Learning and Parsing Dependency-Based Compositional Lexicon For Broad-Coverage Language Understanding

ALAN ZAFFETTI

azaffett@umass.edu

Abstract

I extract logical forms for words based on their position in dependency graphs rendered from the Penn Treebank. Weights for a probability model are learned for words over Stanford Dependencies relations (2008), and token tags. With a basis in linguistic theory, I map these classes to types, and their respective *underspecified logical forms*. I apply this to implementing a weighted CYK chart-parser, generating all semantic parses in $\mathcal{P}$time.

Introduction

Semantic parsing maps input in the form of natural language (NL) to outputs in complete, formal meaning representations (MR) (Wong, Mooney, 2007). While the motive is ultimately to prepare a model excelling at the task of *semantic parsing*, we must first find a way to efficiently learn words and their meanings. Therefore, there are a few parts to this project:

1. Learn the words;
2. Construct a probability model;
3. Implement a chart parser;
4. Present results, quantitative analysis such as *coverage*, and correctness.

For part (1), I apply a “bootstrapping” approach (e.g. Swier, Stevenson, 2004), which includes an expert lexicon on *underspecified logical forms*, given the 38 *dependency relation classes* (see: appendix A; de Marneffe, Manning, 2008). The lexical rules are similar to previous work involving handcrafted lexical templates (Zettlemoyer, Collins, 2005).

A dependency relation can be read directly from a dependency graph. However, to generate the dependency graphs, must convert examples from a phrase-structure treebank, introducing error (de Marneffe, MacCartney, Manning, 2006). Since this project is not about generating dependency graphs, we are using the Stanford Dependencies phrase-structure conversion tool available online[1].

For part (2), after we construct tree-to-graph translations, we learn a structured probability model, following Zettlemoyer, Collins (2005) by taking *word identity* as the lone feature type. The result is a distribution to each dependency relation given each word. This distribution gives great insight into the semantic role for each sense of a word among the example sentences. By mapping these against the expert lexicon, we generate the weighted logical forms for individual words.

Unlike prior work, I will not be learning an augmented grammar (e.g. CFG; Wong, Mooney, 2007), but learning from simpler relations in the dependency grammar, and an expert corpus. Model parsing will rely entirely on typed lexicon; applying rules functionally based on a simple type system, and the principle of *compositionality* (Heim, Kratzer, 1998), which states “*the meaning of a complex expression is determined by meanings of its parts, and the rules to combine them*.”

For (3), implementing the chart-parser will follow naturally. The algorithm is a slight adaptation from CYK, designed to work over types rather than POS tags, following Zettlemoyer, Collins (2005). A weighted version of the chart parser will derive top-K logical forms from the raw sentences.

For (4), since construction of a test set from sentences to logical forms is tedious and time consuming, I will not be able to assess my models accuracy against many test samples. Most of my results section then will be devoted to analyzing model *coverage* against linguistically interesting constructions. If time permits, I will attempt to create an

[1] https://www.github.com/dmcc/PyStanfordDependencies

extensional system for “fill in” logical expressions using a structured database. This will allow me to try my model at question answering, and open the door for future work improving the model using question and answer pairs alone (e.g. Liang, Jordan, Klein, 2011).

References

[1] Marie-Catherine de Marneffe and Christopher Manning. 2008. Stanford typed dependencies representation. In *Proceedings of COLING’08, Workshop on Cross-Framework and Cross-domain Parser Evaluation*, pages 1–8.

[2] Wong, Y. W. and Mooney, R. 2007. Learning synchronous grammars for semantic parsing with lambda calculus. In *Proceedings of the Association for Computational Linguistics*.

[3] Robert S. Swier and Suzanne Stevenson. 2004. Unsupervised semantic role labeling. In *EMNLP*.

[4] Zettlemoyer, L.S. and Collins, M. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorical grammars. In *Proc. 21st Conf. Uncertainty Artif. Intell.* (UAI-05), pp. 658–666

[5] de Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC-06*.

[6] P. Liang, M. I. Jordan, and D. Klein. 2011. Learning dependency-based compositional semantics. In *Association for Computational Linguistics (ACL)*, pages 590–599.