Project Milestones

Setup

In this course, we are going to cover a variety of NLP techniques. As such, I would like you to develop an application of any technique(s) to a topic that interests you. You may work alone, or you may find a partner and share the project (as long as you clearly define how the work is divided). Since there is a mix of undergraduates as well as graduate students in this course, here is how I expect the workload to be shared if you form a partnership:

1. If both partners are undergraduate students, I expect the workload to be split 50/50.
2. If both partners are graduate students, I expect the workload to be split 50/50.
3. If one partner is an undergraduate student and the other is a graduate student, I expect the graduate student to perform 2/3 of the work.

I do not want you "double counting" your project with another project for another class (directed-study is an exception). It is perfectly fine (and even encouraged) for projects to *relate* to another project from another class, as long as the boundary is clearly defined.

Milestone 1: The Idea

For the first milestone, I would like you to form a partner if you choose to have one, and to pick a problem to work on. The deadline for this milestone will be on Wednesday 09/18/2024 at 11:59pm EST. It is perfectly ok if you don't know how to solve your problem yet: the deadline is intentionally set too early. We just want to know what problems you're interested in! Its ok to change your problem later on. You will submit a project proposal document that is between 1/2 - 1 page per undergraduate student, and 1-2 pages per graduate student. Graduate students should format the proposal like the introduction/motivation section of a conference paper (i.e. use the ACL Latex template which you can find here). Your project member(s) should be in the author list. Your proposal should address the following points:

1. What are you trying to do? Describe your goal **without** using any jargon.
2. Write (or draw) an example of what input/output or user interaction would look like.
3. If your project is successful, how will it be helpful to people?

Please submit a pdf on gradescope by the deadline. When grading, we may ask you to fix problems in your proposal and resubmit.

Miltstone 2: Data and Baseline

For the second milestone, you should have everything you need to start experimenting with solutions to whatever problem you've chosen. Describe to us what you've done so far. Page guidelines are the same: 1/2 - 1 page per undergraduate student, and 1-2 pages per graduate student. Graduate students should format their descriptions like the data/baseline sections of a conference paper. Be sure to address the followiong points:

1. What data will you use? Include URLs or LDC catalog numbers (don't worry if your data doesn't have one). Describe what preprocessing steps you used to prepare the data, or mention if these steps were done for you.
2. What metric(s) will you use to evaluate performance?
3. What baseline methods will you compare against? Baselines are something you should implement in about an hour or so. Hopefully you can get access to others' code if you plan on using it for a baseline.
4. How well does your baseline perform (using your evaluation metric(s))?

Please submit your description on gradescope by 10/23/2014 at 11:59pm EST. For some of your projects, these criteria will not apply. Please talk to me (Andrew) if your project belongs in this category.

Milestone 3: Poster Presentation

I will post a LaTeX file that I used to create a poster presentation when I was a masters student. I find that LaTeX is very useful for creating posters, but you do not need to agree with me, or even use LaTeX for this! Your posters should address the following points:

1. Goal: What are you trying to do (no jargon)?
2. Method: How are you trying to do it (briefly)?
3. Experiments: How well does it work (so far)?
4. Conclusions: What did you learn from your experiments (so far)?

The poster session will be on 12/09/2024 from 7-9pm (sorry. I will try to get this changed if I can). Singe-person teams will present during either the first or second hour; two person teams will present during both hours. If you want us to print your poster for you, we can! Instructions on the size, etc. of what to submit will be forthcoming (I am still working with the department on this). We will provide all of the materials (such as easels, poster board, etc.) for you to mount your posters. Please submit a pdf of your poster to gradescope by 12/10/2024 at 11:59pm EST.

Miletstone 4: Project Report (Paper)

Length guidelines: 2-4 pages for undergraduate students, and 4-8 pages per graduate student. Graduate students should format their report like a conference paper using ACL style which can be found here. You should address the following points:

1. Goal: What are you trying to do (no jargon)? Give an example of inputs/outputs or user interaction. How might it be helpful to people?
2. Method: How are you trying to solve this problem? What are existing approaches to the problem?
3. Experiments: What data did you use? What metric(s) did you use to evaluate your approaches? What baseline method do you compare against? How well do your methods perform compared to the baseline, and why?
4. Conclusions: What did you learn from your experiments?
5. Replicability: Submit (or include links to) all the code that you wrote and all the data that you used.

Please submit your report as a pdf on gradescope by the last day of class (12/10/2024) at 11:59pm EST.