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 are tweets of twitter, 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', 'language', 'tweets created time'.

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 text of each tweets, we should do latent semantic analysis on the texts we get, 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 the data we have.

Clusterring

For mining the relationship between each data, we would cluster data the opinion and location of each data. We may use Kmeans and GMM to do clustering.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

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 of most of this question, after we finish sentiment analysis of the data we get. And through our analysis, it is easy to see the big picture and trends on Brexit.

Expected Results

1. We might validate Twitter could correctly predict the Brexit vote result.
2. We expect that there is a large population that regrets the result of the voting.
3. We cannot make assumptions until we analyses tweets from each country.
4. Show a plot of changing variation and try to show the changes actually related to major socio-political events (like the immigration crisis, bombing in Pari, 2015 UK general election, greek bailout referendum, etc.).
5. Remain camps clustered around major cities and Leave camps around the rural and country side.