Context-dependent Syntax Tree Based Neural Model on SParC

Bo Pang, Tao Yu, Dragomir Radev

LILY Lab, Yale University

LILY Lab

Introduction

We present SParC, a dataset for cross-domain Semantic Parsing in Context. It consists of 4298 coherent question sequences(12k+ individual questions annotated with SQL queries), obtained from controlled user interactions with 200 complex databases over 138 domains. We provide an in-depth analysis of SParC and show that it introduces new challenges compared to existing datasets. SParC (1) demonstrates complex contextual dependencies, (2) has greater semantic diversity, and (3) requires generalization to new domains due to its cross-domain nature and the unseen databases at test time. We experiment with two state-of-the-art text-to-SQL models adapted to the context-dependent, cross-domain setup. My work includes: 1. Collect data 2.Collaborate with Tao to define and build a baseline model to experiment on our task. And the future work includes: 1. build up our large-scale and cross-domain corpus for new dialogue system task. 2. Develop baseline models on corpus to prove usability.

Data Collection

We create the SParC dataset in four states: selecting interaction goals, creating questions, annotating SQL labels, and reviewing the data.

Interaction goal selection To ensure thematic relevance within each question sequence, we use questions in the original Spider dataset[1] as the thematic guidance for constructing meaningful query interactions.

Question creation 15 college students with SQL experience were asked to come up with sequences of inter-related questions to obtain the information demanded by the interaction goals.

SQL annotation After creating the questions, each annotator was asked to translate their own questions to SQL queries. All SQL queries were executed on Sqlite Web to ensure correctness.

Data review and post-process We asked students who are native English speakers to review the annotated data. Each example was reviewed at least once. The students corrected any grammar errors and rephrased the question in a more natural way if necessary.

Turn #	CD-Seq2Seq	SyntaxSQL-con
1 (422)	31.4	38.6
2 (422)	12.1	11.6
3 (270)	7.8	3.7
$\geq$ 4 (89)	2.2	1.1

Table 3. Performance stratified by question turns on the development set.

Model	Question Match		Interaction Match	
	Dev	Test	Dev	Test
CD-Seq2Seq	17.1	18.3	**6.7**	**6.4**
SyntaxSQL-con	**18.5**	**20.2**	4.3	5.2
SyntaxSQL-inp	15.2	16.9	0.7	1.1

Table 2. Performance of various methods over all questions (question match) and all interactions (interaction match)

Dataset	Context	Resource	Annotation	Cross-domain
SParC	✓	database	SQL	✓
ATIS (Hemphill et al., 1990; Dahl et al., 1994)	✓	database	SQL	✗
Spider (Yu et al., 2018c)	✗	database	SQL	✓
WikiSQL (Zhong et al., 2017)	✗	table	SQL	✓
GeoQuery (Zelle and Mooney, 1996)	✗	database	SQL	✗
SequentialQA (Iyyer et al., 2017)	✓	table	denotation	✓
SCONE (Long et al., 2016)	✓	environment	denotation	✓

Table 1: Comparison of SParC with existing existing semantic parsing dataset

D_1 : Database about student dormitory containing 5 tables.
C_1 : **Find the first and last names of the students who are living in the dorms that have a TV Lounge as an amenity.**

Q_1 : How many dorms have a TV Lounge?
S_1 :
```
SELECT COUNT(*) FROM dorm AS T1 JOIN has_amenity
AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity
AS T3 ON T2.amenid = T3.amenid WHERE
T3.amenity_name = 'TV Lounge'
```
Q_2 : What is the total capacity of these dorms?
S_2 :
```
SELECT SUM(T1.student_capacity) FROM dorm AS T1
JOIN has_amenity AS T2 ON T1.dormid = T2.dormid
JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid
WHERE T3.amenity_name = 'TV Lounge'
```
Q_3 : How many students are living there?
S_3 :
```
SELECT COUNT(*) FROM student AS T1 JOIN lives_in
AS T2 ON T1.stuid = T2.stuid WHERE T2.dormid IN
(SELECT T3.dormid FROM has_amenity AS T3 JOIN
dorm_amenity AS T4 ON T3.amenid = T4.amenid WHERE
T4.amenity_name = 'TV Lounge')
```
Q_4 : Please show their first and last names.
S_4 :
```
SELECT T1.fname, T1.lname FROM student AS T1 JOIN
lives_in AS T2 ON T1.stuid = T2.stuid WHERE
T2.dormid IN (SELECT T3.dormid FROM has_amenity
AS T3 JOIN dorm_amenity AS T4 ON T3.amenid =
T4.amenid WHERE T4.amenity_name = 'TV Lounge')
```

Figure 1. One question sequence from the SParC dataset.

Goal Difficulty	CD-Seq2Seq	SyntaxSQL-con
Easy (483)	35.1	**38.9**
Medium (441)	7.0	**7.3**
Hard (145)	**2.8**	1.4
Extra hard (134)	**0.8**	0.7

Table 4.Performance stratified by question difficulty on the development set

Reference

[1] Tao et al. 2018c. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-sql task. In EMNLP

[2] Tao et al. 2018b. Syntaxsqlnet: Syntax tree networks for complex and cross-domain text-to-sql task. In Proceedings of EMNLP. Association for Computational Linguistics

[3]Suhr et al. 2018. Learning to map context-dependent sentences to executable formal queries. In NAACL.

Method and Experiment

To benchmark the difficulty of our dataset, we experiment with two state-of-the-art semantic parsing models extended to the cross-domain, context-dependent setup. Here I will present one of them, SyntaxSQL[2].

SyntaxSQLNet is the syntax tree based neural model for the complex and cross-domain Spider text-to-SQL task introduced by Yu et al. The model employs a SQL-specific syntax tree-based decoder with SQL generation path history and table-aware column attention encoders. Since it only handles single-turn complex questions, we provide the model with a short interaction history so as to consider context dependencies.

The results of our experiment on SyntaxSQL as well as Seq2Seq[3], another model we experiment on are shown in Table 1~4. We use the exact set match metric to compute the accuracy between gold and predicted SQL answers. As Yu et al.[1] , we decompose predicted queries into different SQL clauses such as SELECT, WHERE, GROUP BY, and ORDER BY and compute scores for each clause using set matching separately. The final exact set matching score is 1 for each question only if all predicted SQL clauses are correct and i for each interaction only if there is an exact match for every question in the interaction.

Conclusion

In this work, we introduced SParC, a large scale dataaset of conversational interactions over a number of databases in different domains. The data features a diverse range of semantic content and contextual dependencies between questions in the same interaction. The associated task introduces a challenge in mapping context-dependent questions to SQL queries in unseen domains. We experiment with two strong semantic parsing systems and, in general, observe relatively low performance, which suggests strong challenges for future research. This, together with our detailed data analysis, examplifies the complexity of the data.

Acknowledgement

1.The work is currently under review for ACL 2019, with a paper named SParC: Cross-Domain Semantic Parsing in Context.

2.I would like to acknowledge Prof. Dragomir Radev for his supervision of my project. Also, I feel grateful to Tao Yu for his guidance and help in the through the progress.