Project Proposal for 10-605

William Cohen

March 5, 2013

1 Overview of the course project

The goal of the project is to give you additional experience using large data for some analytical task. You should work in a group of two or three students (if you want to work in a group of one or four, please clear this with William.)

The major milestones for project-related tasks are:

- By 12:00pm Wed 3/20: Hand in an *initial proposal*. The initial proposal should describe: the team of students working on this project, the data you plan to use, a description of the analysis that you plan to perform, and a description of how you plan to evaluate your work.
- By 12:00pm Tues 4/4: Hand in a *final proposal*, revised based on comments you will have received from the instructors.
- By 1:30pm Fri 5/3: Hand in the *final project writeup*, which will be in a conference-paper format paper 4-8 pages long.

We may also allocate 1-2 class periods to a poster session, or talks, so you can present your work in that format.

2 The initial proposal

Your initial proposal should be no more than two pages long, and should say:

- Who will be working on the project. Project teams should be two or three students; again, if there's a reason that's not possible, please talk to William.

- What dataset or datasets you plan to use. You should verify that the data is publically available (and actually has all the data that you expect it to have). I strongly suggest using a previously-published and readily-available dataset for the project. If you want work on a novel dataset, that's ok, but **you must have finished all data collection by 3/20**, when this initial proposal is due. A short description of the dataset is also helpful.

- What you plan to do, and how you plan to evaluate it. I strongly suggest picking a well-defined problem that has an existing baseline method for solving it, and then focusing the project on getting the baseline results, and then improving or analyzing the method baseline method.

- What prior research (that you know of) is most relevant.

- What computational resources you will use.

There are some project and dataset ideas posted on the class wiki.