How Would You Say It? Eliciting Lexically Diverse Data For Supervised Semantic Parsing

Abhilasha Ravichander, Thomas Manzini, Matthias Grabmair, Graham Neubig
Jonathan Francis, Eric Nyberg

Language Technologies Institute, Carnegie Mellon University

Motivation

Building dialogue interfaces for real-world scenarios often entails training semantic parsers starting from zero examples.

How do we build datasets that capture the variety of ways in which users phrase their queries?

Introduction

- Orienting a dialogue-capable intelligent system is accomplished by training its semantic parser with utterances that capture the nuances of the domain.
- Previous work[1] proposes a methodology that generates example natural language utterances for logical forms, which are then paraphrased by crowdworkers.
- Three main limitations:
  - Seed utterances may induce bias towards the language of the canonical utterance, specifically with regards to lexical choice.
  - Generic grammar suggested cannot be used to generate all the queries we may want to support in a new domain.
  - No check on the correctness or naturalness of the canonical utterances themselves, which may not be logically plausible.

Method

Lexicon — Manual

FOOD[bread] →
FRIDGE[refrigerator] →
FOOD_STATE[expired state] → expired state
FOOD_STATE[count] → count

Grammar — Manual

FRIDGE[x] → FRDG_NP["x"]
FOOD[x] → FD_NP["x"]
FD_NP[x] in the FRDG_NP[l] → FD_SING[(None,l,x,"getFood")]
what is the FOOD_STATE[r] of the FD_SING[x] → Q[(None,None,x,"checkState-"+"r")]
Q[x] → ROOT[x]

Canonical & Logical Forms — Automatic

what is the expired state of the in the ?
→ ROOT["(None, 'refrigerator', 'bread', 'getFood>checkState-expired state')"]

Crowdsourced Paraphrases — Crowdsourced

ROOT["(None, 'refrigerator', 'bread', 'getFood>checkState-expired state')"]
"is the bread in the refrigerator moldy?"
"did the bread go bad?"
"is the bread in the refrigerator expired yet?"
"is the bread in the fridge bad?"

- 195 Turker Sessions over three days.
- Each logical form shown to 5 Turkers for paraphrasing.
- Each Turker asked to enter a total of 60 paraphrases.
- 8294 unique paraphrases collected over 948 logical forms.

Results and Discussion

Representation	Vocab Size	TTR	Lexical Overlap
Text (Wang et al., 2015)	291	.044	5.50
Text-Image (ours)	**438**	**.066**	**4.79**

Table 1. Comparison of data creation methodology of [1] and this work.

Dataset	NL Types	MR Types	NL/ MR Ratio
GEO	283	148	1.91
ATIS	934	489	1.91
JOBS	387	226	1.71
OVERNIGHT	1422	199	7.14
SMARTHOME (Ours)	1356	83	**16.33**

Table 2. Number of word types in the language compared to the logical form. Larger ratio indicates more lexical diversity for the same complexity of the logical form.

System	SMARTHOME	OVERNIGHT	GEO
Jaccard	18.0%	24.82%	40.7%
Neural Reranker	30.3%	41.91%	60.2%
Seq2Seq[2]	**42.1%**	75.8%	85.0%

Table 3. Test accuracy results of different systems on the SMARTHOME dataset as compared to OVERNIGHT and GEO

Most errors stem due to the following types of queries in SMARTHOME, which are not present in OVERNIGHT or GEO:

- Singular and plural forms (eg. radio/radios)
- Unseen semantically equivalent phrases (eg. *Does Bob not have energy* should be mapped to the logical form for *Is Bob tired*).
- Indirect phrases (eg. D*o i need to change the lights in the living room* not mapping to the logical form for living room lights not working correctly).
- Complementary Terms (on/off).

Conclusion

- A mixture of text and images elicits more lexically diverse paraphrases from crowdworkers with limited loss of correctness.
- SMARTHOME dataset for semantic parsing.
- Domain, cardinality and complementary formulations also contribute to difficulty.

References

1) Wang, Yushi, Jonathan Berant, and Percy Liang. "Building a Semantic Parser Overnight." ACL (1). 2015.
2) Jia, Robin, and Percy Liang. "Data recombination for neural semantic parsing.", ACL 2016.
3) Jekaterina Novikova, Oliver Lemon, and Verena Rieser. Crowd-sourcing nlg data: Pictures elicit better data. In Proceedings of the 9th International Natural Language Generation conference, 2016.

Acknowledgements

This research is supported by Robert Bosch LLC as part of a joint project with CMU on the development of a context-aware, dialog-capable personal assistant.

Dataset: https://github.com/oaqa/resources
E-mail: {aravicha, tmanzini}@cs.cmu.edu