Towards Semantic Parsing in Dynamic Domains

Kyle Richardson and **Jonas Kuhn**
Institute for Natural Language Processing
University of Stuttgart
{kyle,jonas}@ims.uni-stuttgart.de

1 Overview

We describe ongoing work in the area of semantic parsing, which is an emerging subfield in NLP that concerns the task of mapping sentences to formal semantic representations. Recent work in this area has focused on using data-driven methods for learning this mapping, both in a supervised setting and in more complex ambiguous learning settings [Mooney, 08]. In the latter learning scenarios, training examples might be given with several possible target semantic representations, the bulk of which don't relate directly to the particular sentence but are instead part of a broader grounded perceptual context. In such a setting, the aim is to model language as being 'situated' in a potentially wide range of observable events.

Well known work by [Chen et al. 2008] on the sportscaster corpus looks at interpreting soccer commentary in ambiguous contexts where several closely occurring (grounded) events are taking place. For example, a naive language observer might hear commentary such as *The purple goalie kicks out to purple3* in the context of several different actions, and at first be uncertain about which event in view is being described. They present a novel bottom-up learning method for accurately parsing unseen game commentary to symbolic semantic representations by 'observing' ambiguous training games, which has inspired a number of subsequent learning studies.

As pointed out by [Bordes et al. 2010], however, the sportscaster corpus has many shortcomings, most notably its lack of lexical ambiguity and small size. *Contexts* are limited to information about events occurring within a very crude window of time around each comment. In a dialogue setting having more background information might be essential. For example, knowing the referent of 'he' in the sentence *he is cooking in the kitchen* requires having knowledge of which individuals are in the kitchen at this time. Similarly, such contextual information is useful for detecting and learning inferential patterns in language.

The *Grounded World* corpus described in [Bordes et al. 2010] gets at some of the issues, and is a set of English descriptions situated within a virtual house. Sentences in the corpus are often ambiguous and employ pronouns, which must be resolved using information about the state of the house (e.g. the location of objects). The corpus, however, was designed largely for doing named entity recognition, and learning is done in a supervised fashion. We describe an extension to this corpus that looks at learning to interpret these descriptions in an ambiguous learning setting.

2 Grounded World*

Utterance: *while he is sleeping in the bedroom*
Original Annotation: - <friend> - <sleep> - - <bedroom>
Observables*: bring(friend,water,toLoc(bedroom))
get(baby,videogame) **sleep(friend,loc(bedroom))**
World State:
location:**bedroom**<'bed', 'closet', '**friend**', …>
location:kitchen<'baby', 'fridge', 'cat',…..>

Figure 1: training example from Grounded World*

The original Grounded World corpus consists of 50k (automatically generated) training sentences, paired with a target set of named entities and a world state description, and 30k testing examples. Inside the simulated house is a fixed set of objects, including, for example, a set of *actors* (e.g. 'father', 'brother'), and a set of *furniture pieces* (e.g. 'couch', 'table'). There is also a fixed set of 15 events, such as *eating* , *bringing*, and *drinking*. For our study, we used a small subset of 7k examples from the training set, and modified the sentences to have syntactic alternations and paraphrases not seen in the initial corpus. The original annotations were expanded to normalized semantic representations, and using the world state information we produced a set of distractor events (the *observables*) intended to represent the background knowledge or uncertainty an observer might have about related or simultane-

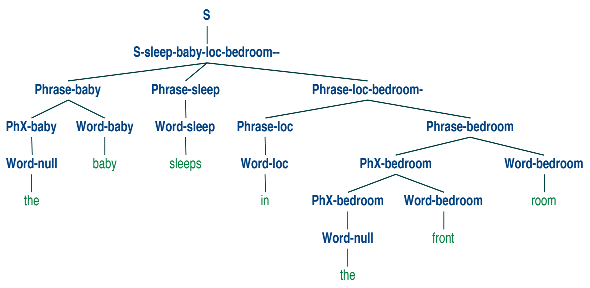

Figure 2: Grounded World* example parse

ous events in view. Figure 1 shows a training example in our Grounded World* corpus, alongside the original annotation. The utterance is *situated* is three separate observable events, two of which are contradictory and represent an observer's uncertainty about whether the friend is involved in the *sleeping* or *bringing* event.

Expanding the relations from the original corpus and *situating* them within larger ambiguous contexts makes the learning task much harder. Given a set of training examples in this narrow domain, we aim to learn, merely from ambiguous observation, how to map novel sentences about the house to their correct semantic representations.

3 Learning

One trend in Semantic Parsing has been to use learning methods that assign rich structure to the target semantic representations, which can be used for finding alignments with latent structures in the language. In many available datasets, the target semantic representations have corresponding *semantic grammars* that produce tree representations. Using ideas from [Wong 2007], [Chen 2008] uses statistical alignment-models for finding alignments between production rules in the semantic grammars and the corresponding words or phrases in the language. In a similar spirit, [Borschinger et al. 2010] recasts the problem in terms of an unsupervised PCFG induction problem, and he develops a technique for automatically generating large PCFGs from the semantic relations in the sportscaster data. In such a setting, the target semantic relations are the *S-Nodes* in the grammar, and the arguments of the relations and relation names are the constituent phrases (in all possible orders) consisting of pre-terminals that correspond to domain concepts. Words in the training data are uniformly assigned to all preterminals and the PCFG weights are learned using EM training and the ambiguous contexts as filters.

4 Experiments

In a pilot study to test our extension to the corpus, we adopt the grammar induction technique used in [Borschinger et al. 2010]. We automatically generate a large PCFG using the total semantic relations in our dataset which includes information about the ambiguous contexts. Following the experimental design in [Chen et al. 2008] and [Borschinger et al. 2010], we perform cross validation by making 4 splits in our 7k sentence set (5k for training, and 2k for testing). We then train on each set using the Inside-Out Algorithm, and evaluate by parsing the remaining unseen sentences and compare each S-node relation to a gold standard.[1] In our initial experiments, we don't consider the world state information, and instead resolve pronouns by choosing the most probable analysis observed in the training.

An example analysis produced after training is provided in figure 2, where the derived S-node relation is *sleep(baby, loc(bedroom))*. In the initial experiments, we achieve an average precision of 77.6 % over the four splits. Most errors relate to pronoun resolution, which had an average accuracy of 37.4%. Further work will look at building a parser that considers world information, building on insights from [Schuler 2001].

References

Bordes, A. et al. 2010 Towards Understanding Situated Natural Language. in *Proc. of the 13th AISTATS*

Borschinger, B. et al. 2011 Reducing Grounded Learning Tasks to Grammatical Inference in *Proc. of EMNLP 2011*

Chen, D. and R. Mooney 2008. Learning to Sportscase: A Test of Grounded Language Acquisition. in *Proc. of ICML*

Mooney, R 2007. Learning for Semantic Parsing. in *Proc. of the 8th CICLing*

Schuler, W. 2001. Computational Properties of environment-based disambiguation . in *Proc. ACL*

Wong, Y. et al. 2008. Learning for Semantic Parsing with Statistical Machine Translation. in *Proc. of HLT/NAACL*

[1] we used Mark Johnson's CKY and Inside-Out implementation available at http://web.science.mq.edu.au/ mjohnson/Software.htm.