# Sentiment Analysis of Political Figures across News and Social Media

Quanzhi Li, Armineh Nourbakhsh, Rui Fang, Xiaomo Liu, Sameena Shah
Research and Development
Thomson Reuters
3 Times Square, NYC, NY 10036
{quanzhi.li, armineh.nourbakhsh, rui.fang, xiaomo.liu, sameena.shah}@thomsonreuters.com

## 1. Introduction

In addition to traditional news media, social media platforms have become an important site for political conversations throughout the world. In this work, we use both news articles from all over the world and tweets from Twitter to analyze public sentiment toward political figures. In the experiment, we analyze four 2016 presidential candidates, Hillary Clinton, Donald Trump, Bernie Sanders and Ted Cruz. But the methodologies and results may be applied to other political figures. This work tries to answer the following questions by analyzing millions of news articles and tweets over more than one month period of time:

- Q1. Is any difference between news articles and social media, in terms of their sentiment toward political figures?
- Q2. In social media, when people express sentiment toward a political figure, what aspects of the politician are they talking about or based on, this politician's personality, policies, family, political party, opponents, or others? Is there any difference between positive and negative messages?
- Q3. In political campaign, if one candidate dropped out of the race, will the sentiment toward this candidate and the remaining ones change? Is there any difference between news and social media?

Most previous studies focus on just identifying public opinions on politicians, especially presidential candidates, or tracking their sentiment change over time [2, 6]. There are also studies on predicting election outcome based on public opinion from news or social media [1, 4]. These studies focus either on news media or social media, and they have not done any cross-source analysis, which will be addressed in our question Q1 and Q3. And none of the previous studies has done analysis on Q2.

These questions, especially Q1 and Q2, will help us understand the difference and hopefully also lead us to ask what cause the difference between the two types of media, regarding political figures. News articles are usually written by the professionals and reflect their opinions, and social media messages, on the other hand, are most generated by and reflect the opinions of ordinary people. The methodology used for Q2 can help related partied get insights on why people like or dislike their candidates or political figures. And it will provide valuable information about the politician's strength and weakness, and what aspects they can improve or put more efforts on. We think, overall, the questions, the analysis methods and findings will provide useful information for political campaigns, political analysts, media researchers and other interested parties.

## 2. Methodologies

### 2.1 Data Sets:

We used two data sets in this study, the GDELT data set and about 60 million tweets from Twitter.

**GDELT:** Its global knowledge graph data tries to connect the people, organizations, locations, and other events or facts in the news media each day into a single network, which captures what's happening around the world, its context, and how the world is feeling about it.

The articles we collected from GDELT are from 04/15/2016 to 05/20/2016. There are totally 13 million articles. We extract two types of metadata from these articles:

- People: all the people appearing in an article identified by GDELT
- Positive and negative sentiment scores for each article. They are used as the polarity value toward these people.

**Twitter:** We collected about 60 million tweets related to the four presidential candidates from Twitter's public streaming API, from the same period of time as GDELT data. Each tweet's sentiment toward the candidate is identified by the tweet sentiment analyzer [3, 5]

### 2.2 Methods

**Tweet sentiment identification:** There may be multiple people appearing in a tweet, and they may have different sentiments, e.g. "Hillary is better than Trump". Using the same sentiment identified for this tweet for both people is not appropriate in this case. So we need target dependant sentiment analysis to get more accurate polarity value for each person in the tweet. A target dependant sentiment analyzer based on [3, 5] is used to identify the polarity for each person in a tweet. This tool is based on general word embedding and sentiment specific word embedding [7, 5]. Both of them are distributed word representations, and are learned based on deep learning technology.

Word embedding is a dense, low-dimensional and real-valued vector for a word. The embeddings of a word capture both the syntactic and semantics of the word. Traditional bag-of-words and bag-of-n-grams hardly capture the semantics of words, or the distances between words. The embeddings help learning algorithms achieve better performance in NLP tasks by grouping similar words together. Sentiment specific word embedding extends the general word embedding model by incorporating the sentiment information into the neural network to learn the embeddings; it captures the sentiment information of sentences as well as the syntactic contexts of words.

**Relationship analysis between sentiment types and affecting factors**: To answer Q2, for each sentiment polarity type, we exploit the word embedding model to find the most related factors (personality, policies, political party, etc) that contribute to that polarity. A word embedding model is built based on millions of tweets in that sentiment category, and then the most related factors are discovered by computing word embedding similarity between the target politician and those factors

## 3. Conclusion:

**Our major contributions** include: 1. Cross source sentiment analysis. 2. Target dependant sentiment analysis utilizing word embedding and sentiment specific word embedding, which are cutting-edge deep learning technologies. 3 Relationship analysis between sentiment polarities and different aspects of a politician. The analysis is based on distributed word representation model.

**Results**: Our findings are displayed through visualization, and so due to the space limitation, we will present our findings and detailed methodology at the SBP challenge poster session.

## References:


1. Choy, M., Cheong, L. F. M., Ma, N. L., & Koo, P. S., A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction
2. Namrata Godbole, Manjunath Srinivasaiah and Steven Skiena, Large Scale Sentiment Analysis for News and Blogs, *ICWSM 2007*
3. Quanzhi Li, Sameena Shah, Rui Fang, Armineh Nourbakhsh, Xiaomo Li, Tweet Sentiment Analysis by Incorporating Sentiment-Specific Word Embedding and Weighted Text Features, *ACM/IEEE International Conference on Web Intelligence 2016*
4. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M., Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. 2010
5. Duy-Tin Vo and Yue Zhang, Target-Dependent Twitter Sentiment Classification with Rich Automatic Features, *IJCAI 2015*
6. Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar and Shrikanth Narayanan, A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle, *ACL 2012*
7. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. *NIPS 2013*.