Implementation of a Twitter Sentiment Analysis Tool

Lean Miguel | 10160542
Computer Science 571

Abstract

Twitter holds a vast amount of data that can be extracted for analysis. The Twitter Sentiment Analysis tool is first given a query, subsequently finds tweets related to that query and displays their sentiment value in a bar graph. A user may also download a CSV file to see each tweet's individual sentiment value. It is created with several different Python libraries, and performs similarly to other lexicon based Twitter sentiment analysis tools that are on the web currently. This paper reports on the creation of the Twitter Sentiment Analysis Tool, as well as others works related to sentiment analysis and how the Twitter Sentiment Analysis tool compares to them.

1. Introduction

Sentiment analysis is the classification of whether a given text is positive, negative, or neutral. Sentiment analysis is a useful tool for companies to identify public opinion on any given topic. Sentiment analysis is powerful for being able to gauge customer satisfaction with one's product. The Twitter Sentiment Analysis tool presented displays a bar graph which shows the general public's sentiment towards the topic with a simple graphical user interface.

2. Related Work

There are currently several sentiment analysis tools that work on Twitter currently available on the web. Some sentiment analysis tools on the web use machine learning to analyse data while some use lexicons.

2.1 Sentiment140

To use Sentiment140, A user must first log into their own Twitter to be able to input their own search query. Sentiment140 uses machine learning algorithms [1] to classify tweets. The classifier that is used by Sentiment140 is a maximum entropy classifier. [2]

2.2 Tweet Sentiment to CSV

Tweet Sentiment to CSV is a simple analysis tool which uses a lexicon [3] (namely TextBlob library) to classify tweets. Tweet Sentiment to CSV also has a feature which allows you to put the searched tweets sentiment in a CSV file.

2.3 SentimentViz

SentimentViz [4], which stands for sentiment visualization, is the most sophisticated of all the online twitter sentiment analysis tools. Instead of just sentiment, SentimentViz presents to the user a map of emotions for the searched query. In

addition to the emotion map, SentimentViz also provides heatmaps and timelines of the tweets. An interesting feature that SentimentViz supports is the TagCloud which groups up the hashtags and puts them into a word cloud separated by emotion scale (Pleasant, Unpleasant, Active, Subdued). The lexicon that SentimentViz uses to classify data is a modified version of the ANEW dictionary. **[5]**

3. Discussion of the Related Work

3.1 Sentiment140

Sentiment140's requires a user to log into their own Twitter account which may prove troublesome for people who don't have an account, or don't want to create an account. Another problem that Sentiment140 has is that the amount of tweets retrieved are always below 100. Sentiment140 also does not show the amount of neutral tweets there are in the query.

3.2 Tweet Sentiment to CSV

Tweet Sentiment to CSV's is a standard sentiment analysis tool that is very easy to use. It also allows user to download a CSV file. A problem that Tweet Sentiment to CSV has is that it only shows how many tweets are positive, negative, or neutral. IT does not display the percentage of tweets. The colors used in the pie chart are not intuitive, where green being neutral, and positive being blue.

3.3 SentimentViz

While SentimentViz is the most sophisticated of all the tools listed above, it is hard to classify the general opinion of the public. You can identify whether a given query is positive or negative, but are not able to identify the most common emotion found within the query. Some of the features that SentimentViz are not intuitive. The heatmap's intended function was difficult to understand. All these problems are very minor in comparison to the functionality SentimentViz provides.

4. Implementation

4.1 Implementation

The Twitter Sentiment Analysis Tool is a graphical user interface created on Python. It uses multiple libraries such as Tweepy for accessing the Twitter API, and Tkinter for creating the GUI. It uses the VADER library to classify the tweets. The tweets retrieved are limited to a 100 as anymore than that causes the result retrieval of the query to be very slow. In terms of ease of use, it is a very simple graphical interface that anybody should be able to use.

4.2 Usage of VADER

There were multiple reasons going with the VADER library as opposed to other libraries like TextBlob. VADER library performed better than other sentiment analysis tools such as the ANEW, and machine learning algorithms when the text inputs were Twitter tweets.[**6**] Vader also has general tools that help accurately identify Twitter text. For example, tweets such as "The food here is great, but the service is terrible", VADER is able to identify that the "but" causes a shift in the sentiment polarity, identifying it not purely as a positive sentiment.
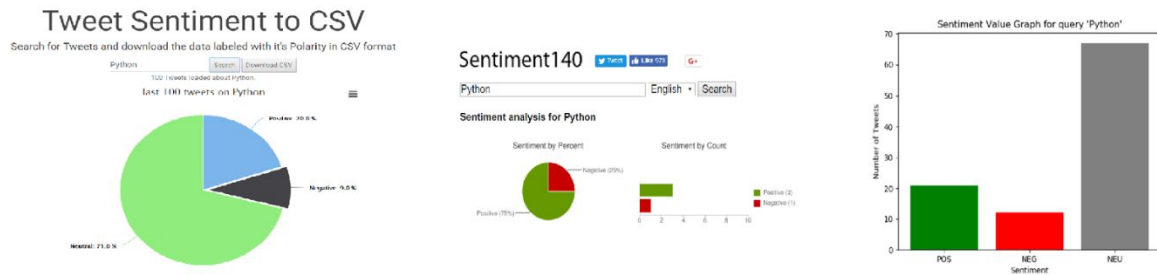
Figure 1.1: Comparison of results for "Python" query between Tweet Sentiment to CSV, Sentiment140, and the Twitter Sentiment Analysis Tool

	Tweet Sentiment to CSV	Sentiment140	Twitter Sentiment Analysis Tool
Positive	20	3	20
Negative	9	1	12
Neutral	71	N/A	68

Figure 1.2: Comparison of results for "Python" query between Tweet Sentiment to CSV, Sentiment140, and the Twitter Sentiment Analysis Tool

4.3 Suggestions

In terms of sentiment analysis, while a user can view the number of tweets related to their respective sentiment, users cannot view the percentage of positive, negative, or neutral. It also cannot modify the amount of tweets to retrieve and is fixed to 100 tweets. A filter for retweets would also be useful to not skewer the results to whatever the majority of retweets is.

4.4 Comparisons with Sentiment140 and Tweet Sentiment to CSV

Sentiment140 is a machine learning algorithm based sentiment analysis tool while Tweet Sentiment to CSV is a lexicon based (uses TextBlob library) sentiment analysis tool. Referring to Figure 1.2, the results generated from Tweet Sentiment to CSV are very similar to the results produced with the Twitter Sentiment Analysis Tool. The results generated from Sentiment140 are difficult to compare as Sentiment140 does not show neutral tweets, and also Sentiment140 also only retrieved 4 tweets in total, which should not be comparable to the 100 tweets that the Twitter Sentiment Analysis Tool retrieved.

4.5 Problems

Since VADER is a lexicon based sentiment analysis tool, queries that have a negative word in them will tend to skew to the negative side of general opinion. For example, querying for the movie Star Wars or Avengers: Infinity War which are currently very popular movies, are being "viewed" by the general public as negative.

This is due to the fact that word "war" is classified by VADER as a very negative word, therefore causing the majority of tweets retrieved to be classified as negative. The results are more normalized if someone were to query "Starwars" instead of "Star Wars".

5. Summary

The Twitter Sentiment Analysis Tool is a tool that queries twitter and retrieves Twitter's general sentiment opinion. Sentiment analysis is used by companies to gauge how their product is being received. There are multiple tools currently out on the web that can perform sentiment analysis on twitter, such as Sentiment140 and SentimentViz. The Twitter Sentiment Analysis Tool is created with Python and uses the Tweepy, Tkinter, and VADER libraries.

6. Future Work

Currently the Twitter Sentiment Analysis tool only has features like analysing query and bringing all the tweets retrieved from the query into a CSV file. Other features that can be implemented into the Twitter Sentiment Analysis Tools are things such as word clouds that identify the most frequently used words in a search. Another feature that can be added is the option to use different classifiers, such as using TextBlob instead of VADER, or even a machine learning algorithm.

7. Conclusion

The Twitter Sentiment Analysis tool is an easy to use graphical user interface that provides the Twitter's general opinion of a given query. The sentiment analysis provided by the Twitter Sentiment Analysis Tool is comparable to other sentiment analysis tools on the web.

7. References

[1]http://help.sentiment140.com/

[2]http://help.sentiment140.com/for-students

[3]https://github.com/rhnvrm/labeled-tweet-generator/blob/master/twitter.py

[4]https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

[5] Estimating Sentiment Section:
https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/

[6] C.J. Hutto, Eric Gilbert. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text - Section 3.2 :
http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf