Sentiment Analysis for Identification of "Subtweeting" on Twitter

A Senior Project submitted to
The Division of Science, Mathematics, and Computing
of
Bard College

by
Noah Segal-Gould

Annandale-on-Hudson, New York
December, 2017

Abstract

The Oxford English Dictionary states that the definition of "subtweet" is "(on the social media application Twitter) a post that refers to a particular user without directly mentioning them, typically as a form of furtive mockery or criticism." In this project, I employ methods of sentiment analysis and natural language processing to identify subtweets.

Contents

Dedication

I dedicate this senior project to @jack, who has willfully made numerous changes to Twitter which inevitably angered millions.

Acknowledgments

x

1
Introduction

1.1 Background

The news and social networking service known as Twitter had over 140 million active users who sent 340 million text-based Tweets to the platform every day by March 2012 [1]. Since Twitter-founder Jack Dorsey sent the first Tweet in March 2006 [2] social scientists, advertisers, and researchers of natural language processing have applied machine learning techniques to understand the patterns and structures of the conversations held on the platform. One such technique is sentiment analysis, which seeks to ascertain the expressed opinion of bodies of text. Sentiment analysis techniques are often treated as classification problems which seek to analyze text and place it into categories such as **positive**, **negative**, and **neutral**. Classification of this kind is useful for companies which seek engagement with their products among customers [3] on social networking websites like Twitter, for political statisticians who seek engagement with the names of candidates among potential voters [4], and for governments or individuals who seek to monitor and control hate speech on the web [5].

On Twitter, a popular way for registered users to communicate with one another is through mentioning the account name of other users in Tweets which can be seen publicly. Links to accounts are embedded within Tweets which feature an at symbol (@) immediately followed by the name of the account. Users receiving mentions of their account names receive notifications

by default on both the desktop and mobile Twitter clients. In contrast to this culture of mentioning users, Twitter can be used for both hate speech and bullying by omitting the at symbol and username of specific users in order to express unsavory feelings about another individual. Indirect mentions of other users which omit usernames have been colloquially called subtweets and typically avoid others' usernames as a kind of furtive mockery or criticism as defined by the Oxford English Dictionary. Through sentiment analysis and application of natural language processing techniques, it is possible to identify subtweets which harass or denigrate individuals on Twitter.

1.2   Sentiment Analysis & Opinion Mining

Academically, sentiment analysis and opinion mining are used synonymously. Both terms were apparently coined in 2003, respectively in "Sentiment Analysis: Capturing Favorability Using Natural Language Processing"[6] and "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews"[7]. In a commercial context, sentiment analysis and opinion mining are useful tools for anyone looking to sell a product. With the vast social networks available, these methods are becoming widespread. Instead of focus groups, opinion polls, and conduct surveys, sentiment analysis and opinion mining programs are increasingly applied to social networking websites to analyze the sentiments and opinions of users toward topics and products.

1.3   SemEval

For various purposes, there exist numerous sentiment analysis and opinion mining methods. SemEval, which started as Senseval in 1998, holds evaluations in the form of competitions in which several tasks and subtasks are assigned. In 2015, SemEval tasks 10 [8] and 11 [9] focused respectively on sentiment analysis of figurative language on Twitter and sentiment analysis of Tweets in general. In 2016's task 4 [10], the focus was on sentiment analysis on Tweets, and again in 2017's task 4 [11] it was the same with the added challenge of working with Arabic-only

Tweets. The subtasks assigned within a particular task are used as problems to which anyone can send a submission as a solution. For example, SemEval 2016 task 4 had the following subtasks:

- (A) Message polarity classification: given a Tweet, predict whether the Tweet is of **positive**, **negative**, or **neutral** sentiment.

- (B) Tweet classification according to a two-point scale: given a Tweet known to be about a given topic, classify whether the Tweet conveys a **positive** or a **negative** sentiment towards the topic.

- (C) Tweet classification according to a five-point scale: given a Tweet known to be about a given topic, estimate the sentiment conveyed by the Tweet towards the topic on a five-point scale.

- (D) Tweet quantification according to a two-point scale: given a set of Tweets known to be about a given topic, estimate the distribution of the tweets across the **positive** and **negative** classes.

- (E) Tweet quantification according to a five-point scale: given a set of Tweets known to be about a given topic, estimate the distribution of the Tweets across the five classes of a five-point scale.

These subtasks are classification problems and submissions are scored as such.

1.4 Naive Bayes

Naive Bayes classifiers are probabilistic supervised learning models which make the "naive" assumption of independence between pairs of features being classified. Sentiment analysis is popularly performed through Naive Bayes.

1.5 TF-IDF

TF-IDF, or term frequency-inverse document frequency, is a statistical representation of how important a single word is for each document in a collection of documents.

1.6 Precision and Recall

In tasks pertaining to text classification, like sentiment analysis, precision refers to the number of correctly labeled items which were labeled as belonging to the positive class and in fact did belong to that class (true positives) divided by the total number of elements which were labeled as belonging to the positive class including ones which were labeled positively either correctly or incorrectly. Recall, then, refers to the true positives divided by the total number of elements that actually belong to the positive class.

2
Methods

2.1 Scikit Learn

Scikit Learn is a machine learning library written for the Python programming language.

2.2 Distribution & Datasets

For this project, I acquired and combined two datasets:

- Alec Go's Sentiment140 dataset [12]. It features 1,600,000 4-point scale classified (**highly negative** to **highly positive**) Tweets. 800,000 of these Tweets are classified as **positive** and 800,000 of them are classified as **negative**. This dataset was not tagged by humans. Instead, it was based entirely on emoticons present within individual Tweets.

- Ibrahim Naji's dataset [13], which features 1,578,628 Tweets. 790,185 of these Tweets are classified as **positive** and 788,443 of these Tweets are classified as **negative**. It was tagged by humans and on a binary (either **negative** or **positive**) scale.

2.3   Confusion Matrices

A confusion matrix is a table which visualizes the performance of an algorithm. In this case, I implemented a Naive Bayes classifier from Scikit Learn on my dataset and included in my results is a confusion matrix of the performance:

```
null accuracy: 50.08%

accuracy score: 79.77%

model is 29.68% more accurate than null accuracy

-----------------------------------------------------------------------------

Confusion Matrix


Predicted   negative   positive   __all__

Actual

negative      324184      72485    396669

positive       88312     309676    397988

__all__       412496     382161    794657

-----------------------------------------------------------------------------

Classification Report


              precision     recall   F1_score support

Classes

negative       0.785908   0.817266    0.80128  396669

positive       0.810329   0.778104   0.793889  397988

__avg / total__  0.798139   0.797652   0.797579  794657
```

2.4   Training, Testing, and Splitting

This is a process of splitting arrays into random train and test subsets according to Scikit Learn's

`sklearn.model_selection.train_test_split`

Bibliography

[1] Twitter Inc., *Twitter turns six* (2012), available at `https://blog.twitter.com/official/en_us/a/2012/twitter-turns-six.html`.

[2] Jack Dorsey, *inviting coworkers* (2006), available at `https://twitter.com/jack/status/29`.

[3] Forsyth Alexander, *How to use Twitter activity to measure the effectiveness of your marketing* (2016), available at `https://www.ibm.com/blogs/business-analytics/how-to-use-twitter-activity-to-measure-the-effectiveness-of-your-marketing/`.

[4] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan, *A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*, Proceedings of the ACL 2012 System Demonstrations **ACL '12** (2012), 115–120.

[5] Twitter Inc., *Hateful conduct policy*, available at `https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy`.

[6] Tetsuya Nasukawa and Jeonghee Yi, *Sentiment Analysis: Capturing Favorability Using Natural Language Processing*, Proceedings of the 2nd International Conference on Knowledge Capture **K-CAP '03** (2003), 70–77, available at `http://doi.acm.org/10.1145/945645.945658`.

[7] Kushal Dave, Steve Lawrence, and David M. Pennock, *Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews*, Proceedings of the 12th International Conference on World Wide Web **WWW '03** (2003), 519–528, available at `http://doi.acm.org/10.1145/775152.775226`.

[8] Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov, *SemEval-2015 Task 10: Sentiment Analysis in Twitter*, SemEval@NAACL-HLT (2015), 451–463.

[9] Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes, *SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter*, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) (2015), 470–478.

[10] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov, *SemEval-2016 Task 4: Sentiment Analysis in Twitter*, SemEval@ NAACL-HLT (2016), 1–18.

[11] Sara Rosenthal, Noura Farra, and Preslav Nakov, *SemEval-2017 Task 4: Sentiment Analysis in Twitter*, Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (2017), 502–518.

[12] Alec Go, Richa Bhayani, and Lei Huang, *Twitter Sentiment Classification using Distant Supervision*, CS224N Project Report, Stanford (2009), 12.

[13] Ibrahim Naji, *Twitter Sentiment Analysis Training Corpus (Dataset)* (2013), available at `http://thinknook.com/twitter-sentiment-analysis-training-corpus-dataset-2012-09-22/`.