CSCI 4342 Project Requirements and Guidelines

SEM 1 2024/25

Project Guidelines

Your research project will study an approach to address some NLP task or problem and will include empirical and/or qualitative evaluation (i.e., human judgement). Keep in mind the relatively short time available and available data when considering the feasibility of your project. Identify intermediate goals to complete this project.

Some guidelines:

- Describe and motivate your task
- Argue why your problem is important
- Argue why your problem is challenging
- Argue why your approach is reasonable
- Describe what you did so far (i.e, experimental setup)
- Show us your data: use examples for all parts of your presentation
- If your task is multi-disciplinary (e.g., NLP + security), describe what aspects of language will play a role
- Description and motivation of your technique and method
- Discuss related work that inspired you and/or justified your approach
- Discuss feasibility given available resources (relying on prior work is valid)
- Identify at least one corpus you will use (more is better!). You may collect and create your own corpus. If you collect your own corpus, describe the processes and do a quick feasibility study
- It is recommended that you **use examples and figures to illustrate the problem and your approach**

Some examples of feasible small research (but not limited to) projects (as per selected areas for paper reviews)

- Language modeling/Large Language Modeling
    - Generating poetry/idioms from language models
    - Generating lyrics from language models
    - Creating abstracts/summary of articles using language models and/or deep learning
    - Using word embeddings to represent words as vectors for semantic tasks
    - Generative AI

- Text categorization
    - Fake news detection
    - News headlines topic classification/clustering
- Text Summarization
    - Research articles/Scientific documents
    - Newspaper articles
- Word Sense Disambiguation
    - WSD for parallel/bilingual texts
    - WSD for poetry generation
- Named Entity Recognition
    - NER in emails
    - NER in minutes of meeting
    - NER in medical documents
- Sentiment Analysis/Text analytics
    - Product reviews
    - Stock prediction
    - Emotion sentiment analysis
- Conversational Agents/Question Answering
    - Frontdesk chatbot
- Machine Translation
    - Medical document bilingual translation
    - Legal document machine translation
- Semantic relation identification
    - Word Embeddings
    - Transfer learning/transformers
- Recommendation System
- Speech Analytics
    - Speech analytics in healthcare : Handles patient inquiries, appointment scheduling, prescription refills, and other services
    - Contact Centers in E-commerce : Handles customer inquiries across multiple channels such as phone, email, online chat, and review management.

***NOTE: Group projects on topics related to email classification, spam filtering or anything similar are no longer allowed due to its overuse.**

Project Presentations (Week 14)

Face 2 Face Presentation (Classroom: Tue, 14 Jan & Thu, 16 Jan 2025).

- Please describe your project and defend it.
- Each group will be allocated a maximum of 15 minutes, followed by an additional 5 minutes of Q&A session. I will stop you after 15 minutes.
- Each group will be randomly assigned a presentation slot and need only to attend their own slot.
- All groups are required to arrive at least 10 minutes before the presentation time to conduct the necessary setup for your project presentation. Please respect the time of others by starting and finishing your presentation on time.
- It is recommended that you use examples and figures to illustrate the problem and your approach. Please present your approach, conducted experiments and analyses of results. You may dedicate less time to the related work.

Technical Report

Due: Sunday, 26 Jan 2025 (11.59 pm)

The technical report is limited to a maximum of **8 pages** (including references) in **ACM** format. The report should be written as a research paper. Please include your slides in PDF form together with your source codes/models and upload them together as a zip file.

This is a short report, so organize your content well. It is recommended to fashion your report after good short papers as per your paper reviews. There are plenty of excellent examples!

Suggested Table of Contents

- Introduction (including problem statement, motivation & objectives)
- Related work (minimal)
- Technical background
- Formal well-motivated description of your approach
- Dataset description
- Experimental setup & implementation/test examples
- Experimental results & Discussion
- Error analyses