055) and the joint project with IBM Research.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
includes a stack of partial derivations, a queue of incoming words, and five types of actions: Shift-\(X\) (\(X\in T = \{\text{entity, relation, category, function, NIL}\}\), where entity and categ-
y nodes correspond to the unary predicates of KB, relation nodes correspond to the binary
predictics of the KB, function nodes correspond to aggregation functions in simple \(\lambda\)-DCS, e.g., count, argmax, and argmin, and NIL are those that have no correspond-
ing items in the KB), Left-Reduce, Right-Reduce, Combine-\(X\), Combine-\(Y\). (1) The Shift-\(X\) action removes the incoming word in the queue \(Q\), pushes it on top of
the stack \(S\) as a new node, and assigns it with a semantic type \(X\in T\). (2) The Combine-\(X\) action pops the top two NIL
nodes from the stack, combines them into a new node assigning a semantic type \(X\) from the set \(T\), and then pushes the new
node onto the stack. (3) The Combine-\(Y\) action puts the top two non NIL nodes from the stack \(S\), combines them by
operation \(Y\in \{\text{Join, Intersection, Aggregation}\}\) and pushes the new node onto the stack. These operations correspond to
the operators in the \(\lambda\)-DCS. (4) The Left-Reduce action pops the top two nodes from the stack and pushes the top
node back onto the stack. (5) The Right-Reduce action puts the top node back from the stack.

### Instantiating Meaning Representation

We apply a probability model to map the meaning representa-
tion \(Q_{ind}\), which consists of \(n\) triples and \(m\) aggregation
phrases, to a structured query \(Q_d\), and we made necessary independent assumptions, approximating \(P(Q_d|Q_{ind})\) as:

\[
P(Q_d|Q_{ind}) \approx \frac{1}{\prod_{i=1}^{n} P(s_i|\text{ind}) \cdot P(o_i|\text{ind})} \cdot \prod_{j=1}^{m} P(d_j|\text{ind})
\]

where the \((s, p, o)\) corresponds to the three parts of a query triple: the subject \(s\), predicate \(p\) and object \(o\), and the \((f, a)\)
corresponds to the two parts of the aggregation function: the operation \(f\) and the argument \(a\). In practice, we use the Free-
base search API\(^1\) to compute the probabilities of mapping the subject and object phrase. We apply the Naive Bayes model
to compute the probability of mapping the relation phrase. To compute \(P(d_j|f_a|\text{ind})\), we propose a distant supervision method to collect co-occurrences of natural language phrases and KB predicates as well as their preferred aggregation functions.

### Experiment

#### Experimental Setup

We evaluate our system on two datasets, the Free917(Cai and
Yates 2013) and WebQuestions(Berant et al. 2013). Specifi-
cally, the Free917 dataset contains 917 questions annotated with
logical forms grounded to Freebase. The WebQuestions dataset
contains 5,810 question-answer pairs, with the same training/testing split with previous work.

#### Main Results

We compare our system with the current state-of-the-art open
domain systems, (Berant and Liang 2014) and (Yao and

\[\begin{array}{|c|c|c|c|}
\hline
& BcFL14 & Yao14 & Our-A & Our-A \\
\hline
Free917 & 68.5\% & - & 71.5\% & 71.7\% \\
WebQuestions & 39.9\% & 42.0\% & 41.1\% & 42.6\% \\
\hline
\end{array}\]

Table 1: Results on test sets of Free917 and WebQuestions

Van Durme 2014). We also include a variant of our model
(Our-A), where we do not consider aggregation operations
during the second phase. In terms of step performances, our
model archives 81.2\% for parsing accuracy in the first step,
and 86.7\% of instantiation accuracy in the second step, on Free917. And for overall performances, as shown in table 1,
our system obtains a system accuracy of 71.7\% on Free917,
outperforming the state-of-the-art systems. On the WebQuest-
ions dataset with much more predicates than Free917, our
system is still able to maintain a competitive performance,
achieving a \(F\)-measure of 42.6\% without extra training on
WebQuestions. This result shows the novelty of our semantic
parser particularly on the meaning recognizing.

### References

Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and
ISWC/ASWC, 722–735.

Bao, J.; Duan, N.; Zhou, M.; and Zhao, T. 2014. Knowledge-based
question answering as machine translation. In Proceedings of the
52nd Annual Meeting of the Association for Computational Linguis-
tics (Volume 1: Long Papers), 967–976. Baltimore, Maryland: Asso-
ciation for Computational Linguistics.

In Association for Computational Linguistics (ACL).

Berant, J.; Chou, A.; Frostig, R.; and Liang, P. 2013. Semantic
parsing on freebase from question-answer pairs. In EMNLP, 1533–
1544.

Bollacker, K. D.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J.
2008. Freebase: a collaboratively created graph database for structur-

Conference on Lexical and Computational Semantics (*SEM),
Cimiano, P.; Lopez, V.; Unger, C.; Cabrio, E.; Ngomo, A.-C. N.; and
Walter, S. 2013. Multilingual question answering over linked data

Collins, M., and Roark, B. 2004. Incremental parsing with the percep-
tron algorithm. In ACL, 111–118.

Kwiatkowski, T.; Zettlemoyer, L. S.; Goldwater, S.; and Steedman,
M. 2010. Inducing probabilistic ccg grammars from logical form
with higher-order unification. In EMNLP, 1223–1233.

Scaling semantic parsers with on-the-fly ontology matching. In
EMNLP, 1545–1556.

of semantic knowledge. In WWW, 697–706.

Unger, C., and Cimiano, P. 2011. Pythia: Compositional meaning
construction for ontology-based question answering on the semantic
web. In Proceedings of the 16th International Conference on Natural
Language Processing and Information Systems, NLDB’11, 153–160.
Berlin, Heidelberg: Springer-Verlag.

and Cimiano, P. 2012. Template-based question answering over rdf
data. In WWW, 639–648.

Yahya, M.; Berberich, K.; Elbassouoni, S.; Ramanath, M.; Tresp, V.;
and Weikum, G. 2012. Natural language questions for the web of

Yao, X., and Van Durme, B. 2014. Information extraction over
structured data: Question answering with freebase. In Proceedings of
ACL.

\(^1\)https://developers.google.com/freebase/