Lessons from Efforts to Automatically Translate English to Knowledge Representation Languages

Chitta Baral

Faculty of Computer Science and Engineering
Arizona State University
Tempe, AZ 85287-8809
chitta@asu.edu

Abstract. Our long term goal is to develop systems that can "understand" natural language text. By "understand" we mean that the system can take natural language text as input and answer questions with respect to that text. A key component in building such systems is to be able to translate natural language text into appropriate knowledge representation (KR) languages. Our approach to achieve that is inspired by Montague's path breaking thesis (1970) of viewing English as a formal language and by the research in natural language semantics. Our approach is based on PCCG (Probabilistic Combinatorial Categorial Grammars), λ-calculus and statistical learning of parameters. In an initial work, we start with an initial vocabulary consisting of λ- calculus representations of a small set of words and a training corpus of sentences and their representation in a KR language. We develop a learning based system that learns the λ-calculus representation of words from this corpus and generalizes it to words of the same category. The key and novel aspect in this learning is the development of Inverse Lambda algorithms which when given λ-expressions β and γ can come up with an α such that application of α to β (or β to α) will give us γ. We augment this with learning of weights associated with multiple meanings of words. Our current system produces improved results on standard corpora on natural language interfaces for robot command and control and database queries. In a follow-up work we are able to use patterns to make guesses regarding the initial vocabulary. This together with learning of parameters allow us to develop a fully automated (without any initial vocabulary) way to translate English to designated KR languages. In an on-going work we use Answer Set Programming as the target KR language and focus on (a) solving combinatorial puzzles that are described in English and (b) answering questions with respect to a chapter in a ninth grade biology book. The systems that we are building are good examples of integration of results from multiple sub-fields of AI and computer science, viz.: machine learning, knowledge representation, natural language processing, λ- calculus (functional programming) and ontologies. In this presentation we will describe our approach and our system and elaborate on some of the lessons that we have learned from this effort.