CS4248 Natural Language Processing: Project Guidelines v1.1 (210226). Refinements to these guidelines are expected. Please ensure you have the latest copy. Accessible at http://www.comp.nus.edu.sg/~cs4248/2020/project-guidelines.html .

CS4248 Project Guidelines

1 Overview

v210226 - clarifies presentation and final report notional rubrics and percentage grading, eliminates video deliverable, and adds Teammates intra-team evaluation procedure. Highlights in yellow versus original guidelines (v210218).

This reference document formally describes the requirements for the project component of CS4248. There are two kinds of projects that can be done: Self-Defined Projects and Staff-Defined Projects.

1.1 Self-Defined Projects

The project is intended to be a limited investigation in the area of natural language processing of your choice. The purpose of the project is to enable you to study an area of your interest in greater detail in a practical way. The project can take on many forms, including but not limited to:

a. Projects that explore the application of natural language processing ideas to an interesting "real-world" problem.
b. Projects that involve a theoretical or empirical study of aspects of a learning method or model.
c. Projects that do an experimental, comparative study of various task-specific NLP methods.
d. Projects that extend previous work (e.g., UROP or FYP) of a project member[1]. *Caution: this type of project may lead to unequal contributions, due to members' prior expertise.*

Doing such a project gives you more flexibility and allows you to work on something of your liking. However, at the same time, this may potentially require some additional effort (depending on your problem) such as data collection or coming up with suitable baselines or such.

1.2 Staff-Defined Projects

The staff may also propose projects suitable for students to try. These are projects that have a dataset and a clear-cut problem that you can attempt. The staff-defined projects are available in LumiNUS and available via the canonical link http://www.comp.nus.edu.sg/~cs4248/2020/project-staff.html

1.3 Common Guidelines and Timeline

You will need to propose a topic for your project in a formal project proposal due in Week 07. The staff and your peers will then review your work by the end of Week 08. After the project proposal, you will be assigned a contact TA that you can use as a resource for questions and advice.

In Weeks 10-11, the staff will meet with all teams for the mandatory consultation. Your team will need to prepare a short presentation deck to present to the staff about the progress of your project.

[1] If you are working with other faculty or students on the project, please mention that clearly in your project proposal. Note that this type of project sometimes results in uneven member contribution, due to the prior expertise of some members. The team will need to pay greater attention to balancing effort with this project type.

You will need to prepare a short presentation video (limited to a max of 180 seconds, geared towards a lay audience about your project) and an A1-sized poster with more technical details. The poster and video will also be peer-graded in a single-blind fashion (see below). Finally, your team will prepare a final project report, which the staff will grade and provide summative feedback (no peer review).

Peer Reviewing: We will be performing peer-reviewing for the project phase of the course. There are several benefits to peer-reviewing. Most importantly, it helps you understand and appreciate work from other students and groups, and it provides more feedback to everybody about the projects.

Peer-reviewing means that each one of you will be given a few submissions of your classmates to read and grade. This essentially involves providing some brief comments to help each other out. Please be as fair and impartial as possible during the reviewing process. We will be having TAs evaluate and provide feedback as well. You will be graded on how well you review other projects and how insightful your comments are. While it may be possible to "game" the peer review to gain a higher grade, the TAs' evaluation will assist us in preventing this.

There will be two phases for the peer-reviewing corresponding to the project proposal, the poster presentation, and the project video and poster. The two phases use different peer-review models, called "double-blind" and "single-blind" reviewing models:

Double-Blind Reviewing: This means that the reviewing team does not know the identity of the project they are reviewing, nor do the authors of the project know who is reviewing them. Even more, a reviewer is not allowed to disclose who he or she is reviewing. As an authoring team, DO NOT put your names or Student IDs on your submission or into your submission filenames.

Single-Blind Reviewing: This means that the reviewer team knows the identities of the authors of the project they are reviewing, but the authors of the project do not know who is reviewing them. Again, the reviewing team is not allowed to disclose who they are reviewing. Single-blind will be applied to the poster and video.

In all models, the course staff will know the identity of everybody.

Further information will be made available before the first peer-reviewing deadline. If you have further questions, we will be happy to address them on the forum. All project submissions are to be submitted via Google Forms which will be released later.

The important dates are:

- Week 07 (4 Mar 23:59): Project Proposals due
- Week 08 (11 Mar 23:59): Project Proposal peer reviews due
- Week 10 or 11: Meet with staff for project consultation.
- Week 13 (15 Apr 23:59): Posters and videos Due
- Reading Week (22 Apr 23:59): Peer grading of the Project Posters and videos Due
- Reading Week (23 Apr 23:59): Project reports due.

For all project deliverables (proposal, project posters, project reports), please follow the organization of the provided template. For Posters, please use your creativity to make your team's submission stand out.

2 Project Proposal (5% of the grade; due Week 07)

The project proposal should outline what you want to do in your project. We will give you feedback on your proposal to make sure the project you are proposing is feasible and appropriate. The proposal should contain the following sections:

Motivation. Explain why this project is interesting and important.

Statement of the Problem/Task. A statement of the problem, issue, or task that you're interested in studying. In particular, try to formulate the key questions (2 to 4 questions is probably a good number) that you will answer in the project.

General Approach. A high-level description of the general approach you'll use to address the questions. This should include how you will evaluate and what evidence you are planning to gather (e.g. how you can answer the questions through experiments on data).

Evaluation. Explain how you will evaluate your project. Propose what your team thinks is a satisfactory project outcome (C grade) and an excellent project outcome (A grade).

Resources. A list of resources you have/need to conduct the project. This includes reading, software, datasets, etc. How are you planning to get these resources?

Schedule. A schedule of work indicating the dates by which you plan to complete components of the project. Make sure the schedule is plausible.

Format: Please use either the proposal.docx file template available in LumiNUS; or preferably, use the the Google Drive document source available at http://www.comp.nus.edu.sg/~cs4248/2020/project-proposal.html to prepare your submission. The project proposal is limited to 1 page. Only one submission per team should be submitted through http://www.comp.nus.edu.sg/~cs4248/2020/proposal-submission.html.

The proposal will be **double-blinded peer-reviewed** and should not contain any team members' names in the document and filename. Name your proposal file with the staff-assigned team ID, e.g. **XX.pdf**. Evaluation guidelines are available below, and are used by your TAs as well as your peers. Please keep the above guidelines in mind when performing the peer-reviewing.

Grading Rubric

Good Title. Does the title:
a) Summarize the main idea of the project in an exciting way?
b) Express the intent of the project concisely?

Statement of the Problem/Task
a) Does the proposal outline a problem statement, issue, or task that the team is interested in studying?
b) Does it formulate a few (2 to 4) questions that the team proposes to address?

General Approach.
a) Does the proposal contain a high-level draft description of the general approach proposed to address the questions?
b) Does it include preliminary plans for evaluation, data gathering? I.e., how the team plans to answer the questions through experiments on data.

Resources. a) Does the proposal give a short list of resources the team plans to use to execute the project (inclusive of readings, software, datasets, etc)?
b) Did the team describe any strategy for getting the resources?

Schedule / Project Feasibility. Does the proposal include:
a) A schedule indicating dates by which the team plans to complete the project components?
b) An assignment of the team members to the deliverables (inclusive of peer-reviewing duties)?
c) Is the schedule feasible given the timeline, expertise, and load of the team members?
d) For projects with possibly too large a data source (e.g., Kaggle projects), does the team propose a way to scope the data or problem accordingly to make it feasible?

Comments. Please add any constructive comments to your peer review. Projects aiming for an excellent outcome (A- or better) need to give constructive comments to their peers.

Integrity / Originality. If the project is self-proposed:
a) Is the project original?
b) Do the project team members give credit where prior work or documents help the team materially with their proposal?
c) Are there any doubts that their proposal is re-used/plagiarised from other sources?

3 Mandatory Project Consultation (occurs in Week 10 or 11, by appointment)

We expect most student teams to have done some introductory reading and work towards their project goals. The interim project consultation is a mandatory, 15-minute time slot in which you'll have to meet with an assigned TA as well as the lecturer to discuss your project.

In preparation, you must provide a set of presentation slides to describe your project and feedback that you wish to get from the course staff. Please practice such that you can conclude your presentation within 10 minutes so as to make use of staff feedback in the final 5 minutes of your consultation slot.

The slides should consist of a (1) title slide with your staff-assigned team ID; (2) motivation and/or statement of the task; (3) key questions to answer; (4) schedule. Teams are advised to make this session useful to you in gathering input or feedback on your project, as well as helping the team assess the quality of your project.

You are encouraged and welcomed to contact your contact TA to get more feedback on your project, and welcome to use the Project forum on Slack to solicit additional feedback. We will try our best to match a suitable TA to your project topic.

For more detail, you may want to look at the indicative interim project consultation slide presentation template http://www.comp.nus.edu.sg/~cs4248/2020/interim-project-consultation.html. You need not follow the template.

4 Project Poster and Presentation (8 + 8 = 16% of the grade; due in Week 13)

As of 26 Feb, we have removed the video deliverables from the project, and added the grading rubric for the presentation. All teams will need to prepare both an A1-size poster and present it to the staff.

Your poster should contain the following:

- Provide motivation for your project, explaining why it is important and interesting,
- Explain your research questions,
- Provide (preliminary) evidence,
- Draw (preliminary) conclusions.

The posters will be **single-blinded peer-reviewed**; i.e., you can include your names on the posters. Please keep these guidelines in mind when performing the peer-reviewing.

Posters are meant for technical presentation and defense of the quality and rigor of the project. The format is an A1-sized single poster, preferably in portrait format. You can be creative with your poster, but it must contain scientific content that helps to defend the decisions made in the project and discuss insights and results related to the theory of the course.

Your team's **presentation** of the poster serves as a primary means of evaluating your project by the staff. You should try to make your posters as self-contained as possible. You should be able to summarize your project (with the help of the poster) within 5 minutes, with sufficient time to answer questions about your project within a 10-minute slot. You do not need to print out the posters.

N.B. all student teams' course outputs (poster, final report) are considered as open-source and will be archived on the course website. If you would like your project held in confidence, please approach our staff; we will be happy to comply, within reason.

For more detail, please access the template for the poster: http://www.comp.nus.edu.sg/~cs4248/2020/poster-template.html.

Poster Grading Rubric

Content (70%)
(As this is the key evaluation criterion, this question counts for 5 times the weight of the other questions for the Poster)
a) Does the poster contain a good summary of what you think the team's report should reflect?
b) Do you think it is easy to glean the key findings of the team's work within 5 minutes?
c) Does it address the motivation/significance of the project, its originality (if a self-defined project), the relevance of the team's work to the project's theme, the quality and persuasiveness of its arguments, and linkage with the concepts taught in the course?
d) Does it address both macroscopic, dataset-wide level performance (e.g., RMSE, Accuracy, F1 measures) as well as microscopic, individual instance level performance (careful error analysis with diagnosis)?
e) Are references appropriately shown on the poster?

Organization (15%)
a) Is the poster well organized with respect to accessing the content?
b) Does it have a good reading flow?
c) Does it utilize space appropriately for the amount of content that is discussed?
d) Are the key mandatory elements placed appropriately (names, title, abstract, affiliation)?

Formatting (15%)
a) Are fonts, color, whitespace, and visual elements easy to read?
b) Are the important highlights easy to find?
c) Are the content items proofread and clear?
d) Images cleared for use and attributed where necessary?

Presentation Grading Rubric

Delivery (30%)
a) Does the team do a good job of having all members involved in the project, in the presentation?
b) Is the delivery of the content clear and generally smooth?
c) Does the team finish their presentation within the stipulated time limit for the presentation?

Content (60%)
a) Does the presentation cover the key aspects of the project up to the presentation week? These factors include the projects originality, significance, relevance, related work, methods, experimentation (macro, micro) and ablation testing,
b) Does the presentation employ both the visual and audio modalities of the presentation medium (Zoom / Wonder session) effectively?

Question Answering (30%)
a) Does the team understand the questions posed by the staff?
b) Does the team answer the questions appropriately? Do they follow up on questions where they do not have sufficient information during the presentation?
c) Do the different team members help reinforce the question answering, and demonstrate that expertise is shared across members of the group?

Extra Perks:
a) Does the team have a live or canned demonstration?
b) Did the team reduce their project work to a usable form for the greater public?
c) Did the team publish useful documentation for the general public to learn about their project or their learning journey.

6 Final Project Report (14% of the grade; due in Reading Week)

The .zip file should include your writeup and the source code of any programs you wrote for your project (don't just include a link to your repository). Include other files if you feel they are appropriate, but obviously, explain their relevance in a README. The submission must be made to **Luminus Files**.

For additional guidance in structuring the report, take a look at the template structure below. Not every project fits into this structure, and you might choose a different structure instead. The most important goals to keep in mind are:

- To motivate your project,
- To make a convincing argument that supports your conclusions,
- To make sure that the reader understands what your project is about and how you came to your conclusions, and
- To make sure that credit is given to all software, literature, *etc*. that helped you in your work.
- Format: Please use the double-column format provided. The progress report should not be more than six pages in length (excluding references). Being concise is a good thing, but do not sacrifice clarity and completeness.

For more detail, please access the template for the final report: http://www.comp.nus.edu.sg/~cs4248/2020/project-final-report.html.

Final Report Grading

The projects will be graded in the same spirit as research papers are assessed (though we do not expect you to do original work at the

same level). Here is a list of things that we will be looking for in your final Project Report (notional percentages added 26 Feb):

- Originality (10%)
- Significance (15%; are the questions you are asking interesting)
- Relevance to course (5%)
- Quality of arguments (35%; are the claims supported, how convincing are the arguments you bring forward)
    - Models (15%)
    - Analysis (20%; e.g., ablation studies)
- Connection to earlier work (10%; scientific literature or lecture materials)
- Clarity of writing (20%; how clearly are goals and achievements presented; organization; supplemental materials)
- Scope/Size (5%; in proportion to the size of group, member expertise mix and evenness of contributions)

Other bonus factors (not factored as deliverables; but may influence your teams grade positively by a maximum of 5%) include creation of supplemental materials: Github code, documentation, demonstration, webpage, applications, blog post and/or videos.

7 Overall Project Grading

The projects will be graded in the same spirit as research papers are assessed (though we do not expect you to do original work at the same level). Your project in total accounts for 5 + 8 + 8 + 14 = 35% of your CS4248 final grade.

Your peer-reviewing performance may be accounted for in the 5% bonus participation grade, shared equally among all members of the team. At the conclusion of the project, all team members will be called to use the TEAMMATES - Online Peer Feedback/Evaluation System for Student Team Projects, developed in SoC. For teams where individuals find themselves not contributing equally / unequally-but-agreed-upon, your inputs to Teammates and/or an interview with staff will be employed to arbitrate credit assignment. Note that CS4248 staff decisions with respect to credit assignment will be final.

Feel free to come and talk to us about the various aspects of your project (in fact we strongly encourage you to) so that we can make sure that you are on the right track. Finally, have fun while doing it; it's meant to be something that you are interested in doing!

Good luck!

~ ~ The CS4248 Staff ~ ~

Credits: Project guidelines based on Cornell's Thorsten Joachims and NUS's Bryan Low and the CS3244 project guidelines.