**CS410 Final Project Proposal 10/25/2020**

Team name: The Tardy Slackers

What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

1. Tom McNulty - Captain NetID: tmcnu3 [tmcnu3@illinois.edu](mailto:tmcnu3@illinois.edu)
2. Wei Dai NetID: weidai6 [weidai6@illinois.edu](mailto:weidai6@illinois.edu)
3. Michael Huang NetID: mh54 [mh54@illinois.edu](mailto:mh54@illinois.edu)

Which competition do you plan to join?

The text classification competition - identifying sarcasm on Twitter

Are you prepared to learn state-of-the-art neural network classifiers?: Yes - of course!

Name some neural classifiers and deep learning frameworks that you may have heard of.

We have done some research to identify good classifiers, but have not chosen exactly which one(s) we will actually use yet. Those we have heard of are

1. Transformer-based models such as BERT, ALBERT, RoBERTa, XLNet, UniLM, Multi-Task DNN
2. LSTMs, GRU, RNN
3. Word Embeddings i.e. word2vec, glove, ELMO
4. CNNs

Describe any relevant prior experience with such methods

Michael has experience with using these architectures. In undergrad, he participated in research projects that utilized LSTMs for NER tasks and also CNNs for some computer vision research. Also at work, he utilizes many transformer-based models to solve text-classification tasks as well as language generation tasks. (Side note: Although he does have this experience, he still finds this project appealing as it allows him to focus on improving other areas he’s been meaning to work on such as writing good software, scripts, and documentation that makes it easy to install, use, and maintain)

Wei has experience with data collection. At work, she utilizes many tools such as Json, goose-extractor to web scrape and crawl to build datasets for further use.

Which programming language do you plan to use? - Python, most likely with TensorFlow for modeling

Additional

If we are able to achieve state-of-the-art performance in a reasonable amount of time, we would like to expand the capabilities of our program to do one or more of the following:

1. Identify a rate of sarcasm or create a sarcasm score on the Twitter accounts of some well-known people such as comedians, politicians, flamboyant business people, etc.
   1. The score would be based on the person’s entire tweeting history - or as much as we can get a hold of.
2. Develop a capability to determine metaphors / idioms based on comparing figurative and literal interpretations of tweets.
   1. These concepts were presented at a recent UIUC presentation by Suma Bhat who is an Assistant Professor in ECE at Illinois
   2. <https://mediaspace.illinois.edu/media/t/1_0db6ad18>
3. Deploy our model with a REST API, so that it is easily interactable and usable by others, and it would also make for a cool demo