CS 4650: Natural Language Understanding
Project Guidelines

Instructor: Kartik Goyal
Teaching Assistants: Jakob Bjorner, Neelabh Sinha, Hrudhai Rajasekhar

April 15, 2024

1 Description

The course project is an attempt to enable students to work on a practical Natural Language Processing problem of their interest. It will allow students to define an NLP problem of their choice and solve it using their understanding of the topic and learning outcomes from the class.

2 General Instructions

- It is to be done in a team of 3-4. You are permitted (but not recommended) to work on the project individually. Individual project submissions will be graded on the same benchmark as group submissions. Please reach out to the instruction team and seek a prior approval if you plan to do so.
- The project has to be in the domain of Natural Language Processing. A project not deemed valid will not be evaluated and 0 credits will be provided for all components.
- It carries 25% of your total grade for this course (in total).
- The problem should have sufficient work to justify the recommended team size of 3-4.

3 Evaluation Components

The evaluation is divided into three parts:

1. Proposal (5 pts)
2. Final Report (12.5 pts)
3. In-class presentations (7.5 pts)

3.1 Proposal (5 pts)

The project proposal should describe your project - the problem, the motivation behind it, intended methodology, expected results and a feasibility analysis of the project.

Requirements for the proposal -
There will be Gradescope assignment where you will need to submit the following -

- Full names and GT Email ID (xxx@gatech.edu) of all team members
- Title of your Project
- A 2 page PDF (excluding references) in ACL template following the below mentioned requirements. A standard template will allow us to evaluate all groups fairly without accounting for difference in presentation format.

- The template has detailed instructions on how to use it. Before submitting the document, make sure to remove the review option to generate the final version, and add all your team members as authors.
- Only one member of the team needs to submit the Gradescope assignment. But, all the team members need to be added to the Gradescope submission to consider them for the evaluation. You can follow this guide to do the same.

Deadline: February 07, 2024, 11:59 PM EST

Components to be included in the proposal -

1. Introduction/Background - Provide an introduction to the topic and a brief literature review to establish the background.
2. Motivation - Describe the importance and exciting aspects of the problem statement of your project.
3. Problem definition and originality - Clearly define your problem statement that you are addressing, and discuss how is it different from the existing approaches.
4. Proposed Methodology - Provide an overview of the methods you are trying to use and the aspects of the project that, assuming ideal outcome, would lead to new knowledge.
5. Potential Results and Comparisons - Discuss the ways in which you will evaluate your implementation. You can mention metrics, qualitative outcomes, baselines you will be comparing your method against, etc.
6. Feasibility Analysis - Describe the most difficult aspects of the project and discuss the project's feasibility given the quarter long timeline.

Please note the following -

- Ground your proposal in existing literature. Include at least 5 references cited appropriately in your project proposal.
- You can use illustrations, tables, etc., to better convey your ideas. Appropriately refer the figures and tables that are added in the text. Please cite the source of any external illustrations, tables, etc., in your document.
- It is necessary to use the recommended ACL template. Violation of this will result in zero credits. No hand-written content in the document will be accepted. Changing any format setting of the template (like font size, column width, etc.) is also considered a violation.
- At any point, feel free to come to office hours to discuss the validity, scope, feasibility or anything else related to your project ideas.

Your project proposal will be evaluated on the following criteria -

1. Gradescope Form (1 point) - Filling the Project Title (0.5 pts) and details of your team members correctly (0.5 pts)
2. Motivation (1 point) - How clear is the motivation? Why is it important to work on the identified problem?
3. Originality/Innovation and Definition (1 point) - Is the problem clearly defined? How original and exciting is it?
4. Planning (1 point) - How well are the proposed methodologies planned and described? Are the potential results and comparisons that are defined relevant and holistic? Has the team planned well on how to approach the project through the course of the semester? Is the amount of work planned sufficient for the recommended group size of 3-4?
5. Feasibility Analysis (1 point) - How detailed is the feasibility analysis? Are most potential difficulties identified and discussed effectively? Is there some plan outlined to deal with those?

Details related to final report and in-class presentations will be shared in due course of time.

3.2 Final Report (12.5 pts)

The final report should describe the problem you are solving, what data is being used, the proposed technique you are applying in addition to what baseline is used to compare against.

Deadline: April 29, 2024, 11:59 PM EDT

Please follow the below **instructions for your final report** -

- It has to be 4-6 pages (excluding references) in ACL template. Please ensure you don't submit the review copy, but the final version of the template in your project. Add your project members as authors.
- The project will be submitted through Gradescope. Links will be added closer to the deadline. The form will be similar to project proposal form structure.
- You need to submit a PDF of your report and your code base that you built and used to complete your project. You can share a public link to your Git repository or a public storage (like GDrive), or zip your code and attach it.
- Please note, add all your group members to the Gradescope submission, along with adding their names in the submission form link. Last time, a few groups didn't do this and we added it manually. This time, due to the grading deadline, we will not be able to do it. Missing this might result in your group members not getting marks for project.
- There will be no late deadline for the project report.
- At the end of your report, include a Contribution table which clearly indicates work of each individual towards every aspect of the project. We will be using this for grading.
- Utilize illustrations and tables to effectively communicate or present your ideas/results. They are mostly better than verbose text.

Your report may follow the following **structure**. However, they are not strict as each project is different and may require different structure. The sections are as follows (all may not apply, but '*' marked are necessary to be there in some form (you can change the nomenclature, order, structure)) -

- *Introduction** - background of the project, motivation, summary of the proposed approach, contributions of your work (clearly highlight this)
- *Related Work** - a brief literature review of the work related to your project
- *Proposed Methodology/Body of the Project/Detailed Description* (divide into subsections as deemed fit)* - complete details describing your project on what you are doing, why, etc.
- *Dataset* - describe your dataset here, along with its properties, patterns you observed, distribution of the dataset, etc.
- *Experiments/Results/Findings** - describe the experiments you did to validate your proposed methodology and lay down your results. Present insightful information and use tables, visualizations, etc., to convey your point.
- *Comparison with other SOTA/baseline/etc.** - Ground your work in existing literature and compare your results with others. Discuss different aspects of the comparison, and what makes your results better/worse.
- *Ablation Study* - Systematically remove components of your proposed design to analyze their effective contribution by showing change in results after doing them.
- *Conclusion** - Conclude your findings.
- *Limitations* - Discuss limitations. Please note, any limitations discussed here will not be judged negatively while evaluating you. It is good to identify the limitations of your work on which rest of the scientific community can build on.

You can also add additional sections based on the nature of your project. The project will ultimately be evaluated based on how well you structure and prepare the report for your case, and not on following/not following the above structure. These are just an indicative.

The **grading rubric for the final project report** is as follows -

- *Clarity (2.5 points)* - For the reasonably well-prepared reader, is it clear what was done and why? Is the report well-written and well structured?
- *Originality / Innovation (2.5 points)* - How original is the approach? Does this project break new ground in topic, methodology, or content? How exciting and innovative is the work that it describes?
- *Soundness / Correctness (2.5 points)* - First, is the technical approach sound and well-chosen? Second, can one trust the claims of the report – are they supported by proper experiments, proofs, or other argumentation?
- *Meaningful Comparison (2.5 points)* - Do the authors make clear where the problems and methods sit with respect to existing literature? Are any experimental results meaningfully compared with the best prior approaches?
- *Substance (2.5 points)* - Does this project have enough substance, or would it benefit from more ideas or results? Note that this question mainly concerns the amount of work; its quality is evaluated in other categories.

Please note, there is no specific marks for data/model/preprocessing, etc., components as different projects are differently designed. You should be in contact with your assigned project mentor TA to keep the expectations mutually aligned.

You should also ensure to adhere to maintaining **academic integrity**. These include (but are not limited to) -

- Not using any figures, tables, content, screenshots, etc. from any other work that is not created by you when creating your report.
- Not using any external work that you are not allowed to use.
- Not citing any external resource that you have referred to in your work.
- Clearly acknowledge any external collaboration in completing your work.
- Disclose any ethics concerns or potential conflicts of interests.
- Not following prescribed format, guidelines of submitting your work.

Violations of academic integrity will be penalized considerably in your report.

3.3 In-class Presentations (7.5 pts)

Presentations will be in-class on the final exam day, which is as follows -

Presentation Day: April 26, 2024, 2:40 PM - 5:30 PM EDT

You basically need to present your project. Include everything that needs to be discussed as per the requirements of project report. However, to meet the time constraints, you can condense it and omit some information that you deem not useful. Please acknowledge that this is the only chance to showcase your work to your classmates. So, there should be an overall continuity and flow in your presentation so that they can understand and learn from your project.

Please **follow the following guidelines strictly** -

- Each presentation should be of 4 minutes (Presentation) + 2 minutes (Q&A). Crossing either of this individually will result in losing 0.5 points per minute. Time calculation will be done as per $ceiling(time\_used)$, i.e., if you use 4 minutes and 5 seconds, it will be counted as 5.
- You need to prepare Powerpoint slides for your presentation. The powerpoint slides will have to be submitted on Gradescope that will be due at 2 PM on the presentation day with no late deadline. You will also need to copy your slides to a PPT deck that will be shared. This PPT deck will be used during the presentation so that there is minimal time lost to load slides, get ready, etc.
- The presentation will be in the order of group number. So, add your slides to the presentation deck accordingly and be ready when your turn comes. Your timer will start 30 seconds after the previous group ends, or whenever you are ready (whichever is earlier).
- You should follow all academic integrity guidelines for this as well.

It is expected that you complete majority of your project by the presentation day, and you invest the remaining days to write your report. So, a presentation should ideally contain your whole project. You are free to take inspiration from the takeaways from your presentation and consider it in your report.

Presentation **grading will depend on the following aspects** -

- *Content* - Content of your presentation (please note, you can't cover everything in presentation, the idea is to identify what are the important aspects you need to cover)
- *Delivery* - Clear and accurate delivery of the content by the project team
- *Effective Use of Presentation Tools* - Effective use of visual and other presentation aids to convey your ideas
- *Q&A* - Concise and accurate answering to the questions asked
- *Time Management* - Adhering to the time constraints provided

4 Project Ideas

Your project can take many shapes and forms. The following are some directions you can take in your project -

- **Dataset Projects**
  - Objective: Define a novel task and develop a comprehensive and enhanced dataset for that task, possibly laying down annotation guidelines along with human and automated annotations.
  - Outcome: A new, high-quality dataset tailored for a specific application or language that is usable in the research community.
- **Analysis Projects**
  - Objective: Some examples of analysis projects would involve conducting thorough analyses on existing datasets/specific models, comparison of multiple models, exploring the learning dynamics, or performing theoretical and empirical comparisons between different algorithms.
  - Outcome: Insightful findings that inform the AI community about dataset quality, model performances, learning behaviors, or an insight into different algorithms.
- **Evaluation Projects**
  - Objective: Identify the limitations of existing evaluation methods for a task of your interest, and develop more reliable and comprehensive methods for evaluating AI models for different applications.
  - Outcome: Advanced evaluation methodologies that offer a more accurate and exhaustive assessment of capabilities and limitations of model behaviors.
- **Modeling Projects**
  - Objective: Develop algorithms and techniques for various NLP tasks for purposes like explainability, performance improvement, efficiency improvement, etc.
  - Outcome: Advanced models that push the boundaries of current capabilities, are more efficient, or address specific challenges in the AI domain.

A project can be at the intersection of one or more of such categories. We are also open to other possibilities.