2012 Presidential Elections on Twitter - An Analysis of How the US and French Election were Reflected in Tweets

Farhad Nooralahzadeh
Politehnica University of Bucharest
Department of Computer Science
and Engineering
Email: nooralahzadeh@gmail.com

Viswanathan Arunachalam
Politehnica University of Bucharest
Department of Computer Science
and Engineering
Email: vishwa.ar@gmail.com

Costin Chiru
Politehnica University of Bucharest
Department of Computer Science
and Engineering
Email: costin.chiru@cs.pub.ro

***Abstract*—The main objective of this paper is to compare the sentiments that prevailed before and after the presidential elections, held in both US and France in the year 2012. To achieve this objective we extracted the content information from a social medium such as Twitter and used the tweets from electoral candidates and the public users (voters), collected by means of crawling during the course of election. In order to gain useful insights about the US elections, we scored the sentiments for each tweet using different metrics and performed a time series analysis for candidates and different topics (identified by specific keywords). In addition to this, we compared some of our insights obtained from the US election with what we have observed for the French election. This deep dive analysis was done in order to understand the inherent nature of elections and to bring out the influence of social media on elections.**

Keywords: Sentiment Analysis, Time Series, Twitter, Elections, Polarity, Hash Tags, Topics.

I. Introduction

In the recent years there has been a huge growth of popularity in the use of micro-blogging platforms, with a conventional example being Twitter. Social media and micro-blogging have become so powerful nowadays, to the extent of playing a crucial role in social revolutions [1]. In the domain of Data Mining and identifying the public opinion by sentiment analysis, Twitter is in the limelight as a hot research topic.

In developed countries such as USA and France, during the elections, electoral candidates use social media with the help of a communication expert to drive their election campaigns and make political gains. There is considerable evidence that even though Twitter is used for social purposes, it has significant use for information dissemination of various kinds, including personal information (this being its major use). It is therefore reasonable to conduct time series analysis on tweets posted by users, in order to understand the mood that prevails during the course of elections. Our research focuses on applying Natural Language Processing (NLP) and Data Mining techniques, on the recently concluded elections in USA and France held in 2012 and gain some useful insights on the election results.

In addition to time series analysis on the candidates and topics, we decided to do a word cloud and hash tag analysis and compare some of our US election results with that of already concluded French election.

The paper is organized as follows: Section 2 covers related work, while Section 3 describes our dataset and its preparation process. Section 4 presents the various experiments including scoring, frequent word extraction, time series sentiment analysis on US elections and its comparison with French elections. Finally, we conclude with Section 5.

II. Related Work

The rising popularity of online social networking services has attracted research into their characteristics and recent work revealed characteristics beyond crawled data.

T.Khot [2], applied k-means clustering method for corpora consisting of a very large number of documents, reaching the conclusion that when the documents' content is very short (as for tweets analysis), it makes sense to actually cluster the words instead of the documents. Therefore, he presented a method that clusters the words using the word co-occurrence as a similarity measure. He used spectral clustering for creating word clusters and do a "search" to get the actual documents. According to his results, spectral clustering using word co-occurrence is an interesting option for faster clustering over a large set of documents but it may not be faster than other heuristics which are used for clustering that avoid computing the distances between every word pair.

O'Connor et. al [3] have connected measures of public opinion measured from polls with sentiment measured from text. They have analyzed several surveys on consumer confidence and political opinion over the 2008 to 2009 period, and found out that they correlate to sentiment word frequencies in contemporaneous Twitter messages. While their results vary across datasets, in several cases the correlations are as high as 80%, and capture important large-scale trends [3]. The results highlight the potential of text streams as a substitute and supplement for traditional polling.

Agarwal et. al [4] examined sentiment analysis on Twitter data. The contributions of their paper [4] are: introducing POS-specific prior polarity features and explore the use of a tree kernel to obviate the need for tedious feature engineering. The new features (in conjunction with previously proposed features) and the tree kernel perform approximately at the same level, both outperforming the state-of-the-art base line. The

978-0-7695-4980-4/13 $26.00 © 2013 IEEE
DOI 10.1109/CSCS.2013.72

work proposed state-of-the-art unigram model as their baseline and report an overall gain of over 4% for two classification tasks: a binary, positive versus negative classification and a 3-way positive versus negative versus neutral one.

Kouloumpis, Wilson and Moore [5] have investigated the utility of linguistic features for detecting the sentiment of Twitter messages. They have evaluated the usefulness of existing lexical resources as well as features that capture information about the informal and creative language used in microblogging. They took a supervised approach to the problem, exploiting the existing hash-tags from the Twitter data for building their training data. Their experiments on twitter sentiment analysis showed that part-of-speech features may not be useful for sentiment analysis in the microblogging domain [5]. These results suggest that more research is needed to determine whether the POS features are just of poor quality due to the results of the tagger or whether POS features are less useful for sentiment analysis in this domain. The features taken from an existing sentiment lexicon proved to be somewhat useful in conjunction with microblogging features, but the microblogging features (i.e., the presence of intensifiers and positive/negative/neutral emoticons and abbreviations) were clearly the most useful. Using hash tags to collect training data did prove useful, as also did using data collected based on positive and negative emoticons. However, which method produces the better training data and whether the two sources of training data are complementary may depend on the type of features used. Their experiments showed that when microblogging features are included, the benefit of emoticon training data is lessened.

The time series approach can also be used to investigate changes in sentiments over time, either to understand the role of sentiments in an event or changes in popularity over time. One large-scale study compared overall changes in sentiments in tweets over time with external social, political, cultural and economic phenomena, finding a connection between offline events and online sentiments [6]. A number of studies focused on sentiments analysis in relation to elections to assess whether it is possible to predict the outcomes [7]. Whilst it seems logical that sentiments expressed in Twitter would reflect the public mood, there is a problem with Spam, attempts to manipulate Twitter for political goals, and different levels of internet use for people of differing political persuasions that makes this difficult in practice.

III. Twitter Dataset

For the purpose of this study, tweets concerning US 2012 election were crawled in October and November, 2012.The dataset was obtained following the next 3 steps:

1) We extracted all the tweets that contained the candidate names, candidate addresses, election related hash tags and candidates hash tags (by crawling the Twitter using *Twitter API*). We collected the full text, the author, the written date and time of the tweets, and we ended up with 196,000 tweets that we stored in a database.
2) We cleaned each tweet by removing stop words, numbers, html references, punctuation symbols and candidate names and addresses. Furthermore we deleted all the re-tweets from the dataset (in order to remove duplicate tweets). Emoticons were found and replaced by the word *posemo* for positive emoticons and *negemo* for negative emoticons. We lemmatized the words from the tweets using [8] and we also stored the cleaned and lemmatized tweets in the database, along with the original ones.
3) We also extracted the candidate names from each tweet and stored it in other column of the database. If more than one candidate was present in the text of the tweet, we marked it with the value +1 candidate. We also marked the tweets where no candidate was mentioned with the value no candidate, in order to be able to filter these tweets later so that they fit to our needs.

Similarly we obtained the French elections dataset collected during the first phase of the French elections. The dataset contains approximately 10000 tweets, collected in about 10 days during the course of elections. We also applied the above mentioned preprocessing steps to this dataset and stored them in a database.

IV. Experiments and Results

Our experiments mainly focused on time series sentiment analysis, finding frequent words and comparing the French and US elections:

A. Time series Sentiment Analysis

Sentiment analysis, also called opinion mining, is the field of study that analyzes people opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities [9]. In our experiment we analysed the sentiment polarity (positive or negative) towards candidates and keywords. We did not have enough tweets for all the candidates and therefore we restricted our analysis only to Barack Obama and Mitt Romney. Moreover, our dataset comprised of tweets collected in about 6 weeks. In our opinion it is not enough time to perform a complete time series analysis with a typical moving average of 3 days, so we reduced it to an one day ratio. And last but not least, because the tweets were gathered manually by the authors, there are unequal amount of tweets per day. Despite all these drawbacks which were also enhanced by the rapid change of public opinions in *Twitter*, we could arrive at some important conclusions.

1) Scoring Functions: Initially three scores were computed to each tweet using three different scoring functions as described below:

score.polarity: This measure uses a list of positive and negative opinion words or sentiment words for English, built by Bing Liu [10]. The score is computed based on the difference between the number of positive and negative words found in the tweet. The advantage of this measure is that the sentiment/opinion words were manually selected by a human judge.

score.sentiment: This metric uses the R package sentiment to classify the polarity (i.e., positive or negative) using a naive Bayes classifier trained on Janyce

Wiebes subjectivity lexicon [11]. The score is -1, 0 and 1 for tags which are negative, neutral and positive respectively. We believe that this measure has the advantage of being able to evaluate all the encountered words (not only those found in different lists).

score.afinn: Computes the tweets' score based on the AFINN list [12], a list of English words rated for valence with an integer between -5 and +5. The words from this list were manually labeled by Finn Arup Nielsen between 2009-2011. Again, we have a human judge evaluating different words, which we consider to be beneficial for our application.

Measures such as score.polarity and score.afinn have values outside the range [-1, 0, 1], and therefore we converted all positive values to 1 and all negative ones to -1, in order to have all the three scoring functions in the same range of values [-1,0,1] for negative, neutral and positive. Finally we assigned the tag as negative or neutral or positive to a tweet, by choosing the mode between the three scoring functions. If the values obtained for the three measures were all different to one another, the tweet was also tagged as being neutral. We consider that using three different scoring functions, substantially increased the accuracy of the final tag for each tweet, or at least it provided a more confident evaluation than using or trusting only one of the scoring measures that we presented above. For a better understanding and also to justify our solution, we present the two tweets in listings 1 and 2 that were well tagged using the three scores, but would have been wrongfully classified if we only relied on using score.sentiment function.

Listing 1. Example of positive scoring

```
Example Positive Tag Assigned

Tweet =  "@BarackObama Good luck in the debate! I'll
     be watching! I really hope you win this year!!
     :)"
score.polarity = 1
score.affin = 1
score.sentiment = 0
Final Tag = Positive
```

Listing 2. Example of negative scoring

```
Example Negative Tag Assigned

Tweet = "RT @mboyle1: Eric Holder's & Barack Obama's
     #FastandFurious killed hundreds of Mexicans,
     including teenagers. Will Candy Crowley ask
     about that?"

score.polarity = -1
score.affin = -1
score.sentiment = 1

Final Tag = Negative
```

Since the research idea was to measure the tweets polarity towards the candidates (i.e., positive/negative), we discarded the tweets that were tagged as neutral from the further analysis.

The time series sentiment analysis was performed using the below ratio on a per day basis.

$$\frac{\text{Number of Positive Tweets - Number of Negative Tweets}}{\text{Total Number of Tweets}}$$

2) Candidates: As stated before, we reduced the analysis to the candidates *Barack Obama* having a total of 97734 tweets and *Mitt Romney* with 64136 tweets. The sentiment analysis is shown in Figure 1 comparing both candidates.

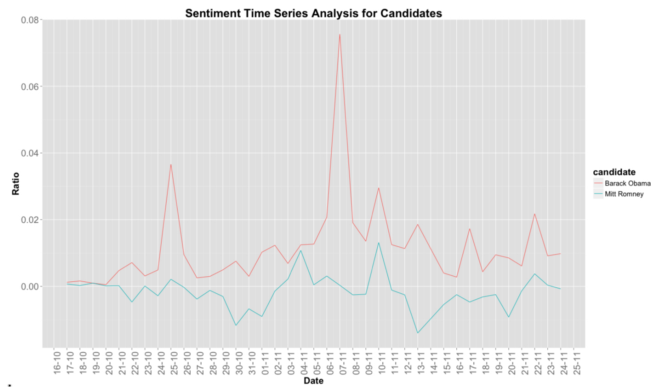

Fig. 1. Sentiment towards Candidates.Red line represents the tweets related to Barrack Obama, while the blue one stands for the Mitt Romney

In general, we see that *Obama* (red line in Figure 1) has better public opinion (at least on *Twitter*) than *Romney* (blue line in Figure 1). The plot describes the number of daily tweets that were positive (above the 0 line) or negative (below it) for each of the candidates. The re-elected president has only positive ratios, while his major opponent has mostly a negative ones. The big peak on November 7th is due to the elections results.

3) Keywords: To do the sentiment analysis we chose the top *abstract* keywords or topics (i.e., non associated to any particular party or candidate) that were also present in every day we gathered the tweets. Topics that were chosen were part of public opinion.

This was done by extracting only the tweets containing a particular pattern. This pattern was chosen from the analysis of the frequent words. For instance, we extracted the tweets containing the pattern *tax*, which was a frequent word in our dataset. By extracting the pattern *tax*, tweets containing the words taxes, #tax, etc were also included in the sentiment analysis.

The keywords selected for analysis were *election, america, job* and *tax* and their time-series sentiment analysis are shown in Figures 2, 3, 4 and 5 respectively. In all the plots we can appreciate the sentiments' volatility (rapidly changing) in such short time periods.

Generally speaking, the term *election* would remain as a positive concept for the American public opinion. This is not surprising, as it is publicly known that the elections is the people's right to choose their representatives in democracies. So even if they have different political ideas, Americans are widely happy having the right of voting.

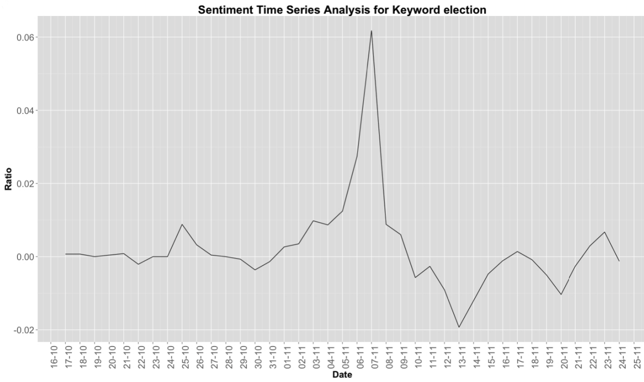

Fig. 2. Sentiment towards keyword 'election'

The positive peak on November 7th is due to the elections' day, when people are mostly congratulating *Obama* for his re-election.

Listing 3. Positive Peak November 7th

```
[1] "#electionday @BarackObama  i'm italian, and i'm
     with you :D good luck"
[2] "#HappyElectionDay tweeps! Hope you voted for #
    Obama! @BarackObama @MichelleObama #TeamObama :D
    "
[3]  "I think @BarackObama will win the elections :)
     #VoteObama #TeamObama #GoObama"
[4] "RT @JJHula: Happy Election Day!!!! Watch my new
     hula hoop-dance video dedicated to Obama!! :D
    @BarackObama #Obama2012 #TeamObama http://t.co/
    LO0ikAOc"
[5] "Left 2 the rest of the world(feedback from FB
    and twitter), @BarackObama would win this
    election hands down,shows how diff Americans are
     :)"
```

The negative peak of November 13th, where people are still reproaching *Romney* his defeat (see listing 4).

Listing 4. Negative Peak November 13th

```
[1] "RT @janiesuds: @mikesacco @MittRomney The
    election is over and you're stupid."
[2] "It was our election to lose, and we lose.
    Something fishy took place #tcot #election #
    romney"
[3] "RT @sherrysamples: FLORIDA YOU HAVE TO BE
    SCREAMING ABOUT THE FRAUD IN UR STATE WITH THIS
    ELECTION!!! Demand recount or REVOTE!
    @MittRomney"
[4] "@MittRomney @PaulRyan are good men! Republicans
     lost election bcuz more people in America want
    something 4 nothing. Capitalism Vs. Socialism"
[5] "now that the election is over mitt romney will
    be as irrelevant as rebecca black"
[6] "@TheNewDeal Tired of hearing how #Obama '
    defined' #Romney in the election. ROMNEY DEFINED
     HIMSELF. Obama just showed him for what he was
    ."
```

This is also an indicator on why *Obama* is better valued (because of his winning) than *Romney* (because of his defeat) as shown in the previous subsection.

The term *America* is positively valued. This is not a big surprise either, it is well-known that Americans really appreciate their country, and even if they have different political

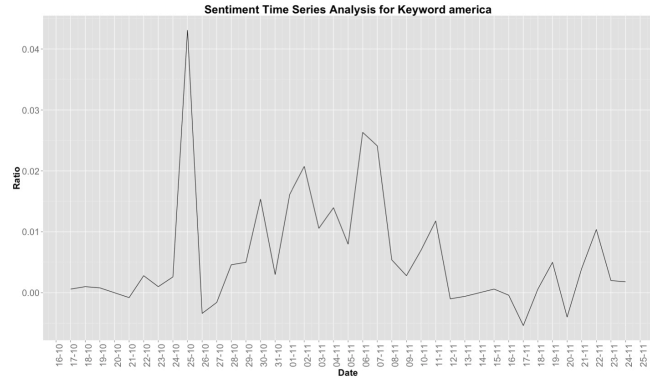

Fig. 3. Sentiment towards keyword 'America'

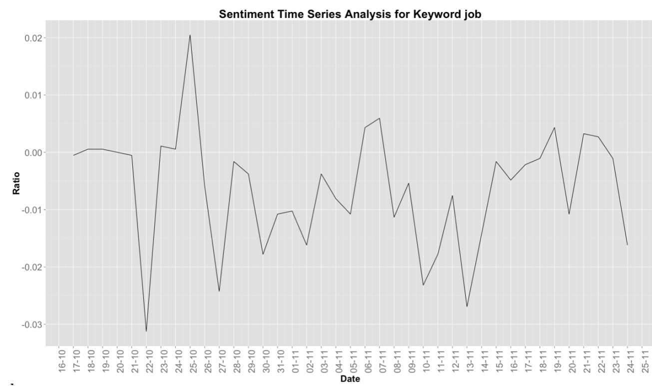

Fig. 4. Sentiment towards keyword 'Job'

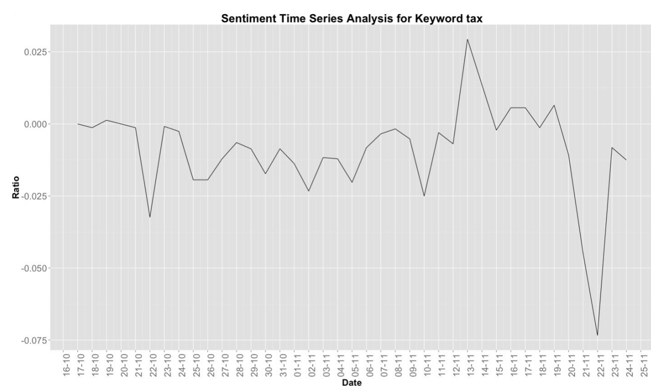

Fig. 5. Sentiment towards keyword 'Tax'

opinions about other issues, they all arrive to a consensus that they want the best for their country, as shown in listing 5.

In opposition to these two patterns, *job* and *tax* show clearly people's concerns. Globally, they are both negatively valued. Today's global crisis has made millions of people all around the world lose their jobs, and this was also happening in US. And who likes to pay taxes?

Listing 5. Peak October 25th

```
[1] "@SinglefiedYue @BarackObama America is proud to
     have you as an American, Yue! :)"
[2] "My mom's not even living in America but she's
    praying for @BarackObama everyday :) Love her so
     much ^^"
[3] "RT @NewYorkPost: ONLY HOPE: For America's
    future, The Post endorses Mitt #Romney for
    president\n  http://t.co/7mEEd5Sq"
[4] "RT @BarackObama: President Obama: 'I've come to
     ask for your help in keeping America moving
    forward.'"
[5] "RT @realDonaldTrump: I am happy to donate $5
    million to a charity Barack Obama chooses.  All
    I am asking is that he is transparent with the
    American people"
[6] "@BarackObama You have my full support Mr.
    President! I still have #hope and I believe in
    you and in #America! I love this country! :D"
```

B. Frequent words

Another main objective of our research was to find frequent words, as it helped us in determining what people were talking about on Twitter during the elections and also helped in ascertaining some important conclusion based on hash tags. Based on the frequent words we performed two different kinds of analysis: Hash Tag Analysis and Word Cloud Analysis.

1) Hash Tag Analysis: Based on the tweets' anatomy, when two hash tags appear together in a tweet,it means that they are related. We extracted the relation between the hash tags by searching our dataset for the pairs of hash tags in a tweet. Every time a pair of hash tags co-occurs, we incremented a counter. Therefore, when two or more hash tags co-occur many times together, it signals that there is a real connection between them. We visualized the relation on a network using Gephi [13] (having the hash-tags as nodes and the edge between them representing the number of times they were found co-occurring in tweets). After that, we discarded the hash tags with candidates' name and hash tags which had fewer degree. The reasons were: 1) candidates' names were the terms which we used to retrieve the tweets, so all of them would have been connected to their names which might have not been useful in our analysis and 2) hash tags with fewer degree would not contribute to community detection in a network. Finally we ended up with 50 hash tags with high centrality degree.

For visualization we applied force-atlas layout in order to align the most connected hubs (hash tags) away from each other and aggregate the nodes connected to these hubs around them. Then, we ranged the size of the nodes by the betweenness centrality in order to emphasize the most influential hash tags in the resulting text graph. Betweenness centrality measure for each node shows how often it appears on the shortest path between any two random nodes in the network. It is an indicator of how important the node is to the overall connectivity of the network and those nodes that connect distinct separated communities together will have a higher measure of betweenness centrality. Finally, we applied modularity algorithm to distinguish the communities in a network. This algorithm scans through all the relations between the nodes, grouping them into communities on the basis of how densely they are connected together. If the nodes are more tightly-knit together than the rest of the network, they are considered to be a part of the same community and these nodes will have same colour. This process is repeated until all distinct communities are identified. Figure 6 shows the resulting graph and the community structure of the hash tags.

Network analysis on hash tags helped us find some important connection between two or more hash tags, that occurred together in a tweet. It helped us detect communities of hash tags. The members of the green color community suggest that they have very strong ties among themselves and that they are against Obama. For example *#benghazi* talks about four Americans soldiers who died in Benghazi-Libya and Obama lied about what happened to them. *#Nobama* also contains tweets against Obama. The red color community is also against Obama but it has weak ties with the green community. On the contrary the yellow colored community is supporting Obama. Moreover the community in light blue is talking about different states, middle east crisis and US governments approaches towards such issues. Pink colored community is neutrally positioned between the communities in red and green.The community in violet color are concerned about the topics *#jobs* and *#economy* which remain central to all other communities.

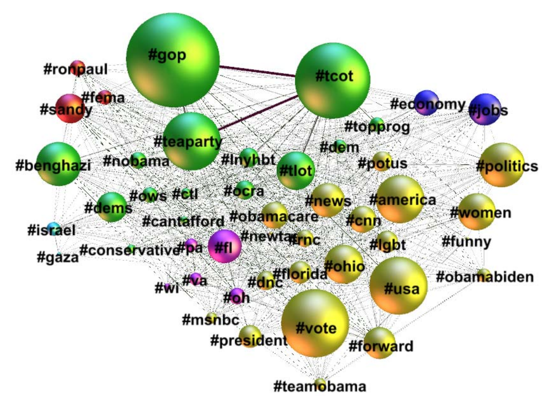

Fig. 6. Hash Tags Network

2) Word Cloud Analysis: In order to do a word cloud analysis, we discarded both the hash tags and the terms used for the retrieving tweets. We tokenized the remaining words (using TreeTagger [8]) and used a bag of words model to determine their frequency. The words having a frequency greater than 100 were considered to be the frequent terms and they had been visualized using the word cloud. The words were colored based on their frequency. From this word cloud, we found out that the most frequent terms people used during the elections were: *President, Vote, America, Job, Tax, Debate, Woman,....* See Figure 7 for the full set of words.

C. Comparison of French and US Presidential elections

In order to compare the two presidential elections, we tried to find the sentiments for each candidates and performed time series and histogram analysis. In addition to this, we also performed a word cloud analysis to see what topics were popular among the candidates and the voters.

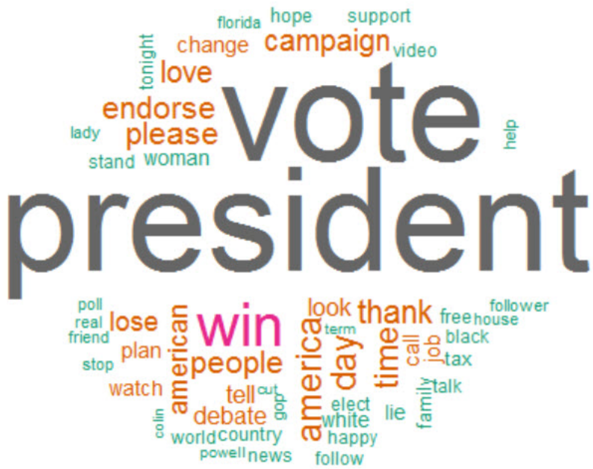

Fig. 7. Wordcloud - US Election 2012

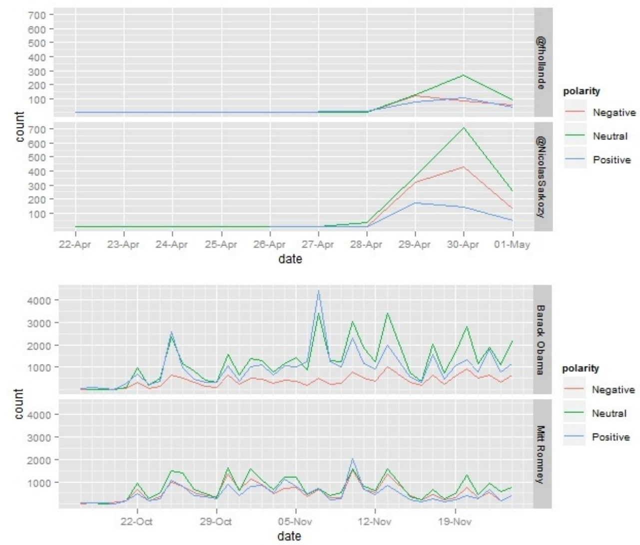

Fig. 8. French and US Presidential - Candidate Sentiments

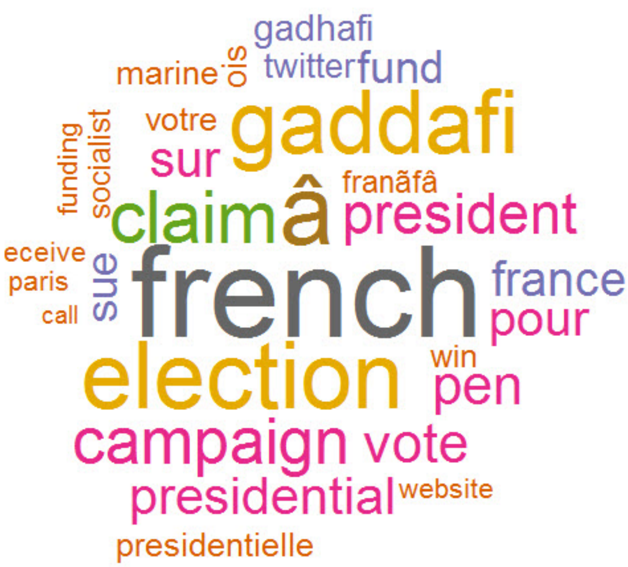

Fig. 9. Wordcloud - French Election 2012

Fig. 10. French and US presidential election - Candidate Sentiments

Due to the difference in the number of tweets collected per day, we normalized it to a weekly average, in order to maintain the consistency of the tweets count. The normalization was applied only to the US-elections dataset, due the inconsistency in tweets' count. Moreover, the French elections dataset was available for only a period of 10 days starting from 22-Apr until 01-May. The time series of French elections suggested that, after the first round of election held on Apr-21/22, only two candidates were significantly competing: *Sarkozy* and *Hollande*. For this reason, we excluded other candidates from the analysis. There was a sudden rise in sentiments due to a speculation by a candidate *Eva joly*, which resulted in positive opinion for *Sarkozy* and less impact for *Hollande*. This trend prevailed for a while but did not last long in favor of *Sarkozy* due to the rise in negative sentiment. Figure 8 shows the time series for both French and US presidential elections.

The word-cloud analysis suggested that the terms the people used most frequently while talking on Twitter in connection with US elections were *President, Vote, America, Job, Tax, Debate, Woman, etc*., while in connection with French elections the tweets were revolving around topics such as *Gaddafi, Receive, Funding, Claim, etc* (Figure 9). There was an allegation that Sarkozy received 42 million pounds funding from Gaddafi, for his election campaign in 2007. This allegation played a decisive role in the case of the French election, leading to Hollande's victory.

The histogram analysis on the sentiments for US elections suggested that *Barack Obama* was ahead of his counterpart Mitt Romney, both in terms of positive and neutral opinions. At the same time, Obama had less negative opinions comparing to Romney. On the contrary the French election data suggested that Nicolas *Sarkozy* was leading against *Francois Hollande* in terms of total sentiments, but *Sarkozy* had more negative sentiments than *Hollande*, which could have led to his decline early in the year. One thing which was common between the French and US election is that only the top two candidates made a mark among the voters, while the remaining candidates were not famous enough or at least less well-known to them. Figure 10 shows the histogram for both French and US presidential elections.

V. Conclusion

In this paper we have presented the whole process involved in time series analysis behind a specific topic and candidates' sentiments from US and French presidential elections held in 2012. We compared the sentiments and keywords for each candidate contesting in French and US elections and found some satisfactory results, which actually reflected the real time results. Based on this, we can ascertain that social media (such as Twitter) works best when candidates hear/see what is being told about them or about national issues and then put themselves in a place to respond accordingly. Even if they cannot solve a problem immediately or address a concern as a whole, the fact that they make their followers (and potential voters) feel as if their voice has been heard is enough to create loyalty and possibly positive results in the elections.

References

[1] http://www.thenational.ae/news/uae-news/facebook-and-twitter-key-to-arab-spring-uprisings-report

[2] T. Khot, *Clustering Twitter Feeds using Word Co-occurrence* (http://pages.cs.wisc.edu/~tushar/projects/cs784.pdf)

[3] B. O'Connor, R. Balasubramanyan, B.R. Routledge,and N.A. Smith,*From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*. In Proc. ICWSM 2010, pp. 122-129, 2010.

[4] A. Agarwal,B. Xie, I. Vovsha,O. Rambow,and R. Passonneau,*Sentiment analysis of Twitter data*, In Proceedings of the Workshop on Languages in Social Media (LSM '11), 30-38. ACL, Stroudsburg, PA, USA, 2011.

[5] E. Kouloumpis, T. Wilson, and J. Moore,*Twitter Sentiment Analysis: The Good the Bad and the OMG.* In Proc. ICWSM 2011, pp. 538-541, Barcelona, Spain, 2011.

[6] J. Bollen, A. Pepe, H. Mao,*Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena.* In Proc. ICWSM 2011, pp. 450-453, Barcelona, Spain, 2011.

[7] J.E. Chung, E. Mustafaraj. *Can Collective Sentiment Expressed on Twitter Predict Political Elections?*. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, 1770-1771, San Francisco, California, USA, AAAI Press, 2011

[8] H. Schmid.*Probabilistic Part-of-Speech Tagging Using Decision Trees.* Proceedings of International Conference on New Methods in Language Processing, pp.44-49, Manchester, UK, 1994.

[9] http://www.cs.uic.edu/liub/FBS/SentimentAnalysis-and-OpinionMining.html

[10] http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

[11] http://people.cs.pitt.edu/~wiebe/

[12] F.Â. Nielsen, *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.* Ed. Matthew Rowe, Milan Stankovic, Aba-Sah Dadzie,and Mariann Hardey, Proceedings, 1st Workshop on *Making Sense of Microposts* (#MSM2011): Big things come in small packages, pages 9398, Heraklion,Crete, Greece, May, 2011.

[13] http://gephi.org/