Racism, ‘Antifa’, and Police Brutality: Sentiment Analysis on Twitter

Maxfield England
6/11/2020

Abstract

Tweets using hashtags and terms surrounding the Black Lives Matter, police brutality and antifa discourse of June 2020, trained with the VADER sentiment analyzer of NLTK, are used to train CNN and MCP models of sentiment analysis, yielding generally decent accuracy (~83%). Both models found the majority of tweets to be generally majority non-negative, but not by a wide margin: about 43.9% of tweets, according to the models, were found to be explicitly negative. This likely reflects both the dark subject matter, as well as the vitriolic discussion being had between participants among the select Twitter tags and search terms.

Background

In 2020, a year many regard as a year of chaos on many fronts, has brought us a global pandemic, rising international tensions on many fronts, countless natural disasters and crises, and now, global demonstrations against police brutality, predominantly committed against people of color. Twitter, widely known as a global platform for (often polarizing) discussion, serves as a great tool for not only participating in but also analyzing discourse in the court of public opinion, especially in regard to politics. Sentiment analysis, used extensively to analyze and categorize vocal twitter communities in the 2016 US election [1], allows for a broader understanding of the ongoing conversations at hand. While sentiment analysis on the Black Lives Matter movement has been done before, such as in 2016 through text mining [2], more research in current events is absolutely vital given the shifting political landscape, and many developments in this tumultuous year. Here, we will take at sentiments that reflect recent events in regard to the Black Lives Matter movement and related ideas.

Problem Statement

Here, we will explore the Bag-of-Words MCP (multilayer perceptron) and CNN (convolutional neural network) models in regard to separating tweets into the camps of negative and positive (or more accurately, non-negative) in sentiment. We will compare their efficacy in sentiment evaluation against our baseline. We will then look at the overall sentiments of these tweets to see what we can garner about the public discussion by the amount of tweets that register as negative.

Methodology

The following analyses were performed on 5,896 tweets collected over a course of seven days, with approximately 800-1,000 collected each day. Using a total of ten search terms and hashtags trending on day one, split between two prevalent ideologies: those in favor of the Black Lives Matter movement and simultaneously against police, versus those in favor of police and against 'Antifa', a purported organization of anti-fascists considered to be aligned with leftist ideals and against the police and military. Sentiments for each tweet are evaluated by VADER, which are used to train two different sentiment analysis models: a multilayer perceptron bag-of-words model, and a convolutional neural network.

Search terms include '#EXPOSEANTIFA', #'alllivesmatter', '#ProjectVeritas', and '#thinblueline', '#BLUEFALL', '#blacklifematters', '#blacklivesmatter', '#cops', and 'police brutality'.

It's important to note that given the nature of both of discussion in general and twitter as a platform, the dichotomous nature of the terms doesn't imply the point of view of the speaker, in each term; hashtags are used in support and dissent of the topic at hand, and are often used sarcastically. A variety of hashtags that represent both perspectives merely serves to capture as wide a view of the conversations surrounding current events, particularly around police brutality against people of color.

The baseline classification for tweet sentiment is determined using VADER (Valence Aware Dictionary and sEntiment Reasoner), an independent sentiment analyzer, "empirically validated by multiple independent human judges". [2] If VADER can be looked at as a strong approximation at what face-value impressions of a social media post may yield, our models could be said to be effectively trained accordingly. If not, though, then we're simply training our models to mimic VADER output to some degree, in the specific subset of tweets surrounding the Black Lives Matter movement and police brutality.

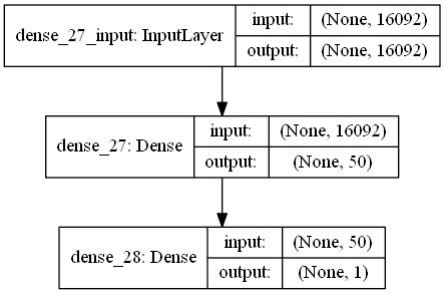
In order to apply VADER results to the CNN and MCP models, the range of values representing the compound score of the tweet (ranging from [-1, 1]) were condensed to a binary ([0, 1], or strictly negative v.s. positive tweets). When evaluating tweets categorized by the CNN and MCP models for data output, only those found to be within 5% of completely neutral (i.e., 0.5) are considered to be 'neutral' tweets; below this range qualifies as negative, and above, positive. When analyzing the non-polarized VADER output, negative values indicate negative tweets, values approximately zero indicate neutral tweets, and positive values indicate positive tweets.

Both models use binary cross entropy loss and the adam optimizer, and a cleaned vocabulary set removing stop words, non-alphabetic characters, and terms that do not appear at least twice in the training set.

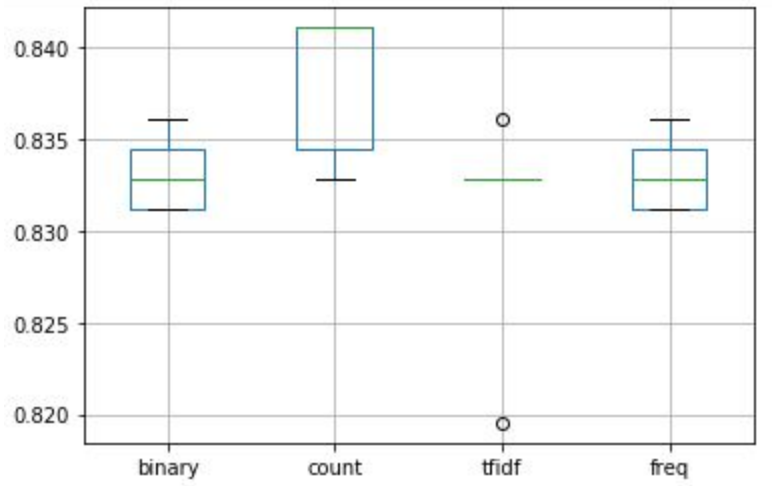
Data

Bag-of-Words:

Model:



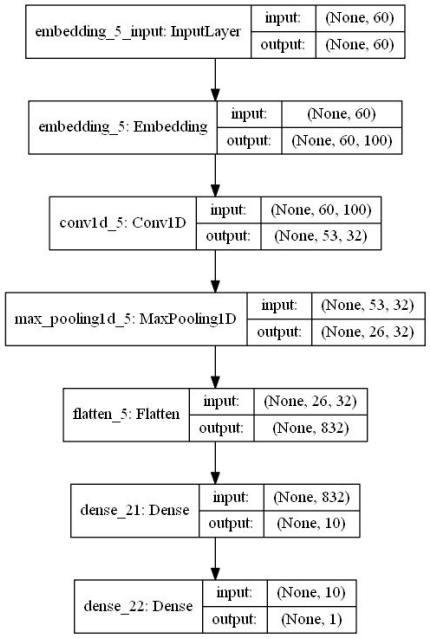
Test accuracy by encoding method:



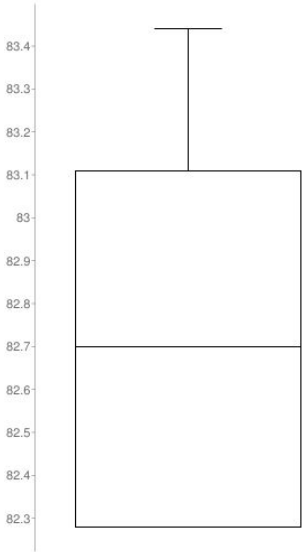
All four encodings perform approximately equivalently, giving an overall accuracy of about 83.1%-83.4% on average across each encoding method.

Convolutional Neural Network:

Model:



Test accuracy across ten trials:



Accuracies varied from 82.3% to 83.4%, with a mean of 82.7%.

Tweet assessment by model (including control assessments)

Model	Positive Tweets	Negative Tweets	Neutral Tweets
VADER Sentiment Analysis (polarized - control)	306	298	N/A
VADER Sentiment Analysis (with neutrality intact)	176	285	143
Bag-of-Words	334	263	7
Convolutional Neural Network	332	267	5

For added insight into the collected data, VADER analyses were also conducted for each of the terms, yielding the following results, using only the VADER's sentiment analysis, based on the compound score generated:

Term	Positive Tweets	Negative Tweets	Neutral Tweets	Total Tweets	Average Compound Score
#EXPOSEANTIFA	136	235	246	617	-0.088
#blacklifematters	250	210	191	651	0.005
#cops	135	247	125	507	-0.170
police brutality	40	515	125	577	-0.611
#thinblueline	57	36	25	118	0.101
#blacklivesmatter	78	88	82	248	-0.023
#alllivesmatter	18	36	24	78	-0.191
#BLUEFALL	144	290	161	595	-0.192
#ProjectVeritas	125	210	143	478	-0.107

Discussion

Upon manually reviewing VADER-assigned scores, a few trends are apparent: many terms regarding protests and violence, as well as mental health and the pandemic, which merely serve as subject matter for a great many of the tweets, appear to flag the tweet as generally negative. Tweets framed in a calm 'matter-of-fact' tone generally tend toward being flagged as neutral. Highly emotive tweets (i.e. expressed in all capital letters) tend to score highly. Here's an example:

Tweet: @realDonaldTrump @SenJohnKennedy @SenBillCassidy GOD SAVE AMERICA. GET RID OF TRUMP. #TrumpResignNow #blacklivesmatter #blacklifematters #BlackLivesMatterUK #blacklivesmatterberlin #BlackLivesMatteritaly
Score: 0.7761

Understanding what 'positive' and 'negative' mean in terms of the data collection is essential in regarding the efficacy of these models. They reflect a combination of the subject matter and the way that that matter is expressed. There are some tweets that I would consider quite positive, that get flagged as negative, such as the following:

Tweet: ✨support my black owned businesses✨ even if it means just retweeting this. Ya girl has been struggling to pay bills after losing her job due to the pandemic. Any kind of support helps! I make handmade pendants ♥#blacklivesmatter #blacklifematters #NoJusticeNoPeace https://t.co/bOTUXFhChG
Score: -0.2698

Also, it's important to note that by virtue of perfectly neutral tweets being flagged as 'positive' tweets before processing in both MCP and CNN models, in order to create a bisected partition of test data, the efficacy of these models must be viewed in a manner that understands this skew. This creates more of a negative-detecting system than one that truly differentiates between 'positive' and 'negative', as only the negative tweets receive a different flag. This also loses a lot of the nuance that VADER provides—for many of the emotive and generally propaganda-esque tweets that fall into the highly positive territory are treated no differently than more neutral-toned ones.

Conclusion

Tweets using hashtags and terms surrounding the Black Lives Matter, police brutality and antifa discourse of June 2020, trained with the VADER sentiment analyzer of NLTK, are used to train CNN and MCP models of sentiment analysis, yielding generally decent accuracy (~83%). Both models found the majority of tweets to be generally majority non-negative, but not by a wide margin: about 43.9% of tweets, according to the models, were found to be explicitly

negative. This likely reflects both the dark subject matter, as well as the vitriolic discussion being had between participants among the select Twitter tags and search terms.

Both models provide similar results, in the ballpark of 83% accuracy, with averages varying less than a percent; and both show a significant amount of tweets in the given topics being negative, although with a moderate non-negative majority (43.5% negative tweets from MCP, and 56.5% non-negative; 44.2% negative tweets from CNN, and 55.8% non-negative).

Compared to the VADER output, which specifies an overall heavily negative trend in tweet sentiments, we see that across the board, there is a strong negative trend to the tweets analyzed. This makes sense for the topic; it's extremely divisive and deals with very dark subjects. Tweets engaging with this these issues are charged with anger, and they concern a grave amount of violence and fear.

Neither model strongly outperforms the other in overall accuracy, and from the counts, it seems that they both made very similar decisions in regard to the test data. Together, they seem to flag a significant amount of false non-negative tweets, compared to the standard VADER provides.

Looking at the VADER evaluations of separate terms, we see that the most strongly negative terms appear to be 'police brutality', '#cops', and '#alllivesmatter'. '#blacklivesmatter' and '#thinblueline' both are the only terms that are, on average, more positive than negative. Tweets under these tags are perhaps more likely to be in support of the topic mentioned; #blacklivesmatter in regard to black people, successes of the movement, and the movement as a whole; while #thinblueline will likely serve mostly as support of the police force and the role they're viewed to maintain in society.

Further Research

While interesting enough at analyzing the broad tone of conversation, more fine-tuned research could illuminate much about the discussion that I was not able to here. Manually assigning tweet sentiment might lead to more refined models that are better fit to analyze the nuances between different posts regarding tough, negative subject matter. I do, of course, also feel concern that by using a different sentiment analysis model (VADER), I may simply be training these models to parrot VADER output. Further, I think different parameters of sentiment analysis (such as political position of the tweeter, assuming tweets are by majority made in good faith without sarcasm or are intentionally misleading, both common on the platform and in the arena of public discourse) could provide interesting data, and allow for some degree of quantification of the online discussion of such topics.

Further, the sample size of tweets is smaller than I anticipated; search terms did not retrieve an even amount of tweets. Controls for data (either an even amount of tweets per search term, or a proportional amount of tweets based on the amount of Twitter discussion that contains the terms) that expand search terms and increase the pool of data operated on could yield different results.

References

1. Caetano, J., Lima, H., Santos, M. *et al.* Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. *J Internet Serv Appl* **9,** 18 (2018). https://doi.org/10.1186/s13174-018-0089-0

2. Ravina, M. (2016, May 31). Sentiment Analysis and The "Black Lives Matter" Movement. Retrieved June 12, 2020, from https://scholarblogs.emory.edu/clioviz/2016/05/31/sentiment-analysis-and-the-movement

3. Hutto, C. (2020, May 22). GitHub: VADER-Sentiment-Analysis. Retrieved June 12, 2020, from https://github.com/cjhutto/vaderSentiment