18-755: Networks in the Real World
Fall 2023

Tentative project title: Twitter Sentiment Analysis Using Language Networks

Project team members and ids:

Zhonghui Cui(zhonghuc)
Yudi Chen(yudichen)
Weigen Chen(weigenc)

Overview of the project topic:

In the age of social media, platforms like Twitter have become a rich source of public opinion and sentiment, and understanding the sentiment behind Twitter comments can provide valuable insights for various applications. This project aims to investigate and analyze the language networks behind a dataset of twitter comments, while harnessing natural language processing (NLP) models to categorize the sentiments expressed in Twitter comments. We will explore the similarities between words when it comes to expressing positive, neutral or negative sentiments and analyze sentiments in Twitter comments on a large scale.

Goal of the project:

One goal of this project is to identify network representation of the English vocabulary, where nodes represent words, and weighted edges represent the sentimental similarity. We aim to develop algorithms and metrics to quantify these relationships and identify meaningful clusters or communities within the vocabulary network. Additionally, we seek to use the characteristics of the networks to achieve the sentimental analysis of a dataset making it accessible to researchers and linguists. We would use text analysis techniques(NLP) to interpret each comment and classify them into three emotions (positive, negative, and neutral).

Dataset and networks:

For this project, we will investigate the sentiment140 dataset. It contains 1,600,000 tweets extracted using the twitter api. The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment. This dataset will serve as our foundational resources for establishing the networks representing word relationships. Our networks will consist of nodes representing individual English words and edges connecting words with similar sentimental characteristics.
Dataset:
https://www.kaggle.com/code/stoicstatic/twitter-sentiment-analysis-for-beginners/input

Division of labor:

Weigen Chen will be responsible for data collection and preprocessing, focusing on acquiring and cleaning linguistic datasets and building the network graph.

Yudi Chen will work on the NPL model to implement sentiment analysis, developing algorithms for measuring word relationships to decide whether the comment is negative or positive..

Zhonghui Cui will handle the creation of the user-friendly visualization tool, ensuring that the network's insights are easily accessible.

All team members will collaborate on fine-tuning the network analysis and the machine learning model. Regular team meetings will facilitate communication and project progress tracking.