Don't Take This Personally: Sentiment Analysis for Identification of "Subtweeting" on Twitter

Noah Segal-Gould   Adviser: Sven Anderson

Computer Science Program and
Experimental Humanities Program, Bard College, May 2018

Abstract

The purpose of this project is to identify subtweets. The Oxford English Dictionary defines "subtweet" as a "[Twitter post] that refers to a particular user without directly mentioning them, typically as a form of furtive mockery or criticism." This paper details a process for gathering a labeled ground truth dataset, training a classifier, and creating a Twitter bot which interacts with subtweets in real time. The Naive Bayes classifier trained in this project classifies tweets as subtweets and non-subtweets with an average F_1 score of 72%.

Background and Motivation

On Twitter, the most common way to publicly communicate with another user is to compose a tweet and place an "@" before the username of that user somewhere in the tweet. Through this method, public discussions on Twitter maintain a kind of accountability: even if one were to miss the notification that they were mentioned in a tweet, one's own dashboard keeps a running list of their most recent mentions. Figure 1 provides examples of the mentions shown on one's own Twitter dashboard.

Figure 1: Mentions Dashboard of User @NoahSegalGould

If an individual sought to disparage or mock another, they could certainly do so directly. But the targeted user would probably notice, and through the search functions of the platform, anyone could see who has mentioned either their own or another's username. Instead, a phenomenon known as "subtweeting" persists in which users of the platform deliberately insult others in a vague manner by making complaints while omitting the targets of those complaints.

The Ground Truth Dataset

This project utilizes a probabilistic **Naive Bayes** classifier to identify subtweets. To train the classifier, we developed a novel approach for acquisition of known subtweets and non-subtweets. This approach for creating a ground truth dataset relies on a particular phenomenon in which Twitter users callout the subtweets of their peers. Figure 2 illustrates the pattern which was observed: a user posts a subtweet which is easily recognized by a peer, and that peer then replies to that tweet in order to complain that the original user was subtweeting or to ask if the tweet was indeed a subtweet.

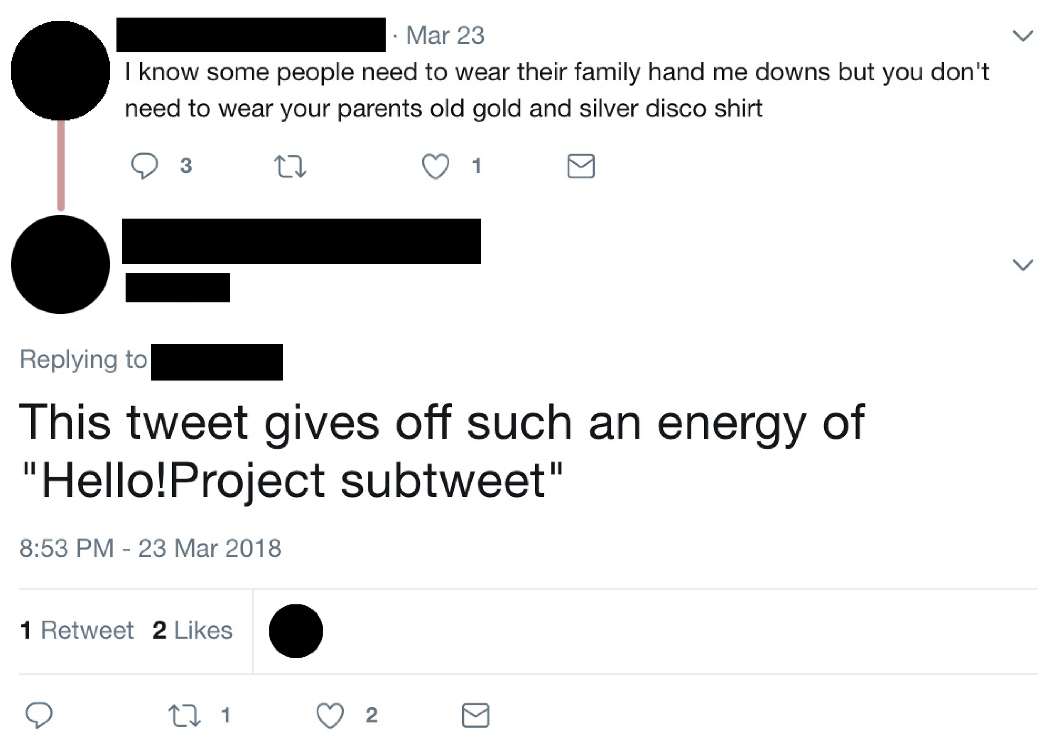

Figure 2: Example of a Subtweet in the Ground Truth Dataset

Training & Testing the Classifier

The **Naive Bayes** classification algorithm makes use of **Bayes Rule** to predict the likelihood that a particular set of features (e.g. words) belong to a particular class [1]. It makes the **naive** assumption that the features are conditionally independent: the presence or omission of a particular feature does not change the likelihood of encountering other features within that class. For each feature f in a document and each class ω in a corpus,

$$P(\omega|f) = \frac{P(f|\omega)P(\omega)}{P(f)}$$

Naive Bayes computes the product of all the predicted probabilities for each feature in the document. The greatest product computed by using **Bayes Rule** on each of the classes in the corpus becomes the predicted class for that document.

	Precision	Recall	F_1 Score
non-subtweets	0.7357	0.6988	0.7166
subtweets	0.7132	0.7490	0.7305

Table 1: Statistics on Performance of the Classifier

In this project, we used k-folds cross-validation to split apart training and testing sections of the ground truth dataset [2]. After the classifier was trained using the training sections, we measured the performance of the classifier on the testing sections in terms of F_1 score, which serves as a weighted average of the precision and recall. Table 1 illustrates these scores for subtweets and non-subtweets.

Each individual testing section was composed of only 10% of the tweets in the entire dataset. By keeping track of the classifier's predictions on this testing section, we accumulated a **confusion matrix** of **true positives**, **true negatives**, **false positives**, and **false negatives** for all 10 folds. Figure 3 illustrates these outcomes in terms of raw counts and normalized over the entire ground truth dataset.

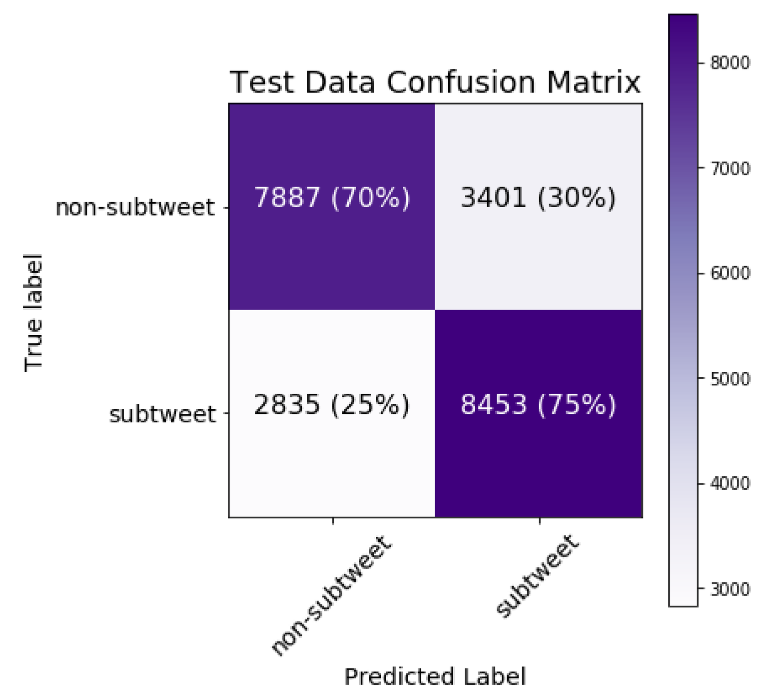

Figure 3: Confusion Matrix on Test Data from Each Fold on the Ground Truth Dataset

The Twitter Bot

After training and testing our classifier, we utilized it in creation of a Twitter bot which interacts with subtweets in real time. Figure 4 shows (censored) examples of these interactions.

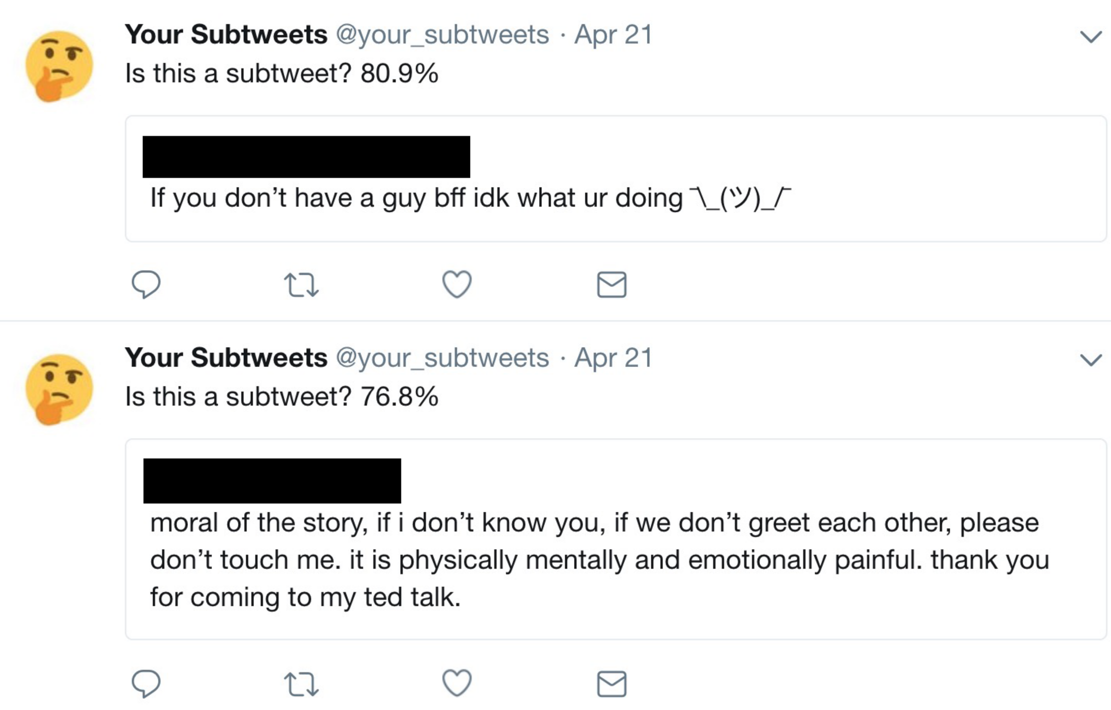

Figure 4: Example of the Twitter Bot Quoting Users' Tweets

References

[1] Harry Zhang, *The optimality of naive Bayes*, AA **1** (2004), no. 2, 3.

[2] Tomas Borovicka, Marcel Jirina Jr, Pavel Kordik, and Marcel Jirina, *Selecting representative data sets* (2012).