New Course Proposal: Advanced Natural Language Processing Seminar

Noah Smith
nasmith@cs.cmu.edu

We propose a new LTI spring course, "Advanced Natural Language Processing Seminar" (proposed course number 11-713). This would be a seminar course with content varying from semester to semester based on what is happening in the field of NLP. Like other Advanced X Seminars in the LTI (MT, 11-734; IR, 11-743; Dialog Processing, 11-716; Statistical Learning, 11-745), this seminar would be a six-credit LTI elective targeted at senior Ph.D. students. To balance the fall/spring course offerings and permit a natural NLP sequence, we propose to offer this course in **spring** semesters, starting in **2009**. The course would meet once per week for 80 minutes.

Course goals: improve participants' knowledge of current techniques, challenges, directions, and developments in all areas of NLP (i.e., across applications, symbolic formalisms, and approaches to the use of data and knowledge); hone students' critical technical reading skills, oral presentation skills, and written communication skills; generate discussion among students across research groups to inspire new research.

Format: In a typical semester, a set of readings would be selected (with student input) primarily from the past 2-3 years' conference proceedings (ACL and regional variants, EMNLP, and COLING), journals (CL, JNLE), and relevant collections and advanced texts. Earlier papers may be assigned as background reading. The format of each meeting will include a forty-minute, informal, critical student presentation on the week's readings, with presentations rotating among participants, followed by general discussion. (Depending on enrollment, presentations will be given by 1-2 students.) Each presentation will receive a grade and instructor feedback. Apart from the presentation and classroom participation, each student will individually write a 3-4-page white paper outlining a research proposal for new work extending research discussed in class - this is similar to the Advanced IR Seminar. We may assign this in tandem with giving a presentation (as in the Advanced IR Seminar), or we may make it a separate exercise.

Example Topics:

- Semantic parsing and transforming sentences into logical representations. Readings: chapter 16 from Jurafsky and Martin (ed. 2, 2008), various papers by Mooney and students (2006-present), Zettlemoyer and Collins (2005, 2007), Pfeffer and Getoor (2005 IJCAI tutorial), Domingos papers on Markov logic (2005-present), Pless and Luger (2001), Baldridge and Kruijff (2002), chapters from Steedman (2000).
- Computational approaches to dependency syntax, including projective and nonprojective models, learning and parsing algorithms, and theoretical results. Readings: Eisner (1996, 1997), McDonald et al. (2005), Nivre (2006), Koo et al. (2007), McDonald and Satta (2007), Smith and Smith (2007), Kuhlmann and Möhl (2007).
- Finite-state morphology, following Beesley and Kartunnen (2003) and/or the first part of Roark and Sproat (2007).

- Nonparametric Bayesian approaches to statistical NLP. Readings: Teh et al. (2006), Goldwater and Griffiths (2007), Liang et al. (2007), Finkel et al. (2007), Haghighi and Klein (2007).
- Computational lexical semantics and distributional similarity. Readings: chapter 20 from Jurafsky and Martin (ed. 2, 2008), Pado and Lapata (2007), Curran and Moens (2002), Snow et al. (2006), Pereira et al. (1993) and LSA papers as background.

Prerequisites:
- *Required*: Algorithms for NLP (11-711) or equivalent at another university, or permission of the instructor
- *Required*: Grammars and Lexicons (11-721) or equivalent at another university, or permission of the instructor

Further notes:
- Like the Advanced IR Seminar (11-743), the Advanced NLP Seminar is separate from the area lab, but can be taken concurrently.
- Grades will be split evenly among the presentation(s), white paper, and class participation.