Wright State University

CS 7900

Final Report

A Sentiment Analysis Application for Tweets

Author:
Daniel McIntyre U00653482
Nathan Rude U00637118

Supervisor:
Dr. Tanvi Banerjee

April 16, 2016

1 Goal

Social media is a very ingrained component of our culture. In this day and age, there are social media users all over the world. In fact, social data is a very rich source of data in todays information age. For this project, we aim to leverage the power of social data from Twitter. As a micro-blogging outlet, users regularly churn out bite-sized chunks of information that can be analysed for information on just about any topic. By aggregating tweets about a certain topic from Twitter, one can get a general sentiment from the populous about that certain topic.

Since tweets can be retrieved in real time, the sentiment of current and trending topics, such as presidential debates and debate topics, can be evaluated to determine how the populous feels about each topic. Understanding and incorporating this user feedback, specifically by examining tweets with a negative sentiment, one can receive the feedback required to move towards a more popular opinion.

2 Objective

We created a web application that allows users to view the sentiment of topics on Twitter. Through the use of two APIs, we 1) retrieved tweets from Twitter and 2) performed sentiment analysis on these tweets and also aggregated the general sentiment about a topic.

3 Approach

This application can be split into two parts: a frontend for the user to select topics and view the sentiment, and a backend for the text retrieval and sentiment analysis to take place.

The user interface will be developed with HTML, CSS and Javascript, technologies that have been learned in class. Using these technologies we can create a dynamic layout that, given a search topic, can display several retrieved tweets with their individual sentiment, as well as the overall sentiment of a topic.

The backend will handle the routing and logic for extracting and analysing tweets. Backend technologies will be used based on compatibility with the chosen APIs. For simplicity, sentiment analysis will be performed at tweet level, rather than at phrase or entity level.

4 Evaluation

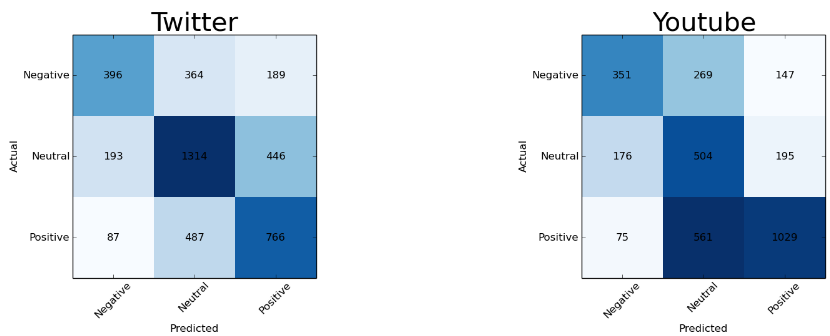
To evaluate the sentiment analysis component of the application we used the following two sub-datasets from the Sentiment Strength Twitter Dataset (SS-Tweet) [4]: 1) Youtube comments and 2) twitter tweets (see Table 1). Using data from two domains allowed us to better evaluate the robustness of our sentiment analysis tool.

To keep the evaluation simple, we performed binary evaluation for this multiclass problem. The testing datasets were labeled with a mean positive and mean negative value for each sample. To calculate the sentiment, we define *sentiment = good - bad* and then threshold it into a class *negative, neutral, positive* based on it's value. To compute precision and recall for a multiclass problem, we transformed the problem into several one-vs-many problems, calculated the precision and recall for those, and then took the average values for the multiclass problem.

The results of the sentiment analysis evaluation can be seen in Table 2 and Figure 1. We can see that the Twitter and Youtube datasets have similar precision and recall values, and this implies that this sentiment module is robust across different text domains. Now while the precision and

Dataset	Samples	Positive	Neutral	Negative
Twitter	4,242	1,340	1,953	949
Youtube	3,307	1,665	875	767

Table 1: Dataset class breakdown



(a) (b)

Figure 1: Sentiment Confusion Matrices

recall values do not seem that high, it's important to realize that not even humans can accurately determine the sentiment of a text 100% of the time. In fact humans can only determine the correct sentiment 70-80% [1][2][3] of the time! This is because humans do not universally agree with each other on any subjective matter.

Examining the confusion matricies in Figure 1 we can see that the algorithm mostly has trouble with the neutral class. With sentiment being such a subjective issue, it intuitively makes sense that there would be more trouble evaluating whether a text is *positive vs neutral* or *negative vs neutral*. The lines can be more gray here rather than the stark contrast that a *positive vs negative* tweet would have.

Dataset	Precision	Recall
Twitter	.55	.58
Youtube	.53	.55

Table 2: Sentiment Metrics

Since the algorithms performs much better than just guessing the class (which would be a 33.3% precision), we conclude that the performance of this algorithm is adequate for the task at hand.

5 Conclusion

References

[1] brnrd.me. On social sentiment and sentiment analysis. http://brnrd.me/social-sentiment-sentiment-analysis. Accessed: 2016-04-16.

[2] brnrd.me. Sentiment analysis: Why its never 100 http://brnrd.me/sentiment-analysis-never-accurate. Accessed: 2016-04-16.

[3] Mashable. How companies can use sentiment analysis to improve their business. http://mashable.com/2010/04/19/sentiment-analysis. Accessed: 2016-04-16.

[4] Hassan Saif, Miriam Fernandez, Yulan He, and Harith Alani. Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold. 2013.