CS410 Team Project Proposal

Team Info

Name	NetID
Xiao Luo (captain)	xiaoluo6
Shi Zhang	shiz4
Wenyan Zhang	wenyanz4
Arindam Chakraborty	arindam4

Topic Overview

Topic

Twitter/X sentiment analysis

Description

We plan to implement a system where users can enter topics of interest, view the relevant Twitter (rebranded to “X”) posts as search results, and view the overall sentiments on the topics.

Task

We broke down the topic into the following sub tasks and distributed among the team members:

1. Front end experience: Develop a front end user interface where users can enter topics of interest, click on ‘search’, and view the relevant tweets and sentiments.
2. Scraping Twitter/X: Access tweets posted during a certain time period
3. Text retrieval: Given a topic of interest, retrieve a ranked list of tweets based on relevance
4. Sentiment analysis: Based on the top X retrieved tweets, analyze the sentiments

Why are we interested in the topic

Twitter provides a platform for users to share information and updates in real-time and is a comprehensive source for staying current with the latest news, trends, events, etc. Therefore, Twitter has become a valuable source of data for sentiment analysis on topics of interest.

Using Twitter data for sentiment analysis can provide insights for individuals, businesses, researchers and organizations. Below are a few examples of use cases where the results from this project can be leveraged in the real world applications:

- **User and customer feedback**: Consumer facing companies oftentimes need to gather feedback on their products or services through various channels. Twitter is one of the channels with rich feedback information, e.g. comments, praises, complaints. Being able to directly fetch the sentiments towards the target topics using Twitter data helps the businesses understand customer satisfaction, identify areas for improvement, monitor the brand reputation.
- **Competitor analysis**: Through understanding the sentiments towards competitors' products or brands, businesses can gain insights into market trends, identify gaps in the market, and improve their competitive strategy.
- **Sentiment-driven advertising**: Advertisers or brands can use sentiment analysis to target ads to specific user segments, ensuring that the ads are more relevant, engaging and effective.

Planned approach

We plan to access Twitter data via Twitter API, process text using packages such as NLTK, retrieval relevant text via packages such as Gensim, and perform sentiment analysis using packages such as Hugging Face.

Besides, we plan to build a front end user interface to allow users to enter the topics of interest and see the retrieved tweets as well as the overall sentiment.

Tools and datasets

Tools & systems:

- Pytorch: deep learning training
- NLTK/spaCy: text processing
- Gensim: topic modeling, document indexing, retrieval by similarity, etc.
- Hugging Face: sentiment analysis

Dataset:

- Tweets data (via Twitter API)
- If we can't get data from Twitter API, we will get Tweets dataset from Kaggle

Expected outcome

We expect to deliver a system where users can enter topics of interest, view the relevant Twitter posts as search results and view the overall sentiments on the topics.

Evaluation of Work

We are going to adopt the following evaluation approaches:

- Use a ground truth Twitter dataset with labeled sentiments to measure the performance metrics of the system (e.g. precision, recall, F1)
- Use a list of test queries to validate if the test queries can trigger the expected results

Programming language

We plan to use Python for text retrieval and sentiment analysis, and create a web page using HTML/JavaScript to display and visualize the results

Workload justification

The sub tasks of this project include front end experience, scraping Twitter/X, text retrieval and sentiment analysis, resulting in ~80 hours of workload for the team.