Twitter Sentiment Analysis of the Brexit result

Surya Vajjhala
U19590925
Email vajjhala@bu.edu

Zhi Dou
U21392913
Email zhidou@bu.edu

Nature of the Dataset

We dataset consists of tweets, from Jan 1, 2016 to Oct 1, 2016

Collection:

We scrape data with tag #Brexit using Twitter API

Format:

Twitter API provides the data in JSON format. These data also called tweets are the basic atomic building block of all things in Twitter. It contains keys like: 'text', 'retweeted_status', 'favorite_count', 'metadata', 'user', 'created_at', 'source', 'retweeted', 'coordinates', 'possibly_sensitive', 'truncated', 'contributors', 'is_quote_status', 'id', 'retweet_count', 'lang', 'place', 'entities', 'favorited'.

For our project, the keys we may need are 'text', 'location', 'lang'(language) , 'created_at'(time the tweet was posted).

Preprocess:

Because the data got from API has more information than we need, we should parse out the raw data extract the attributes we would use and save it into cvs file for later usage.

Expected Analysis

Latent Semantic Analysis

In order to get information from each tweet, we should do latent semantic analysis on the texts , to break texts into words, to remove meaningless high frequency words, and to reduce the dimension we may need for further step.

Sentiment Analysis

To get the attitude of each user towards Brexit, we should implement sentiment analysis on each tweet.

Clustering

We would cluster the data based the general opinion and location by using K-means and GMM clustering techniques.

Application

Based on our work, we could analysis the trends in the Brexit vote. We could answer questions like below:

1. Prediction of the Brexit vote based on twitter data.
2. Does UK regret the Brexit?
3. General opinion of others nations (more specifically other countries of EU) towards the result of the referendum.
4. A timeline of variation of public opinion leading up to the vote.
5. Demographic insights of the Brexit vote.

We could get the answer for most of these questions, after we perform sentiment analysis on the data. And through our analysis, we could see the big picture and trends on Brexit.

Expected Results

1. We might validate that Twitter could have correctly predicted the Brexit vote result.
2. We expect that there is a large population that regrets the result of the voting (based on the tweets after the result).
3. We cannot make assumptions until we analyse the tweets from each country.
4. Show a plot of changing variation of people's opinion and try to correlate it to teh major socio-political events (like the immigration crisis, bombing in Paris, 2015 UK general election, Greek bailout referendum, etc.).
5. Remain camps clustered around major cities and Leave camps around the rural and country side.