Tracking Sentiment Analysis through Twitter

Thomas Carpenter and Thomas Way
Applied Computing Technology Laboratory
Department of Computing Sciences
Villanova University, Villanova PA 19085
thomas.way@villanova.edu

Abstract – *Social media continues to gain increased presence and importance in society. Public and private opinions about a wide variety of subjects are expressed and spread continually via numerous social media, with Twitter being among the most timely. The ability to quantify and evaluate society's perceptions is increasingly critical to the success in the marketplace. Sentiment analysis is an approach that can be used to computationally measure perceptions regarding topics based on selected, textual source material. This paper reports on the design of a sentiment analysis tracking algorithm and its implementation in the form of a web-based application that can quantify sentiment contained in Twitter feeds and track the sentiment as it changes over time. Results are provided that evaluate the tool for use in performing tracking of sentiment analysis as perceptions change over time.*

Keywords: Sentiment Analysis, Twitter, Opinion Mining, Sentiment Tracking.

1   Introduction

People have always been interested in what others think. It is human nature to seek the opinions of others, and with the availability of vast online resources for expressing opinion, it is increasingly feasible to automate the process of discerning widespread opinion or sentiment on any given topic, idea, product or person. With the dramatic growth of social media in the past 10 years, an aggregated study of social media statistics that may be beneficial to business found that over 80% of active online users make use at least one form of social networking and nearly 23% of time spent by those users online is spent on social networking sites [6]. More specifically, the study found that 51% of Facebook users and 64% of Twitter users are more likely to buy from the product brands they follow.

It is well known that social media sites such as Twitter and Facebook are frequently used to express, or post, opinions about a particular topic of interest to the poster. These opinions can be a rich source of feedback to the marketplace if they can be gathered and analyzed in a timely and meaningful way. The field of sentiment analysis, which includes opinion mining, provides a variety of approaches that can attempt to manage and make sense of this large and widely distributed resource of opinion. [9]

While sentiment analysis of one form or another has been investigated as early as 1979 [1], research in the area has increased dramatically since 2001. Since that year, with concurrent improvements to machine learning and NLP techniques, the widespread use of the Internet, the ease of publishing material online, and the recognition that understanding trends in public opinion expressed online can be a valuable source of information to business and researchers [9].
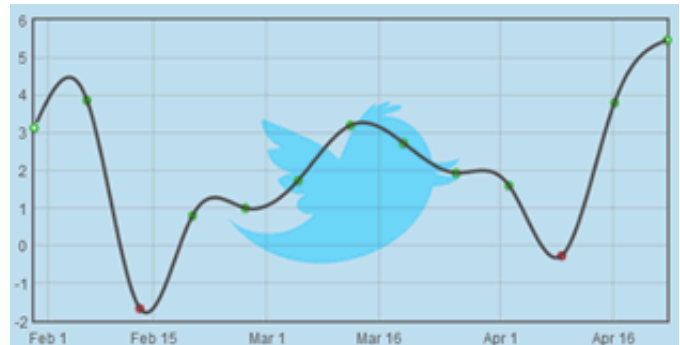

Figure 1. Example sentiment tracking graph.

In this paper, the design of an algorithm that performs sentiment analysis over time using data from Twitter is introduced and a web-based application called the Villanova Analytical Sentiment Tracker (VAST) that implements the algorithm and generates sentiment tracking graphs (Figure 1) is described. Results of preliminary tests of this sentiment tracking approach are discussed, and plans for future development of an online sentiment analysis research tool are presented.

2 Quantifying Sentiment

The challenge of performing sentiment analysis tracking on a large set of source material, such as that available from online social media and over a given period of time, can be broken down into two complementary tasks: Sentiment analysis and sentiment tracking.

2.1 Sentiment Analysis

Sentiment analysis is an active area of research that makes use of natural language processing techniques to quantify an expressed opinion or sentiment within a selection of text. [8] Common uses of sentiment analysis include managing and analyzing review-related online sources, computationally verifying partial content as it relates to overall content in a source material, assessing marketplace reactions to business practices and products, measuring public perceptions of political figures and celebrities, and a wide range of diverse uses [9]. There are a wide variety of algorithms used for performing sentiment analysis, from simple word occurrence counting to machine learning [9], with the selection of one or more approaches typically dependent on characteristics of the material being analyzed [9].

The advent and widespread use of social media, with its emphasis on expression of individual opinion, has provided a rich source of material for large-scale sentiment analysis [2]. Social media has increasingly become a tool for spreading interest in a product, person or idea, and for general promotional use within the online marketplace [5]. If there are negative blog posts or tweets about a product or business, there is a good chance that the business will notice a decrease in their sales. Key among the reasons why this is the case is because those who tend to use online resources frequently to gather and post opinions tend to place significant weight in the information they gather [3]. Thus, it is in the interest of the online marketplace in all its forms to pay attention to the opinions expressed online via social media.

Although sentiment analysis is prone to the same difficulties that general natural language understanding approaches are, the problems are well-understood and solutions to manage these challenges and provide statistically meaningful results exist [2,9]. The use of Twitter as a time-stamped data source has been identified as a realistic and substantive resource [7], as microblogging in general has found a place in online brand marketing [4].

2.2 Sentiment Tracking

Periodically recording the results of Sentiment Analysis over time provides a mechanism for assessing opinions about a topic of interest and quantifying the impact of outside forces, such as current events, popular trends and marketplace competition. Sentiment Tracking gathers the output of Sentiment Analysis at a desired time granularity and over a desired time duration to provide time-based sentiment data for a desired topic.

This longitudinal data can then be analyzed for simple trends or can be compared against other external events. For example, performing daily sentiment analysis on a prominent politician over a period of one year can be juxtaposed with the activities of that politician. This juxtaposition can be analyzed to attempt to quantify the impact of those activities or other current events on the public perceptions expressed via social media.

2.3 Applicability of Sentiment Tracking

An approach such as sentiment analysis and tracking can be appealing to a wide variety of audiences. The general public may be interested in checking what people are saying about their favorite pop star. Celebrities, and the people who hire them and pay them, can see what a large group of people think of them. Politicians may want to know what people are saying about them during an election. Sports analysts want to know what percentage of people are supporting a certain team during a game, which is an opinion that can fluctuate even while the game is happening. Businesses may desire to know if people are upset with them during the latest fall in stock prices. These factors and many others provide strong motivation for the implementation of a flexible, easy to use, sentiment tracking tool.

3 Sentiment Tracking Tool Design

The tool we have designed to perform sentiment tracking using Twitter is called the Villanova Analytical Sentiment Tracker (VAST) tool. The VAST tool is composed of two parts: tracking and evaluating. The tracking portion of the tool provides the ability to quantify the trend of "positive" or "negative" tweets over a specified period of time. The evaluation part of the tool involves the performance of sentiment analysis to attempt to discern a negative or positive opinion in each tweet.

3.1 Motivation for Tool Design

The purpose of developing the VAST tool was to create the ability to track any specified topic on Twitter over any given period of time within the lifespan of Twitter. While there are other tools easily found with a Google search that perform Twitter analysis in real-time, that was not the intent of the design of the VAST tool. Rather, we have aimed to develop a platform for research and further development of sentiment analysis algorithms and sentiment tracking approaches that would be useful to researchers in many fields.

3.2 Description

The VAST tool allows a user to enter a search term, a start and end date, an interval period of time, the number of tweets for each period, and a threshold accuracy level (Figure 2). The start and end dates refer to the overall time period to track the search term. The interval period refers to how often the user wishes to return results within the two dates (daily, every 2 days, weekly, bimonthly, monthly, or yearly). The number of tweets variable refers to the number of tweets, positive or negative (neutral should not be returned, if possible), that the user wishes to return for each time interval.
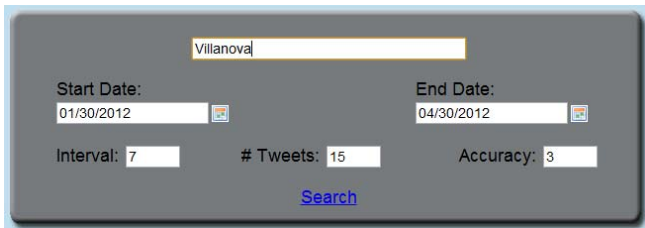

Figure 2. VAST tool control panel interface.

For example, if a user inputs a weekly time interval with 15 tweets per week, the system will return the 15 tweets, classified as either positive or negative. These tweets will be what are determined as the most popular tweets (most frequently re-tweeted or other measurement of popularity) and with the highest sentiment values, for each week in the overall date interval.

Finally, the accuracy parameter refers to the threshold the system uses for determining if a tweet is positive, negative, or neutral. If the accuracy is set to 3, for instance, then the system will only return tweets that have a sentiment value outside the range from -3 to +3. The higher the accuracy is set, the longer the system will take to find the desired number of tweets with that accuracy level. In preliminary test runs, it was determined that an accuracy level of 3 produces acceptable results in terms of speed, quantity of output and quality of output (as determined by qualitative, manual analysis).

Because some sentiment tracking results can take quite some time to produce (minutes to hours), the VAST tool is designed to display results dynamically. As the result of analysis is completed on each time interval, the results are immediately displayed. The user of the VAST tool thus is able to visualize and evaluate the results in an ongoing fashion, enabling interactive analysis as more results are displayed or early termination of the results appeared incorrect in some way.

3.3 Classification Approach

Through experimentation with a number of sentiment analysis classification techniques, it became clear that in order for the VAST tool to perform tracking in the way we proposed it was necessary to develop an algorithm specifically for the needs of the tracker (Figure 3, following page). The more common language classification techniques described in the literature and used for sentiment analysis were either too inefficient or did not provide exception handling for situations where our approach required customized analysis. The resulting algorithm makes use of five word lists: ExtraWordList, MainWordList, NegationList, AdvList, IntensifyList.

The ExtraWordList is the first step of the algorithm. It holds a combination of two types of words or phrases. Its first function is to hold special cases involving the search term inputted to check for in the tweets. For example: [SearchTerm + " beat"] should actually be a positive sentiment because the tweet is most likely referring to a team or politician winning some sort of competition.

Table 1. Sample phrases and their weighted score from ExtraWordList.

Phrase	Score	Phrase	Score
Go + SearchTerm	+4	Don't care	-2
SearchTerm + Beat	+2	No thanks	-5
Beat + SearchTerm	+2	Pretty much	+1
Killing	-3	Does stuff right	+2

```
FOR every Tweet in the List returned from API:

    FOR each word in the Tweet:
      IF the word is in the NegationList
        NegNum = -1
      ELSE IF the word is in the AdvList:
        advNum = 1
      ELSE:
        IF the word is in the MainWordList:
          IF Word value is Positive:
            PosCount = PosCount *negNum+advNum
          ELSE:
            NegCount = NegCount *negNum-advNum
          IF NegNum == -1:
            Append "NOT "+ Word to SentimentWordsList
          ELSE IF AdvNum == 1:
            Append "Adv  "+ Word to SentimentWordsList
          ELSE:
            Append Word to SentimentWordsList
        #Reset Variables
        NegNum = 1
        AdvNum = 0

     FOR word in IntensifyList:
       IF word is in Tweet:
         IF tweet is Positive
           PosCount = PosCount + 1
         IF tweet is Negative
           NegCount = NegCount - 1

    # Add as Positive tweet
    if posCount+negCount  >  accuracy:
           positive ++
            Append Tweet and Tweet Data to Tweet Array
    # Add as Negative tweet
    if posCount+negCount  <  -accuracy:
           negative ++
            Append Tweet and Tweet Data to Tweet Array
    # Add as Neutral tweet to separate Neutral Array
    else:
             neutral ++
             Append Tweet and Tweet Data to Neutral Array List

If there aren't Enough Positive and Negative Tweets to return
      Append Highest Sentiment Valued NeutralTweets to Tweets Array
If there are too many tweets in Tweets Array
      Return the first (numTweets) tweets from Tweets Array
```

Figure 3. Pseudo-code algorithm for Sentiment Tracking in VAST tool.

However, when applying the MainWordList (Table 2) later, this tweet may be considered negative because "beat" is considered a negative word. By adding phrases like this, it is possible to even out or offset the value of the tweet to portray the true sentiment to the tweet. The second type of entry in the ExtraWordList is similar to the first, but without the SearchTerm. The MainWordList does not deal with phrases, only single words.

Table 2. Sample phrases and scores from MainWordList.

Phrase	Score	Phrase	Score
Adorable	+3	Sluggish	-2
Adore	+3	Slut	-5
Kill	-3	Smart	+1
Killing	-3	Smarter	+2
Miracle	+3	Smile	+2
Misbehave	-2	Smiled	+2

The ExtraWordList is used to add in phrases that may invert the results of a negative word to actually be positive and vice versa. For example the phrase "pretty much" will return a small positive value since the word pretty is considered positive. Since this really should just be neutral, the phrase "pretty much" was added to the extra words list with the value of -1 in order to offset the +1 value that will be associated with the word "pretty" later on.

The reason that the algorithm is able to use phrases in the ExtraWordList and not the MainWordList is due to the method of finding the words in the tweet. Since ExtraWordList is very small, the system can efficiently loop through each word/phrase in the list and check if the whole phrase is in the tweet. If the system were to do this on the MainWordList it would have to run through the MainWordList of thousands of entries many times, which would significantly hurt performance. Instead, the algorithm loops through each word in the tweet (a relative small value of N), checking for occurances found in the MainWordList using a much more efficient hashed dictionary lookup.

Tweets can include a wide variety of symbols, letter, numbers, and punctuation. Therefore, before performing analysis using the MainWordList, it was necessary to remove all characters that were neither alphabetical nor numerical.

Once the non-alphanumeric characters are stripped out, the algorithm iterates through each word in the tweet and quantifies its sentiment. This sentiment quantification loop make use of three of the lists. First, the algorithm checks to determine if the word is a negation word found in the NegationList (Table 3). If it is, the algorithm will negate the NegNum variable used in its sentiment score calculation so that if the next word in the tweet is positive it will negate it, and vice versa for a negative word value.

Table 3. Sample from NegationList.

Not
Doesn't
Wouldn't
Cannot
Can't

Second, the algorithm checks to determine if the word is in the AdvList (Table 4) and if found the algorithm modifies the AdvNum variable used in its sentiment score calculation from 0 to 1. The effect of this adjustment is that if the next word is found in the MainWordList, its score will be weighted as a little more positive or negative.

Table 4. Sample from AdvList.

Really
Very
Extremely
Clearly
Certainly

Finally the algorithm checks if the word is in the MainWordList and if it is, it performs the follow action:

```
PosCount = PosCount +
    (wordValue * negNum) + advNum
```

Or,

```
NegCount = NegCount +
    (wordValue * negNum) - advNum
```

If there was not a negation word or adverb before the word in question then the extra variables will have no bearing on the result since it would read something like:

```
PosCount = PosCount +
    (wordValue * 1) + 0
```

However, if, for example, there is a negation word it would negate the positive value and read:

```
NegCount = NegCount +
    (wordValue * -1) + 0
```

The final List that is used in the algorithm is the IntensifyList (Table 5). This list consists of words that portray strong emotion, but could be either negative or positive depending on the rest of the tweet.

Table 5. Sample from IntensifyList.

Wow
OMG
Holy Sh#t
Unbelievable
Insane

A commonly occurring example of the use of an intensifier is the word "wow." For example, the following are positive and negative tweets, each using the word "wow" as an intensifier.

An example of a positive tweet:

Wow! I can't believe we won the Super bowl! Go Giants!

An example of a negative tweet:

Wow! You're crazy if you use Google Drive!

These words should still be applied as they certainly declare sentiment, but it is likely best to wait to apply them until the overall sentiment of the tweet is determined, as they serve primarily as amplifiers or intensifiers. Once the algorithm has completed analysis to determine the overall sentiment of an entire tweet, it checks to see if the IntensifyList words are in the tweet. If they are, the algorithm adds +1 to the value if it is positive or -1 to the value if it is negative. In this way, the score of the word corresponds to the sentiment of the rest of the tweet. Often, the use of an intensifier can tip the score to clearly positive or negative for tweets that are borderline neutral.

At this point, sentiment analysis of the tweet is complete and the algorithm performs a final check to determine if the tweet is within the threshold accuracy level. If the sum of the PosCount and NegCount values is greater than the specified threshold, then the tweet is quantified as a positive tweet. If the sum of PosCount NegCount value is less than the threshold, then the tweet is quantified as a negative tweet. If the sum of PosCount and NegCount do not clearly make the tweet positive or negative, the tweet is stored in a list of neutral tweets, which is only used in the event that there are too few positive and negative tweets to meet the desired count for the current interval.

4   Results

Based on preliminary results (see example result sets in Figures 4-7) of running approximately 50 terms through the VAST tool, the accuracy of the sentiment analysis portion of the algorithm for correctly quantifying the sentiment of each tweet is in the range of 65-85%. This accuracy level was surprisingly good, given that our expectations were that accuracy would be low due to the difficulty with interpreting the small size of tweets and the well-known inaccuracies inherent in many sentiment analysis approaches in general. For example, the use of sarcasm in a tweet can turn a seemingly positive sentiment into a clearly negative one:

I love you, Obama! Please take more of my Money!

For a sentiment analysis system to analyze these words, even applying more sophisticated techniques such as using N-grams or machine learning, it will almost always come out positive. Humans still have the edge in sentiment analysis, it seems, at least where sarcasm is concerned.

Result Set 1
Search Term: **Villanova**
Dates: 03/01/2012 – 04/30/2012
Weekly Interval
15 Tweets per interval
Accuracy of **3**
Correct: **76**
Incorrect: **22**
Unsure: **8**
Irrelevant: **9**
76/115 =
66% correct

Figure 4. Result Set 1 from VAST tool.

Result Set 2
Search Term: **Villanova**
Dates: 03/01/2012 – 04/30/2012
Weekly Interval
15 Tweets per interval
Accuracy of **4**
Correct: **97**
Incorrect: **16**
Unsure: **5**
Irrelevant: **8**
97/127 = **76.4% correct**

Figure 5. Result Set 2 from VAST tool.

Result Sets 1 and 2 had identical search options except for the accuracy. Result Set 1 had an accuracy of 3 and returned only a 66% success rate, while example 2 had an accuracy of 4 and returned a 76% success rate. This may be an indicator that the accuracy threshold is a key element to the accuracy of the algorithm, which was the intent of the accuracy threshold. While the use of this threshold slows the algorithm slightly, the improved accuracy appears to be worth the extra time.

Result Set 3
Search Term: Google Drive
Dates: 04/24/2012 – 05/03/2012
Daily Interval
15 Tweets per interval
Accuracy of 3
Correct: 59
Incorrect: 6
Unsure: 8
Irrelevant: 5
59/78 = 75% correct

Figure 6. Result Set 3 from VAST tool.

```
Result Set 4
Search Term: 'Somebody that I used to know'
Dates: 01/30/2012 – 04/03/2012
2 week Interval
15 Tweets per interval
Accuracy of 4
Correct: 87
Incorrect: 6
Unsure: 6
Irrelevant: 0
Neutral: 8
87/107 = 81.3% Correct
```

Figure 7. Result Set 4 from VAST tool.

The data from Result Sets 3 and 4 (Figure 5) demonstrate that the overall algorithm performs accurate-enough sentiment analysis given its simplicity. Further improvements are planned to improve accuracy and efficiency of the algorithm.

5   Conclusions & Future Work

Sentiment Analysis and Sentiment Tracking holds significant promise for analyzing the perceptions of people who use social media. With the pervasiveness of social media, such as Twitter, Facebook, and the growing popularity of blogging in general, these forms of analysis are viable techniques for quantifying the sentiment of the public at large on a given topic. A tool that can efficiently and accurately quantify human emotion is a challenge to implement. The algorithm that is implemented in the VAST tool reported in this paper attempts to perform sentiment tracking, and has achieved reasonably good accuracy in preliminary experiments.

Research in the area of sentiment analysis is active, and we plan to continue to develop the VAST tool and provide it as a research platform. The VAST tool and algorithm it implements can be made more efficient, and we plan to investigate other sentiment algorithms that are especially accurate for shorter source texts such as tweets.

Other planned extensions of the VAST tool include incorporating GeoLocation data to segment sentiment based on geographic region, data export to provide a means for analysis of sentiment tracking results with other software, and making use of Google search APIs to broaden the scope of sentiment tracking beyond Twitter to the Internet at large. Finally, we plan to collaborate with colleagues in other disciplines. For example, as this is an election year, we intend to coordinate sentiment tracking with colleagues in Political Science to determine the post-mortem, and possibly even predictive, applications of sentiment tracking in an election.

6   References

[1]   Jamie Carbonell. Subjective Understanding: Computer Models of Belief Systems. PhD thesis, Yale, 1979.

[2]   Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. Large-scale sentiment analysis for news and blogs. ICWSM, Boulder, Colorado, 2007.

[3]   Paul Hitlin and Lee Rainie. The use of online reputation and rating systems. Pew Internet & American Life Project Memo, October 2004.

[4]   Bernard J. Jansen, Mimi Zhang, Kate Sobel and Abdur Chowdury. Micro-blogging as online word of mouth branding. In CHI EA '09: Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems, New York, 2009.

[5]   Lillian Lee. "I'm sorry Dave, I'm afraid I can't do that": Linguistics, statistics, and natural language processing circda 2001. In Comptuer Science: Reflections on the Field, Reflections from the Field, pages 111-118. The National Academies Press, 2004.

[6]   Phil Mershon. Social Media Stats for Small Businesses. Social Media Examiner, Nov. 8, 2011. Available at: http://www.socialmediaexaminer.com/26-promising-social-media-stats-for-small-businesses.

[7]   Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the seventh international conference on Language Resources and Evaluation (LREC), May, 2010.

[8]   Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79-86, 2002.

[9]   Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1-135, 2008.