Title
Evaluation of Political Sentiment on Twitter

Permalink
https://escholarship.org/uc/item/5j82j35h

Author
Tavabi, Leili Leili

Publication Date
2015

Peer reviewed|Thesis/dissertation
Evaluation of Political Sentiment on Twitter

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Computer Science

by

Leili Leili Tavabi

2015
Abstract of the Thesis

Evaluation of Political Sentiment on Twitter

by

Leili Leili Tavabi

Master of Science in Computer Science

University of California, Los Angeles, 2015

Professor Song-Chun Zhu, Chair

With the increasing level of access to online political discourses, made possible by the social media networks, a systematic analysis of political speech becomes more critical. The available data could be useful in analyzing important controversies and estimating the general public’s opinion in important political events. A deeper insight enables us to track the public’s opinions to detect the shift of sentiment and discuss on potential reasons why the shift has happened.

Twitter has thus far, shown to be the most effective social network for political analysis. Tweets are limited to 140 characters which lead the users to make succinct statements. The simplicity of text along with the relatively high level of accessibility has made Twitter, a popular target for social analysts.

In this study, we have provided a mechanism for online and automatic collection of Twitter data surrounding different people and events. The data is maintained in an organized time based fashion which makes it easy for future references. We have applied sentiment analysis to estimate political popularity and track the sentiment over time. Sentiment scores have been defined for multiple politicians in order to allow for a comparison of popularity.
The thesis of Leili Leili Tavabi is approved.

Junghoo Cho

Francis Steen

Song-Chun Zhu, Committee Chair

University of California, Los Angeles
2015
# Table of Contents

1 Introduction .................................................. 1

2 Data Collection ................................................. 3

3 Sentiment Analysis ............................................. 7

4 Results .......................................................... 9

5 Related Work .................................................... 15

6 Conclusion ....................................................... 17

References ........................................................ 18
List of Figures

4.1 Obama & McCain’s Tweet Volume(2008) . . . . . . . . . . . . . . 9
4.2 Obama & McCain’s Twitter Sentiment(2008) . . . . . . . . . . . 10
4.3 Ratio of Obama’s Tweet Volume to Romney’s(2012) . . . . . . . 11
4.4 Obama vs Romney’s Twitter Sentiment(2012) . . . . . . . . . . . 11
4.5 Obama vs Romney’s Public Polls’ Favorability . . . . . . . . . . . 12
4.6 Obama’s Twitter Sentiment vs Huffington Post . . . . . . . . . . . 13
4.7 Potential Candidates’ Tweet Volume(2015) . . . . . . . . . . . . . 14
4.8 Potential Candidates’ Twitter Sentiment(2015) . . . . . . . . . . . 14
LIST OF TABLES

2.1 General information about the data set . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Professor Song-Chun Zhu and Professor Francis Steen for co-advising my project as a part of a big and ongoing project on social media analysis.

I would like to offer special thanks to Professor John Cho for his valuable feedback on the project.

I would also like to thank Junseok Joo and Weixin Li for their incredible help and support and without whom this project wouldn’t have been possible.
CHAPTER 1

Introduction

People’s opinions change over time. Being able to accurately detect this change enables us to analyze the potential causes. It’s of great importance to know the main reasons for the dynamics behind general public’s opinion. For example whether a change of sentiment is derived by how an event is represented in the mass media or the true actual event. How influential different networks are in the distribution of news, and if some of them are representing biased information. Being able to provide quantitative analysis for these questions, will help gain deeper insight into the system of social information transfer within a society.

As the complement of an ongoing project of news media analysis, we decided to have a study on Twitter, as a representative of the mass audience. We used Twitter’s text to find the crowd’s dominant sentiment. Since 2011 Twitter has added the feature for containing images in tweets using their uniform resource locators(URL) which makes it possible to perform image sentiment analysis in conjunction with text.

By using sentiment analysis on Twitter’s text, we’ll be able to track the public’s opinion over time and measure its correlation with publicly available polls. The public polls are gathered from various sources and using multiple observations.

The goal of this project was to build a platform, which provides an easy way of collecting Twitter’s data and performing basic analysis for people with little or no coding experience. The platform could be set to a start by specifying the subject of interest and the required time span. It will then crawl the data and
maintain the dataset in an organized structure. The images associated with the
tweets will be downloaded simultaneously and everything would be prepared for
different types of basic statistics including twitter volume, the most dominant
sub-topics, emergence of new sub-topics at certain time points, and last but not
least sentiment score of each subject over time.
CHAPTER 2

Data Collection

The initial aspect of this study was the collection of the data set of tweets mentioning various politicians or important political events. The goal was to perform a long-term analysis on influential politicians and political events starting from 2008. Since both historical and real time data was crucial for building a complete data set, different crawling methods needed to be deployed in association with different sources of Twitter data. The Twitter search API only allows for real time search queries to the tweets posted within the past week of the crawl. This accessible set of data consists of only up to 1% of the whole number of tweets and is retrieved using a random sampling policy. This approach was sufficient for the collection of online stream of data, whereas it couldn’t be used for any time prior to one week of the crawl. We adopted this API to crawl the 2016 election data surrounding 22 potential election candidates. We started the collection process from Feb 05, 2015 until present. The crawler will proceed automatically until the election day on November 08, 2016 to retrieve the complete set of available data. Since Twitter Search API is limited in only accessing 7 days worth of tweets, we needed to use a third party provider which allows access to historical data for the past few years. For this purpose we used Topsy, which is a real-time search engine for social posts. Topsy grants fire hose access to all twitter data dating back to 2006. The retrieved tweets from Topsy are ranked using a proprietary social influence algorithm that measures social media authors on how much others support what they are saying. Topsy retrieves a volume of data dependent on the length
of time one wants to devote to the collection process. It provides full access to all the tweets in case one has infinite time for crawling data. The crawler requires the specification of the query as a unique or set of keywords combined by logical operators. The query could be a combination of words, hashtags or statements. The crawler also requires the time span to be stated by the start and end date of the range for which the user wants to crawl. The crawler then starts the data collection using the suitable API for that time span.

Given the focus of the project, which is the Presidential Elections in the United States, we collected the tweets surrounding the election candidates for 2008, 2012 and the potential candidates for 2016. This includes the tweets surrounding Barack Obama and John McCain for the 2008 election, Barack Obama and Mitt Romney for 2012, and finally 22 potential candidates for the 2016 election including but not limited to Hillary Clinton, Joe Biden, Jeb Bush, Ben Carson.\footnote{Joe Biden, Hillary Clinton, Al Franken, John Kerry, Martin O’Malley, Jim Webb, Luis Gutierrez, Tim Kaine, Jeb Bush, Ben Carson, Chris Christie, Lindsey Graham, George Pataki, Rick Perry, Scott Walker, Marco Rubio, Rand Paul, Bobby Jindal, John Kasich, Mike Pence, Rick Santorum, Bernie Sanders} For the 2008 and 2012 elections, I collected the data beginning from each of the respective election years until the end. For the upcoming election in 2016, the collected data set includes the tweets mentioning either of the candidates starting from February 05, 2015. The crawling process will continue automatically until the actual election date in November 08, 2016, therefore the collected data will expand an approximate range of 1.5 year prior to the election date and provide enough insight for various types of analysis. Another featured data set is the collection of tweets mentioning the current president, Barack Obama, from 2008 until 2015, which allows for a long-term analysis of Obama’s popularity and feedback among the Twitter users during his years of presidency.

Each tweet is returned as a JSON object consisting of various fields, including information such as hashtags, image URLs, additional external URLs, the number of times the tweet has been retweeted or favored by other users, replies to the
tweet, date and time of the post. It also includes some information on the author as when he/she has joined twitter, where his/her base location is and etc. The tweet objects have expanded over the years as new fields have been added. Some of these fields are highly useful whereas some are not guaranteed to be reliable. For example retweet and favorite counts could be useful features in measuring how much support one tweet has received. Retweeting is the action of reposting a tweet while referencing its original author and favoring is the same as liking the tweet. These two fields could presumably be used as the importance weights of a single tweet, but they have shown to be unreliable; meaning the reported numbers are not always equal to the original values. Twitter’s policy while retrieving the data, is to prioritize speed over exactness. This sometimes comes at the cost of reporting untrue numbers for the retweet and favorite counts. Therefore in the current available search API, the retweet and favorite counts are sacrificed by the crawling speed, which makes them impossible to use.

Hashtags are highly important in Twitter. They allow easy detection and tracking of important subjects during their lifetime. New hashtags are born everyday while others disappear. They have different lifetimes, as only few of them tend to last for long periods while most of them appear suddenly and disappear shortly after. In order to collect all possible data surrounding a subject, it would be beneficent to be aware of all the existing hashtags within that domain. This is usually very hard to detect manually because it requires a brute-force search among the whole set of data. Hashtags are included as a separate field within the tweets JSON object and would therefore be easy for the crawler to detect new hashtags while crawling the tweets. Therefore the query can be dynamically optimized over the crawling process. This additional feature will help achieve a more complete and inclusive collection.

During the collection process, the data has been fully maintained in an organized structure consisting of the posting date, text and keyword of interest. For
<table>
<thead>
<tr>
<th>Year</th>
<th>Crawling API</th>
<th>Tweet count</th>
<th>Image count</th>
<th>User count</th>
<th>Influential user count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Topsy</td>
<td>1,429,012</td>
<td>0</td>
<td>133,889</td>
<td>5</td>
</tr>
<tr>
<td>2009</td>
<td>Topsy</td>
<td>4,022,465</td>
<td>0</td>
<td>495,030</td>
<td>10</td>
</tr>
<tr>
<td>2010</td>
<td>Topsy</td>
<td>4,148,910</td>
<td>0</td>
<td>562,016</td>
<td>24,574</td>
</tr>
<tr>
<td>2011</td>
<td>Topsy</td>
<td>4,258,305</td>
<td>2,600</td>
<td>651,102</td>
<td>125,909</td>
</tr>
<tr>
<td>2012</td>
<td>Topsy</td>
<td>4,365,743</td>
<td>29,061</td>
<td>446,212</td>
<td>162,717</td>
</tr>
<tr>
<td>2013</td>
<td>Topsy</td>
<td>4,339,913</td>
<td>72,018</td>
<td>654,333</td>
<td>225,207</td>
</tr>
<tr>
<td>2014</td>
<td>Topsy</td>
<td>4,343,228</td>
<td>191,328</td>
<td>630,722</td>
<td>248,710</td>
</tr>
<tr>
<td>2015</td>
<td>Twitter &amp; Topsy</td>
<td>21,479,297</td>
<td>668,264</td>
<td>3,709,976</td>
<td>624,017</td>
</tr>
</tbody>
</table>

Table 2.1: General information about the data set

the tweets including images, images have been downloaded separately for image processing and analysis. Relatively few Twitter users also include geographical information about their posting locations or origins, which makes it possible for analysis on each state’s political stands and biases but challenging due to limited amount of geo-location data. Table 2.1 contains some general information about the collected data set, including the number of tweets and number of users posting the tweets. Influential users are those having more than 1000 followers.
CHAPTER 3

Sentiment Analysis

All words are associated with different types of emotions. Robert Plutchik who is a famous psychologist, proposed the Theory of Emotion which is one of the most influential classification approaches for general emotional responses. He considered there to be eight primary emotions by showing each to be the trigger of behavior with high survival value. These basic emotions are listed in the following: Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust. In the work from [5], Mohammad et. al have considered these 8 emotions while adding two general emotions of Positive and Negative to form 10 different emotion dimensions. They have used these 10 dimensions to annotate a huge data set of about 14,200 word types as a word-level emotion association lexicon where each word could be associated with either none or multiple dimensions. They used Amazon’s Mechanical Turk as their crowd sourcing platform and had each term annotated by 5 different people in order to gain accurate annotation results. Each entry within the annotated dictionary is a mapping between a word and a boolean list of size 10. Each element of the list represents whether not the word is associated with a certain type of emotion. The dictionary consists of 14,200 entries, some of which having the same stemmed equivalent. Therefore along the original dictionary, we maintained a stemmed dictionary in which the sentiment vector of each entry is the average of the sentiment vectors for all the words belonging to the same stemmed equivalence class.
We use both of these dictionaries to find the tweets’ sentiment by splitting each tweet into tokens and removing the stop words. The tokens are then matched against the original dictionary. If there exists a match for the original word in the original dictionary, the corresponding emotion vector is taken as its sentiment. If not, the stemmed dictionary will be checked to find a match for the stemmed version of the word. The number of the words that are found in either dictionary will be saved per tweet. The sentiment score of each tweet is computed by taking the mean of the sentiment scores of the words found in either dictionary. The overall sentiment score for any time span on the subject of interest will then be found by taking the mean of the sentiment scores for tweets within that period. Since the sentiment results could be noisy to some extent, the results are smoothed over long time spans and then compared against publicly available polls.
CHAPTER 4

Results

The collected data was analyzed in different aspects of tweet volume, emotional sentiment and dominant keyword associations. Some interesting results were collected from the analysis.

We compared the tweet volume for the 2008 election candidates, Obama and Romney. The results have been smoothed, but the original charts can also be seen in the background. This chart indicates an initial increase of both Obama and McCain’s tweet volume by the end of Aug, 2008. One reason might be Obama’s big speech on August 28, 2008 in Denver where he took the fight on McCain. The last presidential debate was held on October 15 when the tweets volumes seem to have met. Obama’s tweet volume has then increased to reach its peak on the election day on November 04 and McCain’s has gradually decreased.

Figure 4.1: Obama & McCain’s Tweet Volume(2008)
Figure 5.2 displays Obama and McCain’s popularity over the course of the 2008 election. The popularity has been measured as the difference of positive and negative sentiments retrieved using the NRC lexicon. It can be seen that Obama’s popularity score has always been higher than McCain’s during 2008. (The sentiment scores are reported from a scale of 0 to 1).

The tweet volume analysis on the 2012 election data was more challenging. Due to the increase of Twitter usage and the large amount of available tweets, we were only able to retrieve a subset of the data and therefore plotted the ratio of the tweet volume for Obama to Romney during the course of 2012 (Figure 5.3). This ratio measure is only informative in the sense that it has always been greater that one; meaning Obama’s tweet volume has been higher during the year. The ratio reaches its peak after the election, on December 16 and as the result of Obama’s speech about the Newtown gun shooting.

Figure 5.4 displays the chart tracking the sentiment score of Obama and Romney during 2012. The scores have been computed the same way as for 2008, by taking the difference of the positive and negative sentiments for each candidate.

We compared these results to Huffington Post’s favorability chart for Obama
Figure 4.3: Ratio of Obama’s Tweet Volume to Romney’s (2012)

Figure 4.4: Obama vs Romney’s Twitter Sentiment (2012)
Figure 4.5: Obama vs Romney’s Public Polls’ Favorability

As it can be seen, these two sets of results don’t correlate. A deeper look at what is causing this difference might be helpful in order to improve the current used method.

A reason why the results are not accurate in this case is the NRC tool’s inability of detecting sarcasm. Here’s an example: Romney’s sentiment score from Twitter reaches its peak at March 20 when his initial primary victory in Illinois was marred by his adviser’s remark on the campaign’s policy, comparing it to an etch-a-sketch. This negative remark was widely spread through social networks by sarcastic tweets along side actual positive tweets focusing only on Romney’s successful speech. Since the NRC sentiment tool is not capable of detecting sarcasm, all such tweets were perceived as positive and hence resulted in a positive peak.

Another interesting result was a long term analysis on Obama’s twitter sentiment from 2008 to 2015. The result is shown in Figure 5.6 along with Obama’s
favorability chart from Huffingon Post. Both charts have been smoothed over time. They are showing a lagged correlation over the long time span.

As already mentioned, the candidates for the 2016 election are also being crawled. The tweet volume and sentiment score for only a few of them who are receiving the highest level of attention are shown in Figures 5.7 and 5.8 respectively. It’s shown in Figure 5.7 that Ted Cruz reaches his first peak of attention on March 23 as he becomes the first major candidate to announce presidential bid for 2016. It can be seen from Figure 5.8 that Ted Cruz is seemingly having the highest level of popularity among the twitter users and Hillary Clinton has been having the lowest during this time span.
Figure 4.7: Potential Candidates’ Tweet Volume (2015)

Figure 4.8: Potential Candidates’ Twitter Sentiment (2015)
CHAPTER 5

Related Work

Within the past decade, the interest toward mining sentiment in text has grown rapidly. This is partly due to the large increase of the availability of documents expressing personal opinions. In particular, sentiment in Twitter data has been used for prediction or measurement in a variety of domains, such as political events, social movements or stock market. For example, Tumasjan2010 [9] found the twitter volume to be good predictor on the outcome of the 2009 German election. They have used a text analysis software called LIWC, developed to assess emotional, cognitive, and structural components of text samples using a psychometrically validated internal dictionary. This software calculates the degree to which a text sample contains words belonging to empirically defined psychological categories. The categories include politically important characteristics such as future-orientation, past-orientation, achievement, tentativeness, certainty, etc.

The work of Wang2012 [10] has developed a real-time system for preforming sentiment analysis on the 2012 U.S. presidential election. They provided an interface to track the overall tweet volumes, most positive and negative tweets and the most dominant keywords. They have used a crowd-sourcing approach to do sentiment annotation on in-domain political data in order to generate data for model training. They have then used a nave Bayes model on unigram features as their classifier. Mejova2013 [4] has also done sentiment analysis on the candidates for the republican presidential nominations in 2012. They have proposed a multi-stage approach for political sentiment classification. They have extracted 1,2 and
3-grams as feature vectors and built their classification models using SVMlight.

Bakiwal2013 [1] performed a supervised classification task on tweets to divide them into 3 groups of positive, negative and neutral. Although they removed the tweets that were labeled as sarcastic by the annotators, they achieved a 61.6% accuracy on the classification, which is still superior to naive unsupervised approaches on tweets.

O’Connor2010 [6] have used measures of public opinion extracted from polls and made a connection with sentiment measured from text. They have analyzed surveys on political opinion and found that they correlate to sentiment word frequencies in Twitter. Their results vary across datasets but they have reached a correlation of up to 80% in some cases.
A system of automatic data collection has been developed in this study, also containing basic features for data analysis. The currently used tool for sentiment analysis is accurate to a high extent but also needs to be improved in order to find robustness to sarcastic statements. This is highly challenging due to the limited size of each tweet and its independence from other tweets in the data set. The data collection process will proceed automatically until the 2016 election date and therefore provide sufficient data for analyzing each candidate’s policy and people’s opinion on its regard.
REFERENCES


