Final Project

LING 539: Statistical Natural Language Processing

Gus Hahn-Powell

Last Revised April 23, 2020

Overview

Only graduate students enrolled for credit are required to undertake a final project. Students are encouraged to work *individually*; however, with special permission and sufficient topic scope, I may permit teams of up to 3 students. There are three options for your final project.

Option 1: Shared task with leaderboard

Construct a statistical natural language processing system for one of the SemEval 2019 shared tasks[1]. Alternatively, choose another task with a publicly accessible leaderboard (AI2 Leaderboard[2], CodaLab[3], Kaggle[4], etc.). Note that for this option, your system's results must be reported on a publicly accessible leaderboard.

Option 2: OpenClass Prize

See `https://openclass.ai/prize` for a detailed description. This is essentially a variant of Option 1.

[1] `https://alt.qcri.org/semeval2019/`

[2] `https://leaderboard.allenai.org/`

[3] `https://competitions.codalab.org/competitions/`. Ensure that the task you have selected has not ended (Under the "Phases" tab, the task should show "Competition Ends: Never").

[4] `https://www.kaggle.com/docs/competitions`

Option 3: Propose your own project

Students are encouraged to apply what they learn in this class to their own research interests.

Workflow

You will manage the various components of your final project using a Git repository (one per team). Rather than waiting until the cusp of each deadline, I recommend you commit and push your work regularly to mark your progress. This also makes it easier for me to provide you feedback as I have access to your repository.

Submissions

Do not email me any drafts. All work must be submitted by committing and pushing to the `master` branch of your repository's remote. At various points throughout the semester, I will review the version you've pushed to `master` to assess progress.

Grading and feedback

Feedback will be pushed to your project's Git repository under the `feedback` directory.

Grades will be posted to the course's D2L site:

- `https://d2l.arizona.edu/d2l/home/869634`

Pulling in upstream changes

As the semester progresses, I may share new resources related to your final project. In order to pull these changes, you'll need to add another remote to your project's repo:

```
# NOTE: you only need to add the new remote `upstream` once.
git add remote upstream "https://github.com/ua-ling-439-spring-2020/final-project-template.git"
```

To pull in any changes, run the following command:

```
git pull upstream master
```

Components

Your final project consists of core components summarized below. Deadlines for these components will be spread across the semester.

Proposal

Your proposal should include the following sections:

- Task description

  Provide a brief overview of the selected task. Include the URL to the task leaderboard. How will you be evaluated?

- Planned approach

  Describe how you intend to approach the problem. For example, what features and libraries do you intend to use? Do you intend to use any supplemental data? Where will you get this data?

- Delegation of responsibilities*

  *Only for teams with more than one member.

- Project timeline

  Provide a project schedule to gauge your progress (results for initial end-to-end system, experiments, data collection/augmentation, etc.)

- Anticipated challenges

  What might impede progress? How will you mitigate these risks?

Project code

- Provide a concise description of its organization
- Include clear instructions for running your code
  - Dependencies and runnables must be containerized using either Docker[5] or Singularity[6]

[5] `https://www.docker.com/why-docker`
[6] `https://sylabs.io/guides/latest/user-guide/`

Presentation slides

Presentation slides (PDF) summarizing the problem, approach, and results. These slides are used in video presentation.

Video presentation

A pre-recorded video presentation (3-15 minutes) summarizing the problem, approach, and results. Your video should be uploaded to a service such as YouTube, Vimeo, or Dropbox by the 9PM on the final day of class (**5/6/2020**). Include the URL for the video in your repository's README. While the video needn't be public, ensure that it can be accessed via the specified URL.

Timeline

What	When	Where
Proposal	2/23/2020 (9PM)	Final project Git repo (`proposal/final-project-proposal.pdf`)
Video presentation	5/6/2020 (9PM)	Uploaded to YouTube, Vimeo, Google Drive, DropBox, etc. Add URL to README of Final project Git repo
Slides (PDF)	5/6/2020 (9PM)	Final project Git repo (`presentation-slides/presentation-slides.pdf`)
Code and instructions	5/6/2020 (9PM)	Final project Git repo

Table 1: Tentative schedule for final project components.