CSCI544: Project Topics & Expections

A primary component of the course is group projects. This document provides more context about expectations and grading criteria for your projects.

1. Project types

The goal of your course project is to solve an interesting problem. It offers you the chance to apply your NLP knowledge on an in-depth application and expand your mastery of using NLP in a focused area.

There are two possibilities for the projects:

1. You select a problem that you are excited about and then solve the problem using NLP. Examples of potential project directions include:

a. Select a routine NLP task and use it in an unexplored domain, e.g., sentiment analysis of Yelp reviews of restaurants in Los Angeles. Solve the routine task and analyse your results, e.g., what features in a restaurant will result in positive/negative experience of costumers or is there any correlation between the location of restaurants and costumer experiences. Consider checking the paper "Predicting wins and spread in the Premier League using a sentiment analysis of twitter" as an example for this type of project.

b. Select an existing algorithm, e.g., Word2Vec, try to use it in several languages to preform several prediction tasks. Analyze the results, e.g., whether the performance depends on the domain language. Consider checking the paper "CNN features off-the-shelf: an astounding baseline for recognition" as a good sample for this type of project. In other words, the goal is to evaluate how good existing tools are in practice to tackle a challenge.

c. Select an existing technique or algorithm in a non-NLP domain, e.g., vision, and try to adopt it for NLP. Explore whether that technique or algorithm can lead to some benefits in NLP problems. Consider checking the pa-

per "Learning to Solve NLP Tasks in an Incremental Number of Languages" as an example.

d. Select an existing technique or algorithm and come up with an improvement or addressing a more challenging setting at which the original algorithm would fail. Check the paper "Two/too simple adaptations of word2vec for syntax problems" as an example.

e. Select a classic research problem in NLP, e.g., unsupervised domain adaptation, and develop a new algorithm to address that problem.

You can see that the above types are not homogeneously difficult in terms of the required novelty and creativity. To make grading fair, expectations for analysis and exploration will be higher if your project does not entail algorithmic creativity (type a. vs type e).

2. Select an NLP paper, implement the proposed work, and regenerate all the reported results in the paper. Additionally, you should also perform exploration tasks:

a. Change the modules in the paper data processing pipeline and study study the algorithmic performance under the new conditions.

b. Explore generalizability of the proposed algorithm beyond the presented results in new settings and scenarios and discover weaknesses of the paper.

For both types of projects, you can select the topic or paper on your own or in consultation with your project advisor. However, you need to ask your advisor's feedback.

Our expectation is that you go beyond the course homework in your projects and solve problems that entail learning skills beyond hoemwork assignments. Don't forget that we will provide feedback on your proposals to align your projects with expectations. Please use the remaining time and the office hours until the proposal deadline to finalize your topics.

Computing Resources

A major aspect that you should make sure to have for selecting your topic is that you make sure you have enough computing resources to perform your projects. Some students may have personal resources, but if you don't have such resources, you can try using free computational resources available for students. Azure and Amazon both provide a limited amount of free computational power for students. Google also has provided this course some

resources. Given that each team has five members, you may be able to aquire up to $1000 worth of computational resources. Some students already have applied successfully to gain these resources.