COSI 217: Topics in Natural language processing: meaning representation and parsing, deep learning

Spring 2019

Instructor: Nianwen Xue

- http://www.cs.brandeis.edu/~xuen
- Email: xuen@brandeis.edu
- Office: 253 Volen
- Office hours: Thursday, 2pm-4:00pm

Teaching Assistant:

Time: 2:00-4:50pm, Tuesday

Location: Goldsmith Math 116

Class website: http://latte.brandeis.edu

Prerequisites: (any one of) COSI 114b, 134a, 137b or instructor consent.

Course objective:

Innovations in deep learning techniques, most notably Transformer-based neural architecture and pre-trained context-sensitive embeddings like ELMo and BERT, have led to many exciting breakthroughs in the past year. These include improved accuracy in Machine Translation, Question Answering, Natural Language Inference, and Sentiment Analysis, to name a few. Even syntactic parsing, after being stagnant for a few years, has seen a major breakthrough. On a well-established benchmark, the Penn Treebank, the parsing accuracy for English is approaching 96% (Kitaev and Klein, 2018), matching or even surpassing human inter-annotator agreement.

This does not mean all the NLP problems are solved however. For example, understanding the meaning by the machine is still an extremely hard problem. One specific form of understanding is to parse sentences into a meaning representation that enables lexical and logical inferences to help downstream applications. The objective of this seminar is to take stock of the state of the art in semantic parsing, study the latest technical advances in deep learning, and explore ways in which these latest deep learning techniques can be applied to advance the state of the art in semantic parsing.

Upon completion of the course, the student will be familiar with different meaning representations and automatic semantic parsing techniques. The student will also have learned to conduct research in this area, which typically involves identifying a technical challenge, developing linguistic representations and statistical models to address it, and writing it up as a technical paper. The student
will also have learned to synthesize existing research on a research topic, identify shortcomings or weaknesses with existing approaches, and propose solutions that are improvements or even disruptive improvements over existing approaches.

Course Description:

With the advancement of syntactic parsing coming to an equilibrium, much of the attention in the field of NLP has shifted to meaning representation and parsing. There are, however, divergent opinions on what a meaning representation should look like and what are the critical components to a meaning representation. In this course, we will focus on examining two general approaches to meaning representation that reflect different priorities on the part of their proponents. The first approach emphasizes meaning representation as a logical inference system, focusing on the proper representation of quantification, negation, tense, etc, constructs that are critical to logical inference. The second approach considers meaning representation as a way to preform (similarity-based) lexical inference, focusing on the proper representation of the predicate-argument structure as well as word senses, entities, and relations. Just as with meaning representations, there are also different approaches to automatic meaning representation parsing (or “semantic parsing”). One line of semantic parsing research has been to develop end-to-end, domain-dependent semantic parsing systems. Semantic parsers of this type take as input a natural language query and a knowledge base, and parses the query into a meaning representation that can be executed to query the knowledge base. The meaning representation used is typically some variation of first-order logic. An alternative approach is the domain-independent approach that parses a natural language sentence into a meaning representation that is not tied to any particular application, but can rather be used for many applications. An example of this approach is the Abstract Meaning Representation (AMR), a graph-based meaning representation that has gained traction in the field of NLP in recent years. We will discuss how to extend AMR to different languages and work towards a Universal Meaning Representation that applies to a diverse set of languages and supports both lexical and logical inference. We will also devote a substantial amount of time going over computational approaches to AMR/UMR parsing, which are rapidly evolving.

This is going to be a seminar-style research-oriented course where students will work on a project, and report their results in the format of a typical research paper that is presented in conferences such as ACL, NAACL, and EMNLP. Students of the course are expected to participate in two ways. The student is expected to identify a research project in the area early on, and periodically report his/her progress in the seminar. By the end of course, the student is expected to write a report that is in the format of a typical conference paper. The student can also select a topic within the broad area of semantic parsing, read and present related research in class, and write a synthesis paper.

Course format:
We will generally use the first half of every class to present papers, and the second half for project reports.

Readings:

1. Selected conference papers and journal articles on meaning representation, automatic semantic parsing, and deep learning

Grading:

- Participation: 40%
- Synthesis paper or project report: 60%

Academic Integrity:

You should finish homework assignments, exams, and project reports on your own unless a project is explicitly stated as a collaborative project. Late projects are subject to grade deduction.