CS 410 Final Project Progress Report

Topic: Stock Sentiment Analysis Using Twitter Feed

Team: Porcupine Tree

Team Captain: Zachary Oldham

Zachary Odham	NetID: zoldham2
Yogeswara Rao Lekkalapudi	NetID: yrl3
Luis Mariano Ovalle Castaneda	NetID: lo22

1. Which tasks are completed?
    a. The creation of the sentiment analysis system has been completed. This was accomplished with a recurrent neural network that was implemented using PyTorch. The system takes arbitrary labeled text strings as input, and uses them to either train or test a neural network, depending on how the data is labeled. Based on preliminary testing using data obtained from Sentiment 140, the accuracy of this classifier is about 70% when trained with 40,000 labeled tweets over 20 epochs. We will attempt to improve the performance of this system for the final product by adding additional preprocessing, and training for longer using a larger dataset.
    b. Testing querying using the Twitter API has been completed. We have figured out how to retrieve tweets using the API, and how to retrieve tweets based on specific keyword searches. We will use this system to retrieve the data we display to the user in the final system.
2. Which tasks are still pending?
    a. Identify socks to query from twitter, map company names to stock code for reverse lookup of the codes for input queries
    b. Build full set of documents for ranking model
    c. Build inverted index for the documents
    d. Build ranking model and evaluate
    e. Create interface to obtain user input and display the results to the user
3. Any challenges/issues being faced?
    a. Understanding twitter query parameters to fetch relevant data was tricky. Searching for a company name vs stock code returned different feeds. For our project, we are only fetching stock related feeds.
    b. We need to improve the performance of the sentiment analysis system. We strongly suspect this can be accomplished by adding preprocessing and training for longer with more data, but this needs to be investigated.