Final Projects
CS 287: Statistical Natural Language Processing

1 Important Deadlines

The final project will consist of the following four aspects:

1. Project Proposal
2. Status Update
3. Final Paper
4. Oral presentation in class

No late days for any course project deadline.

These deliverables will be due at the following deadlines:

Aspect	Deadline
Project Proposal	3/25, 5pm
Status Update	~~4/15~~ 4/22, 5pm
Oral Presentations	~~4/26~~ 5/3, in class
Final Project	~~5/2~~ 5/12, 5pm

2 Goals and Scope

This course has been devoted to coverage of research topics in modern natural language processing. The final project is the capstone of the course and a chance for you to engage with a difficult research problem in the area. Students are expected to design and carry out final projects working in teams of two. You can use the Piazza forum to find partners. As you have already implemented several research systems, it is no longer a stretch to build one of your own :).

A crucial element of the project will be reading and responding intellectually to contemporary research in the field. Much of the class has been devoted to developing vocabulary for reading papers on these topics. With this in mind, we will evaluate the project heavily based on their processing of related work. Before jumping on a project be sure to first find related references. As a proxy we recommend looking at papers from:

Natural Language Processing

- ACL/NAACL/ EMNLP

Deep Learning/ Machine Learning

- ICML/NIPS/ICLR/AISTATS/UAI

Additionally, NLP has several shared tasks which contain data and evaluation on a specific problem:

- SemEval/CoNLL Shared Tasks/KDDCup

It has also become common to post new papers on the Arxiv. Note though that these papers are often not peer reviewed. Be sure to be aware of this and keep it in mind.

Plausible projects include: reimplementing and extending past papers, applying algorithms from the class to new domains, presenting comprehensive analysis and results to new language domains, and experimenting with new techniques on optimizing or building large-scale systems.

If you are having trouble finding a topic yourself, a list of possible course ideas is given on theprojects wiki. Their content and scope are meant to be suggestive, not definitive. The teaching staff would be delighted to talk with you about possibilities. Each suggestion has an associated TF; he would be a good person with whom to speak first, but feel free to contact any of us. After the proposal, each final group will be assigned a course mentor who will be the best point of contact throughout the project.

We will evaluate your project on the concepts it investigates and the results and on how well it demonstrates your comprehension of the concepts, techniques and issues we have covered in the class. The project grade will incorporate evaluation of the proposal, presentation (oral or poster) and final paper quality. As there is no final exam for CS 287, your final project is the major integrative element of coursework. It will account for 20% of your overall course grade.

3 Deliverables

3.1 Project Proposal

To ensure that you choose an appropriate project, you are required to turn in a 1–2 page project proposal. The proposal should begin with a clear, unambiguous statement of your topic, and include all of the following:

- the motivation of the project
- the research question being asked
- the experimental setup
- clear empirical metrics and baselines
- proposal for evaluating the success of the project
- a bibliography of three (3) current references for the project

The proposal will be graded on mainly on completeness. It will allow us to assign you a mentor and to propose early suggestions for related work and experiments.

3.2 Status Update

To ensure that you are on track with the project and to identify any issues on time, you are required to submit a short (3-page) report with a status update. The report should describe the problem you are working on, the progress you've made so far, and any problems you came across that you would like to get help with.

3.3 Presentation

The oral presentations will be on April 26th. These presentations are a chance to explain your problem and approach, showcase what you've accomplished, and get advice on surmounting any hurdles you've encountered. Students are expected to attend both presentation sessions, as they provide an opportunity for you to learn from each other.

Oral presentations will be allocated 10 minutes, and should be focused on key issues. You are encouraged to bring less than 4 slides, because often just one diagram or chart can explain the essence of your idea and save lots of presentation time. You need not prepare fancy graphics; just come prepared to explain your topic and share what you have discovered.

3.4 Report

Finally, you must submit a written report on your project and the complete, well-documented source code for it. The report should be a maximum 10 pages in length (this is actually quite short, so you should be judicious with your details and graphics).

The report format should be exactly the same as the homework assignments for the class and use the same Latex template. Your goal should not be to innovate in the format, but to use the standard we have had in class, and present details in a clear and, hopefully, obvious way.