Trump VS Clinton: Twitter Edition

Extended Abstract

Author:
John Halloran
Marquette University
john.p.halloran@marquette.edu
https://github.com/johnnyguitar95/DataScience_Marquette

Abstract

Hillary Clinton and Donald Trump were the 2016 Presidential Candidates. Social Media played an affluent role in shaping these candidates from their public image to their stance on policies. Twitter provides an easily accessible medium to look at a candidate's profile, and gather data from their tweets, to explore the popularity of tweets, their most commonly used words and their word associations.

1. Introduction

The election dominated the year 2016 on all media platforms. The election had 24/7 coverage throughout the entire year on television, radio, and the internet. The most interesting aspect of this campaign season was the use of social media. Social media has not seen as much use in the previous elections since platforms like Facebook and Twitter were not necessarily mainstream as they are today. Their use in this campaign appears to be critical to the candidates successes.

The two leading candidates, Donald Trump and Hillary Clinton had different ways of using the popular social media platform known as Twitter. These two candidates had incredibly different personalities and ideologies.

Twitter allows us to peek inside their minds and get glimpses into themselves, but also how they ran their campaign. The words they use greatly reflect their characters. It also is interesting to examine the use of hashtags. Hashtags are a way to create trends on Twitter, when people use a common hashtag, we can get get a list of tweets that use that hashtag; it relates the tweets. Hashtags allow us to track the popularity of the candidates tweets.

2. Experiment

2.1 The Dataset

The datasets of Donald Trump's tweets were provided by Abbas Alidina on Crowdbabble[1] and Hillary Clinton's tweets were provided by Andrew Ginns on GitHub.[2]

Using Python and it's extensive libraries, we can import all the data to and perform extensive cleaning on the data.

Originally the date formats were not in the same format so the first step was to organize the dates to a common format. The date format chosen was MM/DD/YYYY. Trump's was originally YY-MM-DD. Conversion code can be found in TweetCleaning.ipynb.[3] The converting code is as follows

```
tmp = donald_tweets['Date'][count]
    tmp = tmp[3:5] + "/" + tmp[6:] + "/20" +
tmp[0:2]
```

[1] Alidina
[2] Ginns
[3] Halloran

Clinton's tweets follow a similar format in the same ipynb file, but also combining the date and time.

```
tmp = hilary_tweets['created_at'][count]
    index = findyear(tmp)
    date = tmp[0:index]
    time = tmp[index+1:]
    dates.append(date)
    times.append(time)
```

2.2 Word Frequency

In Word_Counts.ipynb[4], Word Frequency counts the number of words most commonly used by the candidates. The words that were gathered followed a couple of restrictions:
1. They were not stopwords, such as *the, a, etc.*
2. The words were at least three characters long
3. The words did not contain @ or #
Rule 3 was ignored when we look at most common mentions, denoted by @, and hashtags, denoted by #.

Tweets were first tokenized, or broken down word by word for each tweet.

```
def tokenize(s):
    return tokens_re.findall(s)

def preprocess(s, lowercase=False):
    tokens = tokenize(s)
    if lowercase:
        tokens = [token if
emoticon_re.search(token) else token.lower()
for token in tokens]
    return tokens
```

A Counter from Collections in Python was used to keep track of words.

2.3 Hashtag Popularity

In Data_Visuals.ipynb, the extent of a hashtag's time of use was examined. By looking over the most popular hashtags over the entire dataset, the times these hashtags were used were gathered and time charts were drawn up.

2.4 Word Co-Occurrences

For CoOccurrences.ipynb[5]In the tweets, words were often used together and can give us some idea as to what the candidates might have associated with each word. A similiar process to Word Frequency was used to tokenize the tweets first.

For this information, a co-occurrence matrix was built as follows,

```
for i in range(len(terms_only)-1):
        for j in range(i+1, len(terms_only)):
            w1, w2 = sorted([terms_only[i],
terms_only[j]])
            if w1 != w2:
                com[w1][w2] += 1
```

This compared all the words used into one singular matrix. Whenever two words appeared together, it's datapoint i,j would increment by 1.

It was also possible to just examine the top 5 words and hashtags and see which words associated the most with them as follows,

```
for search_word in search_terms:
    while count < numRows:
        terms_only = [term for term in
preprocess(donald_tweets['tweet'][count])
                      if term not in stop and
len(term) > 3
                      and not
term.startswith(('#', '@'))]
        if search_word in terms_only:
            count_search.update(terms_only)
        count = count + 1
```

3 Results

It is clear, Donald Trump prefered using Twitter. He has a massive number of tweets and he tweets far more frequently than Hillary did. The main aspect of his tweets was the use of repitition. We look at his most popular words, hashtags, and mentions,

[('Trump', 1008), ('Thank', 754), ('great', 557), ('Hillary', 537), ('people', 377)]

[4] Halloran

[5] Halloran

[('#Trump2016', 592), ('#MakeAmericaGreatAgain', 506), ('#MAGA', 98), ('#DrainTheSwamp', 77), ('#AmericaFirst', 74), ('#VoteTrump', 71)]

[('@FoxNews', 272), ('@CNN', 231), ('@megynkelly', 107), ('@foxandfriends', 100), ('@DanScavino', 90)]

Then looking at his word associations with these words, these are the top five pairings.

[(('Donald', 'Trump'), 250), (('Clinton', 'Hillary'), 209), (('Crooked', 'Hillary'), 203), (('AMERICA', 'MAKE'), 173), (('AGAIN', 'AMERICA'), 165)]

We can see there is frequent pairings of Trump talking Hillary and how she is 'Crooked.' Trump ran his campaign on the promise of making America Great Again. His hashtag use if far more common throughout long periods of time, as evidenced by **Figure 1**. The graphs can also be found in Data_Visuals.ipynb.

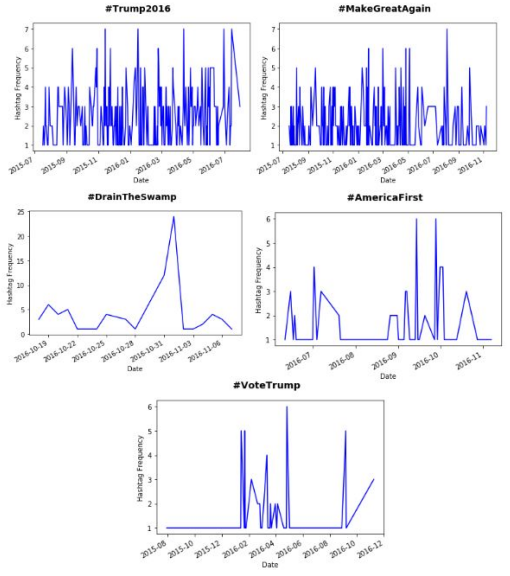

Figure 1

Trump's hashtag usage is far more frequent and more enduring than Hillary's hashtag.

Hillary does not use trending words or words frequently enough compared to Trump. Here are her most frequently used words, hashtags, and mentions.

[('president', 253), ('America', 208), ('vote', 200), ('make', 197), ('people', 180)]

[('#DebateNight', 54), ('#DemsInPhilly', 48), ('#VPDebate', 46), ('#RNCinCLE', 36), ('#debate', 29)]

[('@POTUS', 176), ('@HillaryClinton', 107), ('@FLOTUS', 101), ('@timkaine', 100), ('@TheBriefing2016', 91)]

These are her most frequent pairing of words,

[(('Donald', 'Trump'), 376), (('Clinton', 'Hillary'), 116), (('Hillary', 'Trump'), 104), (('Hillary', 'president'), 81), (('America', 'Trump'), 67)]

Now comparing this to her timing of hashtags, they are less frequent and used for short periods of time in **Figure 2**.

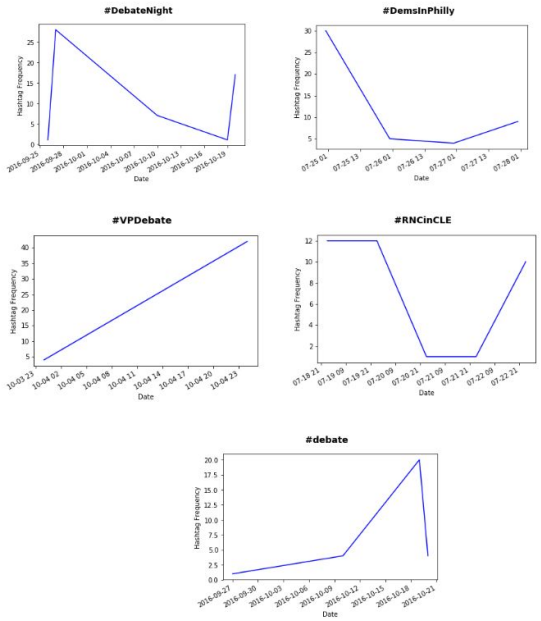

Figure 2

4. Conclusion

From these results, give us a glimpse at Trump and Clinton's approach to social media. Trump uses more repetition to drive his point across. He uses words that have a significant ring to them. The most common of these words being "Great." This probably causes a chest pounding reaction from followers. It makes Trump look confident.

Clinton, to contrast does not take the trending part as seriously. She works within small timeframes, using Twitter as a way to speak out on current events. She does not create trends, she just worries about the convention that is happening tomorrow.

References

1. Alidina, Abbas. *Donald Trump's 11 Best Tweets (and 7,000 Losers)*. Retrieved February 14, 2017 from Crowdbabble: https://www.crowdbabble.com/blog/the-11-best-tweets-of-all-time-by-donald-trump/.
2. Bonzanini, M. *Mining Twitter Data with Python*. Retrieved March 8th, 2017: https://marcobonzanini.com/2015/03/02/mining-twitter-data-with-python-part-1/.
3. Ginns, Andrew. *Idp-twitter*. Retrieved February 14, 2017 from GitHub: https://github.com/andrewginns/idp-twitter/blob/master/HillaryClinton_tweets.csv.
4. Halloran, John. *DataScience_Marquette*. Retrieved March 9, 2017 from GitHub: https://github.com/johnnyguitar95/DataScience_Marquette/tree/master/Notebooks.
5. Kelly, John. *What's With All Trump's Talk About "Draining the Swamp."* Retrieved April 11, 2017 from Slate: http://www.slate.com/blogs/lexicon_valley/2016/10/26/why_do_trump_and_his_supports_keep_talking_about_draining_the_swamp.html