COMP8420 2024 S1 Major Project (Assignment 3)

Qiongkai Xu (qiongkai.xu@mq.edu.au)

Due: Tuesday June 11th 2024 (during the exam period)

This assignment assesses the following unit learning outcomes:

- UL01: Identify the key natural language processing applications that match current and emerging industry needs. implement natural language processing applications.
- UL03: Implement natural language processing applications using common tools and libraries used in industry.
- UL 04: Design natural language processing applications using advanced deep learning techniques.
- ULO5: Apply natural language processing methods and techniques to industry applications using real data.
- ULO6: Apply good practice in the development, monitoring, and deployment of natural language processing systems.

It is expected that students will need approximately 30 to 50 hours to complete the project as a group (of at most 2 students).

All members in the group should have significant contributions to the project, presentation and report.

It accounts for 40% of the overall assessment for this unit.

Start this assignment as soon as you can! Find a reliable partner for your project. Attend most of our lecturers to find ideas on meaningful real-world tasks. Choose a suitable project wisely. Make (even tiny) pieces of progress every week. If you leave this project until the last minute, you won't have enough time to complete it, even if you can program at 1000 lines per hour and you have a server with 100 GPUs.

Essential Timeline:

- Presentation: Week 13 Practice workshop. (May 31st 2024).
- Code & Report: Exam period. (June 11th 2024).
- Q&A Sessions: Week 10 and Week 11 Practice Workshops. You need to form your teams and pinpoint your task selection by the Practice Workshop in Week 10. If you cannot attend the workshop that week, please send an email to inform TA & Qiongkai about team members and project title by Week 10.

1 Introduction

In this Major Project (a.k.a. Assignment 3), you are invited to assume the role of an analyst, consultant, or engineer tasked with the development and implementation of a Natural Language Processing (NLP) project for an IT company specializing in NLP solutions.

Your primary objective is to pinpoint a tangible challenge or need faced by an IT firm in the realm of NLP services. This entails deliberating on (1) the specific types of NLP services the company aims to offer, and (2) the potential hurdles they may encounter when incorporating NLP technologies into their operations.

Following this analysis, your mission is to leverage the technologies and methodologies covered in this course, possibly in conjunction with additional tools, to craft a viable solution to the identified issue. Your deliverables will include a comprehensive project that addresses the challenge, accompanied by a detailed report elucidating the problem-solving process.

2 Project Submission and Presentation Guidelines

We anticipate that each group will fulfil two primary obligations:

1. **Deliver a Presentation**: A brief presentation, spanning 3-4 minutes, is expected to be delivered during the Practice Workshop in Week 13 (Friday May 31st 2024). Should scheduling conflicts arise, proposals to present in Week 12 are allowed. It is imperative that all group members actively participate in this presentation.
2. **Submit Comprehensive Project Documentation**: This includes various components to be submitted during the examination period (Tuesday June 11th 2024): (a) a **report** on the project; (b) complete **codebase** of the project; (c) the utilized **resources**.

Detailed Submission Requirements:

1. Presentation:
    - Each group must present for 3-4 minutes at the Week 13 Practice Workshop. Alternatively, presentations in Week 12 may be requested.
    - At least one participation from the group is required.

2. Project report (4-6 A4 pages), which should be submitted to iLearn.
    - Project title and team member details (name, student ID, contact emails, etc.).
    - Introduction of the project scope: the real-world challenge your project addresses.
    - Methodology, highlighting the technologies implemented, datasets utilized, and evaluation methods for performance.
    - Experimental results: demonstrate the project's effectiveness, possibly in comparison to alternative solutions.
    - Contribution breakdown: explain each member's role and workload distribution.
    - Links of your project (Public Github project is suggested) and your model. Make sure your markers can easily view and download your project and models.
3. Source code and datasets, which should be publicly available to the markers (e.g., maintaining a public Github repo and submitting an URL for docker images).
    - Source code: Structure your project as a series of small programs or Jupyter notebooks. Include a README file to guide users through your project.
    - Datasets: Document the source of your dataset (for reused datasets) or the collection process (for original datasets). Submit a portion of the dataset (less than 2MB) for verification purposes.
    - Models: If your models are too large for direct submission, provide a download link within our submission.

Note that:

- You should **NOT** trivially recycle the assignments or any related materials completed in past or concurrent ongoing courses.
- Any actions in violating MQ's policy academic integrity policies are strictly prohibited.

3 Rubric

Presentation (10 marks)

- Scope of the project (5 marks) (0-1: not an NLP problem; 2-3: relevant NLP problems in real-world application; 4-5: the project should be useful in real-world business or research)
- Clarity of presentation (5 marks) (0-2: the project is not explained well; 3-4: good explanation; 5: clear explanation)

Report (10 marks)

- Clarity of report (5 marks) (0-2: the project is not explained well; 3-4: good explanation; 5: clear explanation)
- Soundness of evaluation (0-1: no evaluation; 2-3: using convincing evaluation methods; 4-5: convincing evaluation and further analysis, such as ablation study)

Project (20 marks)

- Novelty of the project (10 marks) (0-4: reusing previous work taught in MQ units; 5-7: relevant to existing work but new implementation given the proper references; 8-10: proposing a completely novel method or framework)
- Effort in the project (10 marks) (add up to at most 10 points: data processing +1 to +2; project management +1 to +2; training models +1 to +3; trying novel methods +1 to +3; evaluation / analysis +1 to +3) [+1: week; +2 good effort; +3: outstanding in the class]

4 Notes

Am I allowed to reuse previous projects or assignments? While you are permitted to employ tools utilized in previous assignments, it's essential to develop a fresh project and solution. The innovation and uniqueness of your work will significantly influence your grade.

Is it permissible to incorporate ChatGPT in my project? Incorporating ChatGPT into your solution is allowed. However, it's important to note that 'project effort' contributes 10 points towards your score. Merely integrating ChatGPT's API without additional development efforts will not suffice for a high mark in this category.

How can I receive feedback? We encourage you to share your project ideas with our lecturers and teaching assistants. We're eager to listen and provide feedback. Time will be set aside during the Workshops in Weeks 11 and 12 specifically for project assistance.

5 Late Submission

Unless a Special Consideration request has been submitted and approved, a 5% penalty (of the total possible mark of the task) will be applied for each day a written report or presentation assessment is not submitted, up until the 7th day (including weekends). After the 7th day, a grade of "0" will be awarded even if the assessment is submitted. The submission time for all uploaded assessments is 11:55 pm. A 1-hour grace period will be provided to students who experience a technical concern (If the submission is overdue between 1 hour and 24 hours, a 5% penalty will also be applied). For any late submission of time-sensitive tasks, such as scheduled tests/exams, performance assessments/presentations, and/or scheduled practical assessments/labs, please apply for special consideration.

URL for details around special consideration: `https://students.mq.edu.au/study/assessment-exams/special-consideration`