

Faraz Heravi
Dr. Slobodan Vucetic
14 November 2019

Twitter Sentiment Analysis

The motivation of doing this project is to analyze emotions of tweets in a simplified extent. We live in a world that wants to process online texts as good, bad, happy, sad, etc. There is also a lot of work done on sentiment analysis on Twitter with a high accuracy. For example, businesses use sentiment analysis to figure out the response to their products or services. The problem that I want to solve is to classify and analyze the Twitter account of @TheTempleNews. The Temple News contains the most recent news about Temple University. By using sentiment analysis, the news of the account can be classified into categories.

The work on this project will be divided into two parts: first gathering data and preprocessing it, and then analyzing the data using machine learning. The first step is to gather a certain amount of the most recent tweets depending on how many tweets we need. I will be classifying each tweet into positive, neutral, and negative. Therefore our problem now is turned into a classification problem. All these classification is purely an opinion of mine, but we can say that for news the positivity or negativity for a certain group (such as Temple students) will be similar. The second step in data processing is to “clean” the tweets and then use them as inputs. So for example, extra URLs, mentions, hashtags, or stop words will be removed from the tweet and only important words will be stored.

The second part includes the use of machine learning algorithms to solve the problem of classification. We will use the data that we gathered and use different machine learning algorithms in order to analyze these word vectors and its corresponding sentiment. Based on my research, the use of Keras will be important throughout this project. The algorithms which can be tested are a few. My goal is the final accuracy of the sentiment analysis. My data will probably not be that large due to the 20 hours which we are given, but I will try to gain at least an acceptable accuracy. After looking at a few sentiment analysis algorithms, I figured that Support Vector Machines, Naive Bayes, or deep learning will be the most useful classification algorithms for this purpose.

During the first two weeks, I will spend time gathering as much data as I can. I will try to preprocess initially in order to be confident in my preprocessing, then start manually putting my results in manually. Then after week 2, I will start vectorizing my data, and then testing the algorithms I mentioned. As of today we have covered a bit of Support Vector Machines, but we have not gone over Naive Bayes algorithm. I will have to study a bit on that and its possible hyperparameters, and then use the algorithm from python library. During the last week I will try to finish my results and report on exploratory data analysis (EDA). Since this project uses word, I would have to do my EDA on list of words and their corresponding possible sentiments. The final report will be written on the methods which I had the chance to use and their corresponding results and accuracy. I will report on the success or failure of my data collection and processing, and then I will try new tweets from the same account in order to test application of my project.

I used the following references to study and think about sentiment analysis in terms of Machine Learning:

<https://monkeylearn.com/blog/sentiment-analysis-of-twitter/>
<https://www.lexalytics.com/technology/sentiment-analysis>