Political Speech in Social Media Streams: YouTube Comments and Twitter Posts

Yelena Mejova, Padmini Srinivasan
Department of Computer Science
University of Iowa, Iowa City, IA 52240
{yelena-mejova,padmini-srinivasan}@uiowa.edu

ABSTRACT

Recently, political sentiment on social media websites has been receiving much attention both in research circles and in the news. However, political sentiment analysis has been largely performed on only a single social media source. It is unclear what outcomes would result if more than one source were used. We present a unique comparison of the textual content of two popular social media – Twitter posts and YouTube comments – over a common set of queries which include politicians, issues, and events. We show Twitter as a stream driven by news and outside sources with 40% share of its content lacking any sentiment, and YouTube as an outlet for opinionated speech. Specifically in YouTube, we find that the author's political stance and the sentiment of the document do not always match, and should be treated separately in analysis of political documents. We also examine the connection between social media sentiment and that of general population by comparing our findings to the Gallup poll, and show that neither discussion volume or sentiment expressed in the two social media were able to predict the republican Presidential nominee frontrunner.

Author Keywords

Social media; Political discourse; Sentiment analysis

ACM Classification Keywords

H.3.3. Information Search and Retrieval:
J.4. Social and Behavioral Sciences

INTRODUCTION

Originally dealing largely with product reviews, social media-driven sentiment analysis (SA) has recently expanded its target to encompass political discourse. However, a limitation of SA research both in the general sphere and in political arena is that the focus is on a single source at a time [3, 4, 5, 9, 14]. For example, Bollen et al. [3] use Twitter to estimate "public mood state". Would the same observations be made if more social forums were taken into the account? Furthermore, when multiple sources *are* compared, such as news and blogs in [8] or reviews and blogs in [1], the topics considered in the sources differ, thus limiting the observations and conclusions that can be made.

In this paper, our goal is to compare two social media sources – Twitter and YouTube. We focus on textual content and thus limit our analysis of YouTube to the comments made on videos. Since our goal is to compare, we focus on a common set of political discussion topics, these include politicians, issues, and political events.

We have several specific goals in this research. The first is to re-examine standard approaches and measures seen in the political discourse literature and test them in our two-social media setting. Discussion volume is one that has been used to estimate political favorability of the crowd [14, 10] and sentiment (polarity) counts is another standard one used for similar purposes [4, 7, 10]. Second, we examine in considerable detail the relationship between political stance (or agreement) taken by the author of the text and the sentiment conveyed. We believe that although the two appear to be similar and have been used for similar purposes [4, 10], there are significant differences that should not be ignored. This analysis is another unique contribution of our work. We also examine several stylistic characteristics of the texts such as the presence of humor, sarcasm, and quotations of outside sources. These are recognized as aspects that complicate the analysis of political discourse [6, 13]. All of these goals are designed to a) understand better the techniques used commonly in analyzing political text and b) understand better the similarities and differences of signals reflected by different social media in the political sphere.

RELATED WORKS

Both Twitter and YouTube have been used widely during various political actions around the world. Thus, from tracking discussions of political debates [4] to predicting election outcomes [14], social media has become a gold mine for political sentiment research. For example, focusing on representation of political figures in Twitter, [12] have developed a way to detect *astroturf* (politically-motivated speech which creates appearance of widespread support for a candidate or opinion). Elections have been studied through the lens of social media: [9] examine the usage patterns of social media by US political parties in the 2010 Midterm Election, whereas [5]

WebSci 2012, June 22–24, 2012, Evanston, Illinois, USA.

look at the conversations surrounding German political parties during the 2009 Federal Elections.

All of these studies focus on only one social media source. Though Lin et al. [8] look at news and blogs to determine whether news sources are biased in favor of covering one political party more than another, they do not compare user-generated social media sources in their analysis. More generally, researchers have recently been addressing sentiment analysis questions that involve multiple sources of data. For instance, Bermingham & Smeaton [1] ask whether it is easier to classify sentiment of short documents like Twitter than longer ones like blogs. Peddinti & Chintalapoodi [11] explore cross-source classifier adaptation from microblogs to reviews, but without topical constraints on the datasets.

DATA COLLECTION

Our dataset consists of Twitter posts and YouTube comments on a set of common topics which are of two types. The first is a politician - issue combination, yielding a total of $13 \times 13 = 169$ combination topis. We also studied 9 event topics. Each topic yielded a query that was executed both on YouTube and on Twitter collecting YouTube comments and Twitter posts for the period of November 16 to 24, 2011.

We implement a two-step approach to collecting YouTube comments. First, using YouTube Search API we collected the top 50 returned videos ranked by relevance. Then, for each video we collected up to 500 most recent comments.

DISCUSSION VOLUME

We first examine the volume of documents – tweets (for Twitter) and comments (for YouTube) – returned for each query. We are interested discussion volume as it has been seen to correlate with popularity [14, 10]. The search described above yielded 27,084 tweets and 40,775 YouTube comments with median of 6.5 tweets and 49 YouTube comments per query. Fifty queries run against Twitter and 22 against YouTube returned zero results (with the intersection of 15 queries). For example, five queries with Elizabeth Warren did not return any results for either source.

First, we examine the volume of topics within the streams. We compute Spearman's rank correlation coefficient between lists of topics in Twitter and YouTube to be 0.566 for politicians, -0.192 for issues, and 0.583 for events. Thus politician and event rankings are more similar across sources than issue lists. We see, for example, that the discussion in the YouTube comments about the three mentioned wars (Iraq, Afghanistan, and Libya) is greater than in Twitter. Also, *Debt Ceiling* topic is much more popular in Twitter than *Tax Reform*, and the opposite is true for YouTube. Because the two topics are related, it could be argued that a mere word choice in the discussions of these issues could result in a divergent results. However, we note that *Immigration* is at the top of the list for Twitter and second from the bottom for YouTube, indicating a drastic difference in the discussion volume in the two streams.

Finally, we compare the Republican Party candidates rankings to those produced by the closest Gallup poll[1], one taken around the same time, Nov 13 - 17. Using Spearman's rank correlation coefficient, we get 0.771 for Twitter and 0.314 for YouTube, showing Twitter to be better at matching Gallup poll ranking. It is interesting to note that this supports the popularity Twitter is gaining for predicting election outcomes. In particular, [14] found that "the mere number of tweets mentioning a political party can be considered a plausible reflection of the vote share and its predictive power even comes close to traditional election polls." However, although fairing better than YouTube, Twitter does not predict the top candidate correctly, as would be a requirement for a successful election predictor.

[1]http://www.gallup.com/poll/election.aspx

CONTENT ANALYSIS

We further analyze the two streams by labeling a subset of the above topics. We label them for sentiment, agreement, writing style, inclusions of links and also explore vocabulary usage.

We choose topics by taking the intersection between the top seven politician, issue, and event lists (ranked by volume) in both streams which have at least 100 comments/tweets in the dataset. Since we have limited resources for annotation, we first identify items most likely to be relevant by using a search engine to rank the documents. For both streams we index the documents (tweets for Twitter and comments for YouTube) using Lemur Indri[2] resulting in an index for each query, for each stream. We then search the index using the query, resulting in a ranked list of documents.

We find that Twitter is a good source of data, with the vast majority of labeled documents being relevant to the queries. YouTube, on the other hand, provides very few documents about both the issue and politician for politician/issue queries, but has a very high accuracy for the event queries. For politician/issue queries, it does capture conversation about either the politician or the issue. For example, over 70% of the documents about topics with Ron Paul are just about the politician and not the issue; same is not the case, for example, for Bachmann queries. Thus, in the sense that only documents relevant to both the issue and the politician being relevant, YouTube gives us a much worse performance – an average of 15% precision (compared to 89% for Twitter). But if we treat discussion about issue only or politician only as also relevant, we get precision of 95% for Twitter and 49% for YouTube. Overall, retrieval of relevant comments for YouTube was a harder task.

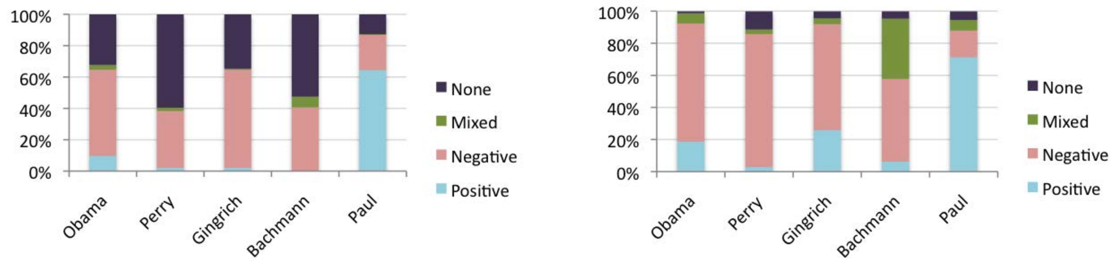
Sentiment

Table 1 shows sentiment expressed by the documents about politicians, issues, or their combination. The two streams differ drastically in the number of documents showing no sentiment (column None). Otherwise, the proportion of positive

[2]http://www.lemurproject.org/

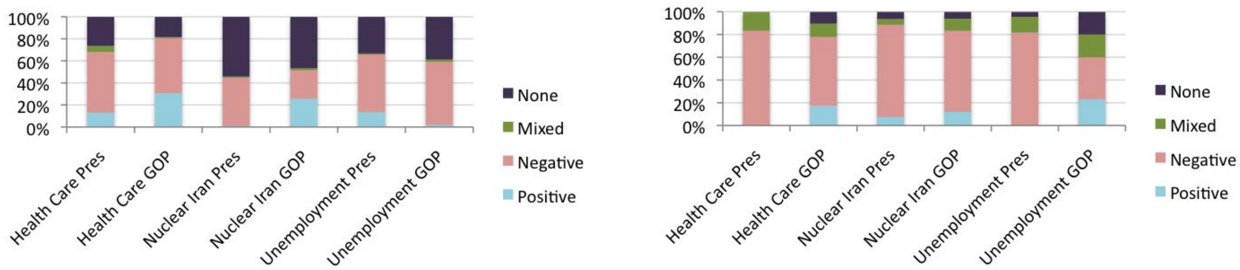
	Pos	Neg	Mix	None
Twitter	17.5	40.6	1.8	40.1
YouTube	27.7	59.0	6.6	6.7

Table 1. Overall sentiment in each stream (percentages)



(a) Twitter (b) YouTube

Figure 1. Sentiment summaries of politicians



(a) Twitter (b) YouTube

Figure 2. Sentiment summaries of issues

to negative sentiment is similar between the two streams, favoring negative about 2 to 1. So there is consistency between the two streams in this regard. As we later show, negative sentiment dominates all discussions, both those about liberal or conservative politicians, and along all of the issues. Thus, it may be the case that the default tone of any political discussion is negative irrespective of medium.

We further examine the sentiment expressed about the issues and politicians by aggregating appropriately. Figure 1 shows the sentiment of documents talking about politicians. The politician getting the most positive sentiment in both streams and by a large margin of difference from the next politician is Ron Paul. This is consistent with the fact that he is known for his active and young base[3]. But there are notable differences between the streams. For example, YouTube shows over 20% positive sentiments about Newt Gingrich, but his support is near zero in Twitter. And, over a third of the YouTube comments about Bachmann express mixed sentiment (compared to 4% for Twitter), showing that in her case the discussion on YouTube can be more complex. Thus the two sources express different sentiment signals for these politicians.

We take a different approach to comparing sentiment across issues. Since the political party may take opposing positions, we divide the data into 2 groups: President (who is considered liberal or centrist) and GOP (who are considered conservative) – all other politicians are in GOP. The summaries are shown in Figure 2.

[3]http://www.huffingtonpost.com/2012/01/12/ron-paul-young-voters_n_1202616.html

Note the overwhelming negative sentiment in YouTube (on average 81% for Pres and 60% for GOP), which is less so in Twitter (on average 50% for issues relative to Pres. and 42% for issues relative to GOP). Furthermore, there is more positive sentiment for GOP side of the issues, except for Unemployment, where in Twitter President gets more positive signals, whereas in YouTube GOP's stance is favored more. This shows that the two media differ in the sentiment signals, and also that some issues may polarize people differently on different social media.

As for discussion volume, we compared the Gallup poll GOP politician rankings to the rankings of our select GOP politicians (ranked using sentiment), and found that neither predicted the frontrunner, and both overestimated the popularity of Mr. Ron Paul. Comparing these ranks to Gallup poll ranking, we find Spearman's rank correlation coefficient -0.199 for Twitter and 0.60 for YouTube. The interesting point here is that with discussion volume Twitter did much better than YouTube (0.771 correlation compared to 0.314), here the performance with sentiment is reversed.

Agreement versus Sentiment

Besides discussion volume and sentiment, another aspect being examined is that of stance taken by the text w.r.t the politician or issue of interest. We use the term 'agreement' for this as it more clearly indicates whether the stance taken by the author of the text agrees or not with the stance of the politician or issue (i.e., the topic). We see for example, sentiment and agreement used for similar purposes [4, 10]. The more positive the sentiments expressed the greater the support inferred and counts of agreement also indicate the overall level of support. Of the two, the problem with sentiment is that it is not enough to identify the sentiment conveyed in the text. It is also important to identify the target of the sentiment and make sure that it is "on topic".

First we look at the number of documents that express sentiment but this sentiment is not on topic. For example, consider the topic of *Newt Gingrich* and the document *"White House is full of liars and old scoundrels, makes me sick! Vote-NEWT12"*. The document conveys negative overall sentiment (given words such as *liars*, *scoundrels*, *sick*) but this is not directed to towards the topic, and hence we regard this sentiment as not being on topic. Had the topic been *the White House* the sentiment would have been on target and negative. We find 56 tweets (7.3% of total) and 176 YouTube comments (24.3% of total) to have sentiment that is off topic. 7% may seem to be a small value however, this indicates that the margin of error is somewhere between 7 and 14% for Twitter when comparing sentiment across two topics. The range is much higher for YouTube.

We now look at the relationship between agreement and sentiment. We define two categories. The first category consists of combinations of agreement and sentiment that are synchronized. These include "agrees + on topic positive sentiment", "disagrees + on topic negative sentiment" and also "neutral stance + on topic neutral sentiment". The second category includes all other combinations of agreement and sentiment.We note that for YouTube only 66% of the documents (480/725) fall into the desired category. The remaining 34% are noisy in this regard. In Twitter, 89% of documents are in the desired category while the remaining 11% are in the noisy category. Thus we see that there is non trivial noise present and again more so in YouTube than in Twitter.

Therefore, the definition of "sentiment" in political discourse should be delineated clearly in order to distinguish between political opinions and emotional states, lest one is misinterpreted as another and inaccurate conclusions are made. These results also indicate that simple lexicon based classification of sentiment is likely to be of limited value in political discourse.

Style

Table 2 shows statistics on various stylistic features of the text. Twitter has many more quotations across all sentiment classes than YouTube. Negative documents have higher chance of being sarcastic, but this is not a very dominant trait in either dataset. Flaming (using inflammatory language) happens more in the Negative ones, though in YouTube it also occurs in other sentiment classes.

	Twitter				YouTube			
	Pos	Neg	Mix	Non	Pos	Neg	Mix	Non
sarcasm	1.2	7.1	0.0	0.3	1.6	5.0	1.6	1.6
humor	0.6	1.5	5.9	0.8	2.0	3.5	6.6	9.7
flaming	0.6	2.3	0.0	0.0	2.7	10.5	3.3	3.2
has quote	55.3	44.4	29.4	50.4	0.0	0.6	3.3	1.6

Table 2. Stylistic features (% of documents in sentiment class)

It has been observed by [13, 2] that people use sarcasm and humor to make their point in ideological arguments, making them more challenging to analyze. However, we show that, even though these are present, they are not dominant in our dataset.

CONCLUSION

In this study we compared YouTube comments and Twitter posts on a set of topics in the domain of politics. Our study indicates several significant differences. The volume of discussion, the amount of sentiment expressed, the nature of agreement vis-à-vis sentiment expressions, all of these show differences across media. Neither medium matches well with Gallup polls. With volume of discussion Twitter appears to have the edge, while with YouTube sentiment does better. A key conclusion is that choice of social medium to analyze determines the results we get. Although we obtain some signals from each that parallel the political world, overall the results obtained across the two media are not consistent. We also studied the relationship between agreement and sentiment and show, for example, that with YouTube we face a greater risk in terms of lack of congruence between the two.

REFERENCES

1. Bermingham, A., and Smeaton, A. Classifying sentiment in microblogs: Is brevity an advantage? *CIKM* (2010).
2. Bifet, A., and Frank, E. Sentiment knowledge discovery in twitter streaming data. *Lecture Notes in Computer Science 6332* (2010).
3. Bollen, J., Pepe, A., and Mao, H. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *WWW* (2010).
4. Diakopoulos, N. A., and Shamma, D. A. Characterizing debate performance via aggregated twitter sentiment. *CHI* (2010).
5. Feller, A., Kuhnert, M., Sprenger, T. O., and Welpe, I. M. Divided they tweet: The network structure of political microbloggers and discussion topics. *ICWSM* (2011).
6. Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., and Konig, A. C. Blews: Using blogs to provide context for news articles. *ICWSM* (2008).
7. Gonzalez-Bailon, S., Banchs, R. E., and Kaltenbrunner, A. Emotional reactions and the pulse of public opinion: Measuring the impact of political events on the sentiment of online discussions. *http://arxiv.org/abs/1009.4019* (2010).
8. Lin, Y.-R., Bagrow, J. P., and Lazer, D. More voices than ever? quantifying media bias in networks. *ICWSM* (2011).
9. Livne, A., Simmons, M., Adar, E., and Adamic, L. The party is over here: Structure and content in the 2010 election. *ICWSM* (2011).
10. O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. From tweets to polls: Linking text sentiment to public opinion time series. *International AAAI Conference on Weblogs and Social Media (ICWSM)* (2010).
11. Peddinti, V. M. K., and Chintalapoodi, P. Domain adaptation in sentiment analysis of twitter. *Analyzing Microtext Workshop, AAAI* (2011).
12. Ratkiewicz, J., Conover, M. D., Meiss, M., Goncalves, B., Flammini, A., and Menczer, F. M. Detecting and tracking political abuse in social media. *ICWSM* (2011).
13. Somasundaran, S., Wilson, T., Wiebe, J., and Stoyanov, V. Qa with attitude: Exploring opinion type analysis for improving question answering in on-line discussions and the news. *ICWSM* (2007).
14. Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. *AAAI* (2010).