Dong Hyun Kim

Professor. Li

Problem Solving and Software Design

Text Mining

**Project Overview**

I decided to use Twitter as a data source for this project, because people actively share their ideas through Twitter at any time. Therefore, I assumed that Twitter can be a very useful source for analyzing trend of people’s thoughts regarding one incident or a person. Since people like to share their ideas through Twitter, I though sentiment analysis is the best tool to read how people think about the famous soccer player Wayne Rooney before his last game as a part of England national team.

**Implementation Plan**

First of all, I had to install re, tweepy, and textblob packages to python. The ‘re’ package allows me to remove any special characters or links from the text. The ‘tweepy’ allows to get access to Twitter API through python. Lastly, ‘textblob’ is used to process textual data I gathered from Twitter.

Then, I set up the whole function to get data and analyze sentiment of people on Twitter. First, I entered my authentication keys in order to get access to Twitter. Then, ‘clean\_tweet’ function through ‘re’ packages removes any special characters or links. Next, tweet\_sentiment checks polarity of the tweet by using textblob package. If it is bigger than 0, it is positive and vice versa. Lastly, get\_tweets fetches and parses tweets. In addition, it stores textual data from Twitter, so that I can actually see tweets on python. Then, tweets will be categorized into positive or negative depending on their polarity. Also, python will calculate percentage of positive and negative tweets.

There are several ways to analyze sentiment of people regarding Wayne Rooney. One way is to only show every tweets about him. However, number makes it easier to see the result intuitively. Therefore, I decided to get percentage of positive and negative tweets; therefore, I can easily see the trend of sentiment among people.

**Results**

|  |  |  |
| --- | --- | --- |
|  | Positive | Negative |
| Percentage (%) | 36.144578313253014 | 15.662650602409638 |
| Tweet | I’m sorry too Coleen, Wayne deserves to put on the  England shirt again, especially for such a great charity. | Does the inclusion of WAYNE ROONEY in the ENGLAND squad for game against the USA make a mockery of INTERNATIONAL FOOTBAL… |

As you can see from the table, there are more positive tweets than negative tweets. Since he is the living legend of England soccer, having more positive tweet is reasonable. Also, around 49% of people have neutral sentiment toward the retirement of Wayne Rooney.

Among tweets, people mostly praise Rooney’s contribution and accomplishments. And they still support Wayne Rooney. Interestingly, some people feel offended to include retired player to international squad only to commemorate. Other negative tweets are about expensive price of tickets. Therefore, people may have negative sentiment on the event itself, not on Wayne Rooney.

**Reflection**

Since I learned about sentiment analysis on the other classes, but it was good to learn new tool and ways to conduct sentiment analysis. Python provides easier method and better results than other tools, if it is well designed. Moreover, I could find a loophole of sentiment analysis. Since python only gets textual data, not the context, sometimes it is hard to analyze whether negativity actually goes toward the person I tried to search for. Like in tweets, some people express negativity at the price of ticket, not Wayne Rooney; however, python still recognizes it as a negative tweet toward Wayne Rooney. As a result, there is a need to improve sentiment analysis to allow the program to understand the context and increase the accuracy of sentiment analysis.