Language and Statistics II
Literature Review Instructions

Noah Smith
School of Computer Science
Carnegie Mellon University

Fall 2006

1 Motivation

Understanding and keeping up with the literature in a particular research area is crucial for long-term success in that area. As a major portion of your grade in this course, you will integrate a large body of existing work into a coherent literature review that will be *useful* to you or to others who want to contribute further in the area.

2 Ground Rules

This literature review is an independent exercise: everything you write should be your own. You are welcome to talk with anyone you like about the literature, including the authors of papers you read and other members of the research community who know the area well—just be sure to give credit in your review to anyone who was helpful. You are expected to read whatever literature is necessary to meet the requirements of the project; it is expected that this will include at least ten articles, and probably no more than thirty. It's often hard to know how big an area is before you start an undertaking like this one, especially if you don't know a lot about the area, so keep in mind that you might need to scale up or down as the semester progresses.

3 Questions to Answer

- First and foremost, what is the research problem? Can it be stated concisely and formally? Does everyone formulate it the same way? Are there substantive differences in the different ways of presenting the problem? How is performance evaluated? Is there controversy about how it should be evaluated? What are the available datasets, and how can they be obtained? This section should be sufficient for a CS graduate student to start to tackle the problem, even if she is ignorant of all the existing approaches to solving the problem and has very little background in language technologies.
- What are the existing solutions? You should focus on empirical solutions, but you should not neglect any technique that is competitive or dominant. How different are these solutions from each other *formally* and *algorithmically*? Researchers do not always make it easy to answer

this question; papers are often written to trumpet what's new without clarifying how much of the underlying solution is really the same. You will need to be critical and thoughtful. Your coverage should include all key work in this area. Imagine that you are trying to protect the reader from being accused of ignorance of important work.

- What are the major obstacles for future work? Is there some resource that would help bring on a major development? What hasn't been explored yet? Are there formal problems that need to be tackled? What ideas do you have for future work on the problem? Any thoughts on how to solve or avoid the current obstacles? Are there seemingly obvious ideas that haven't been tried?

You need not carry out *any* implementation for this literature review. You are welcome, however, to present useful statistics or examples from key datasets—whatever will help the reader to understand the problem. Be creative.

4 Deadlines

September 12 Write a one-page proposal describing your topic (what you know about it already) with an initial reading list of 5–10 articles. Feel free to set up a meeting with the instructor before writing your proposal. Submit in pdf by email to the instructor by 5pm.

October 16–20 Mandatory meeting to discuss your progress. Email instructor to set up a time to meet.

November 10 Draft due. Submit in pdf by email to the instructor by 5pm.

November 28, 30 Oral presentations and discussion. You'll give a 20-minute talk summarizing your review. This should be polished and "conference-quality."

December 8 Final version due. Submit in pdf by email to the instructor by 5pm.

5 Suggested Topics

These may be too large or too small; that's why we have monthly checkpoints to make sure the review ends up being reasonable. Feel free to propose something else. The review should focus on either a coherent *problem* (question answering) or a major approach to solving a major problem (syntax-based MT).

- Question answering
- Textual entailment and paraphrase
- Coreference resolution
- Summarization
- Sentiment classification
- Natural language generation

- Morphology induction and modeling
- Syntax-driven machine translation
- Syntax-based language modeling
- Data-oriented parsing (DOP) and machine translation
- Natural language interfaces to databases
- Finite-state parsing
- Optimality Theory

6 Final Note

It is sincerely hoped that the document you produce will be useful to you or to others in the future. If the area is closely related to your own research, you may want to recycle parts of your literature review in other reports or publications, or even your thesis. Even if the topic is not of special interest to you for future work, you are encouraged to put your report online for the use of others in the area. Consider it an act of scholarly community service.