ANALYTICS EXPERIENCE 2016

#analyticsx

Clinton vs. Trump 2016: Analyzing and Visualizing Tweets and Sentiments of Donald Trump and Hillary Clinton

Siddharth Grover
Oklahoma State University, Stillwater, Oklahoma

ABSTRACT

The United States 2016 presidential campaign has seen an unprecedented amount of media coverage, numerous presidential candidates, and acrimonious debates over wide-ranging topics from candidates of both the Republican and the Democratic parties. Twitter is the dominant social medium for people to understand, express, relate and support the policies proposed by their favorite political leaders. In this poster, we analyzed the sentiment of the tweets posted by Hillary Clinton and Donald Trump on their Twitter feeds. We also identified the most frequent policy-related keywords used by these candidates, along with the twitter handles they most frequently mentioned in their tweets. We discovered that Donald Trump was more concerned with media coverage, as he frequently tweeted at and mentioned media handles and used social space to negatively talk about the other presidential nominees. Trump also falls short on positive sentiment and had an overall negative Twitter sentiment. Hillary Clinton, on the other hand, used the same space to discuss some of her current policies and events. Though Clinton did not show the same ability to drive the mainstream media narrative on Twitter as Trump, she did generate more overall positive sentiment.

METHODOLOGY

We extracted about 200,000 tweets accessing the live stream API of Twitter, using a java program, mytwitterscraper, which is an open source real-time Twitter scraper. The timeline for the analysis was from April 2016 to June 2016. We concentrated on @realdonaldtrump and @hillaryclinton Twitter handles and also on trending hashtags like #trump2016 and #clinton2016. We also collected information on the number of followers, retweets, and "favourited" tweets for both candidates.

Text mining and sentiment analysis were performed to help focus on the following key points to better make the comparison between both candidates.

- Most used phrases on Twitter
- Which Twitter handles candidates most tweeted at
- Policy key words mentioned on Twitter
- Comparing sentiments on Clinton and Trump's tweets

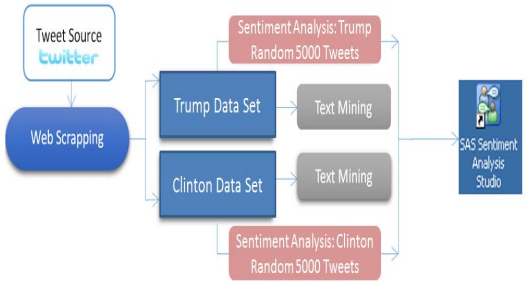

Figure 1. Text Mining and Sentiment Analysis Framework

ANALYSIS

TEXT MINING

Text from 200,000 tweets was extracted and saved as text files using the Text Import node of SAS® Enterprise Miner. Many of the tweets were re-tweets and redundant, and were therefore removed to get a wide variety of topics for better analysis. Text mining was initiated by parsing the data to find tokens (terms), parts of speech tags, entities, etc. We ignored parts of speech, which filters prepositions, determinants, auxiliary verbs, etc., along with numeric values and punctuation, as these contain far less information. Data was then filtered using the Text Filter node and "IDF" was set as the term weight and default setting for the frequency. We built two concept links around official twitter accounts of both candidates to reflect what are they are talking about.

The concept link in Figure 2 shows the association between terms in the tweets Hillary Clinton made. There is a strong relationship among the terms "status", "hillaryclinton", "supporter" and "love", which might imply she tweets about her love for her supporters. Terms like "campaign", "money", and "voter" might imply asking voters to vote or donate money for her campaign.

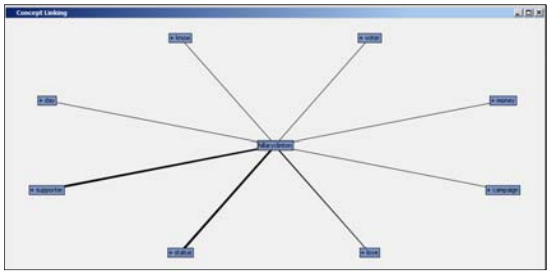

Figure 2. Concept link of @hillaryclinton

The concept link in Figure 3 shows the terms mentioned in the tweets by Donald Trump. There is a strong relationship among terms "foxnews", "cnn", "trump2016", and "hillaryclinton", suggesting that these might be most mentioned terms in Donald's tweets. This might imply that Trump frequently tweets at media handles and also talks about terms like "american", "gun", and "potus."

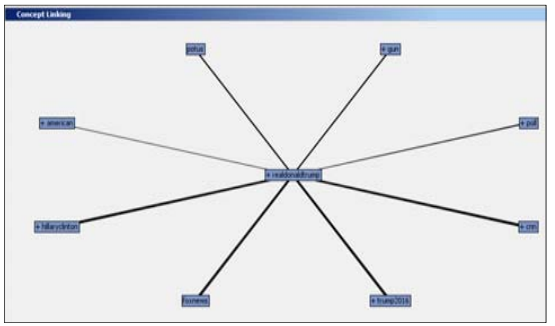

Figure 3. Concept link of @realdonaldtrump

ANALYTICS EXPERIENCE 2016

#analyticsx

Clinton vs. Trump 2016: Analyzing and Visualizing Tweets and Sentiments of Donald Trump and Hillary Clinton

Siddharth Grover
Oklahoma State University, Stillwater, Oklahoma

TEXT MINING RESULTS

ANALYSIS OF MOST RECENT TWEETS

We did an analysis on the most recent original 3000 tweets from both Hillary Clinton and Donald Trump.

Figure 5. Word Cloud of most frequent terms by Trump & Clinton

TRUMP	%	CLINTON	%
@CNN	16%	@POTUS	36%
@FoxNews	14%	@billclinton	20%
@foxandfriends	8%	@realDonaldTrump	16%
@nytimes	7%	@BernieSanders	10%
@JebBush	7%	@HFA	4%

Figure 5. Most frequent twitter handles mentioned

TRUMP	COUNT	CLINTON	COUNT
Terror	133	Guns	250
Immigration	78	Health	150
Jobs	78	Taxes	83
Taxes	53	Immigration	80
Guns	13	Education	53
Education	13	Foreign	53
Health	8	Vets	28

Figure 6. Count of Policy Focused Key Words

METHODOLOGY

SENTIMENT ANALYSIS

We focused on performing sentiment analysis on tweets posted from official Twitter handles of both candidates - (@realdonaldtrump and @hillaryclinton). From the data collected for text mining, we extracted two random samples as modeling data sets with 5,000 tweets each. We further extracted two additional sets of random data sets with 5,000 tweets that would be used to test the results. We used the most recent tweets for data exploration and initial trends before diving into sentiment analytics. We then built different statistical models on those 5,000 tweets in both the modeling data sets using SAS Sentiment Analysis Studio to identify the sentiment distribution for tweets posted by Trump and Clinton. Different rules were specified for positive, negative, and descriptors in accordance with the data collected. Finally, all of the models were tested against the test data to see how they'll hold and predict overall sentiment expressed by both candidates.

SENTIMENT ANALYSIS RESULTS

Per our sentiment analysis, the model built for Trump shows 94% precision for positive sentiment and 90% precision for negative sentiment. The model built for Hillary shows 85% precision for positive sentiment and 95% precision for negative sentiment. Overall model precision for both models is above 90% as shown in Figure 7.

	Model Results		
	Positive Precision	Negative Precision	Overall Precision
Trump	94%	90%	92%
Clinton	85%	95%	90%

Figure 7. Model Comparison

As seen in Figure 8, overall sentiment associated with Hillary is **positive**. Sentiment distribution is at **37% Positive**, **35% Neutral**, and **28% Negative**.

Figure 8. Clinton Sentiment Analysis Output

We can see in Figure 9, overall sentiment associated with Trump is **negative**. Sentiment distribution is at **47% Negative**, **27% Positive**, and **25% Neutral**.

Figure 9. Trump Sentiment Analysis Output

CONCLUSION

In the world of real-time information, Twitter plays an important role in disseminating news and opinions. In order to capture the varying political views of leading presidential candidates, we performed text mining and sentiment analysis on Hillary Clinton and Donald Trump tweets during the time period from April, 2016 – June, 2016. We created concept links to understand association between terms mentioned in the tweets by both candidates. We discovered that Donald Trump frequently tweeted at and mentioned media handles in his tweets and the tweets had an overall negative sentiment. Hillary Clinton, on the other hand, had more policy focused keywords and frequently tweeted more to the political establishment. Clinton fell short on engagement, measured through re-tweets, but generated more overall positive sentiment.

REFERENCES

1. Chakraborty, Goutam, Murali Pagolu and Satish Garla. November 2013. Book Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS® SAS Institute.
2. SAS Institute Inc. 2014. Getting Started with SAS® Text Miner 13.2. Cary, NC: SAS Institute Inc.