CMSI 5370 Final Projects

Dr. Korpusik, mandy.korpusik@lmu.edu

Due: Monday, May 1 at 11:59pm PST

The final project lasts from 3/14 until the end of the semester. You will be presenting your final project to the rest of the class (**everyone must speak!**) on Tuesday, April 25. I know it seems like a long time from now, but you'll only have less than two months to complete a full NLP system and written report in the style of an academic research paper. **Don't leave it all to the end!** Remember: sleep solves bugs :)

Project Schedule

Deadline	Deliverable
Tues, 3/21	Proposal due
Tues, 4/11	Milestone due
Tues, 4/25	In-class presentations
Thurs, 4/27	Peer- and self-assessments
Mon, 5/1	**Final project due!**

Project Proposal — 10 points

You may work on teams of your choice (roughly 2-3 students). Please email or Slack me your preferences for teammates, if any, before class on Tuesday, 3/14. Your first assignment is to propose an idea for a final project. Here is your chance to apply what you have learned in this course to a real-world application of your choice! Here are a few possible types of projects:

- New dataset: Apply an existing model (e.g., BERT or GPT-2) to a new task (e.g., medical domain).
- Analyze models: Is the model (or data) biased or ethical? Compare models (e.g. is an RNN or Transformer better for a given task?), or study/visualize what the model is learning (i.e., interpretability).
- Participate in a competition (e.g., SemEval, Kaggle, etc.).
- Combine domains: Images + NLP (e.g., image captioning, visual question answering), speech + NLP, reinforcement learning (RL) + NLP (e.g., RL for playing text-based games, or RL for dialogue).
- New model: Design a novel NN layer, loss function, optimizer, etc.

Here are sample projects from other universities (this is just for ideas!):

- Princeton: https://www.dropbox.com/s/xeg8cizwkhcf5ed/all_presentations.pdf?dl=0
- MIT: http://courses.csail.mit.edu/6.864/fall_2016/past_projects.html
- Stanford: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports.html

No matter what project you pick, you should be able to describe two extremes. At the low end, you should describe the bare minimum functionality necessary for the project to demonstrate its intention. You will be expected to get at least this far. At the high end, you should come up with possible extensions to make the project more functional or interesting. A good project idea is one for which it is easy to reach the bare minimum, but has lots of room to grow in many different directions.

By the start of class on Tuesday, 3/21, you should submit a 1-page project proposal (one per team) that 1) describes your **specific** goal or hypothesis. For example, a good goal is "Reimplement XYZ paper and verify the results on ABC tasks,"whereas an idea that is too vague is "Use BERT to analyze financial news."2) In addition to the hypothesis, you should describe your proposed approach. What dataset will you use? What baseline will you compare to? Which model(s) will you implement? How will you evaluate the models? 3) You should also have a section that lays out your schedule for the next few weeks, and maps out where you expect to be each week from now until the project presentation and final report. You should break the time down into milestones, including the initial design, baseline, evaluation, and extensions. You should also assign tasks to each team member. 4) Finally, include a brief literature review of at least two related papers. **Be sure to proofread your proposal, since part of your grade will be for writing style.**

Milestone — 10 points

By class on Tuesday, 4/11, you should submit a 1-2 page informal progress report (one per team). Your Python 3 code should be up and running (i.e., you should have a trained baseline model and preliminary results in a table and/or figure), and should be pushed to your team's GitHub repository. There should be a history of clear commit messages, branches for each team member, and merged pull requests from these branches into the master branch. Start with your hypothesis and literature review from the proposal. You should then explain what data and code base you are using, as well as how you pre-processed the data and divided it into training, validation, and testing sets. Conclude with a brief discussion of the preliminary results.

In-class Presentations — 10 points

You will have 8 minutes to present your project to the class, followed by 1-2 minutes of Q&A. You should describe your goal or hypothesis, give a demo (if applicable), and explain the data you used, models you implemented, and results you have so far. You should avoid talking about low-level details of the code.

Make sure that everyone in the group presents something. By this point in time, you should all understand how your code works. If something was a mystery, you should have been asking your teammate to explain it to you! **The presentations should be fun and engaging, and you should rehearse.**

Some other ideas for things to talk about, if you have extra time:

- Your team dynamics: Who did what? How well did you communicate with each other? How did you resolve (or not resolve) differences in coding style, conceptual design ideas, or personal issues?
- What would you do next?
- Did you stay on schedule and accomplish the extensions, or was this the minimum viable product?
- What did you learn?
- What do you suggest for improving this class project next year?

To encourage practice, **teams who have consulted with peers or with the Writing Center will receive a 5% bonus on their presentation grade.** As evidence, please have the Writing Center submit a report, or please have your peer team submit their notes (e.g. what went well, what could be improved).

Self-Assessments — 5 points

Each member of the team should submit on Brightspace a short paragraph assessing yourself and your teammates. I will give you time to do this on the last day of class, in addition to course evaluations. You should indicate at least one code file that you had primary responsibility for, and which you feel best represents your coding, documentation and creative abilities. **Be sure to tell me what grade you would assign to yourself and each of your teammates.**

Final Report and Code — 40 points

The hard deadline for projects is Monday, 5/1 at 11:59pm PST—there will be NO extensions. Submit the following to your team Github repository:

- Your **well documented** code. You should have comments at the top of every file, as well as describing each class and function.
- A complete user's manual in the **Readme** for how to run your code. The user's manual should be designed for someone who does not know or need to know how your project works. Instead, it should contain enough information to enable others to run your code.
- Acknowledgements for any code or libraries used in your project that wasn't explicitly written by a member of your team.

Submit to Brightspace a pdf of your final final report (one per team) in the style of a conference submission. It should begin with an abstract and introduction, clearly describe the specific goal or hypothesis, present technical details, give results, compare to baselines, provide analysis and discussion of the results, and cite related work. I recommend using the Association for Computational Linguistics conference's provided style sheets: https://acl2020.org/calls/papers/#paper-submission-and-templates. Your paper does not need to be a certain length, but be sure to cover the requirements in sufficient detail.

The Writing Center

The Academic Resource Center (ARC) provides free, peer-to-peer writing support for undergraduate and graduate students working on projects in any discipline and at any stage in the writing process. We believe that all students can benefit from setting time aside to talk with another writer about their writing. The ARC's tutoring schedules are available to students with WCONLINE. You can schedule your own appointments from anywhere and at any time! To log in, view the schedule, and make or modify appointments, sign into myLMU and click "Writing & Course Tutoring" (under "Academics").
academics.lmu.edu/arc | 310-338-2847 | tutoring@lmu.edu | @lmuarc

YOU WILL LIKELY FIND YOUR WRITING CENTER VISIT MORE VALUABLE IF YOU GO EARLIER THAN THE NIGHT BEFORE SUBMISSION.