Summer 8-29-2017

A Sentiment-Change-Driven Event Discovery System

Lili Zhang
Kennesaw State University, lzhang18@students.kennesaw.edu

Ying Xie
Kennesaw State University, yxie2@kennesaw.edu

Guoliang Liu
CareerBuilder, gavin.liu@careerbuilder.com

A Sentiment-Change-Driven Event Discovery System

Lili Zhang
Kennesaw State University
lzhang18@students.kennesaw.edu

Ying Xie
Kennesaw State University
yxie2@kennesaw.edu

Guoliang Liu
CareerBuilder
gavin.liu@careerbuilder.com

ABSTRACT

We present a system that automatically discovers important events that have significantly driven people's sentiment changes towards a target using Twitter data (i.e. tweets). This system can also provide the time, importance, and description of events that are associated with people's sentiment changes. In this system, a sentiment classifier is used as the sensor to detect the time points of those changes. It is also used as the filter to effectively eliminate a considerable amount of noisy information and select the most informative tweets to be further analyzed for event descriptions. Discovered events are described from the following aspects, 1) the most important tweets ranked by tweet-based TextRank algorithm, 2) the topics generated by the nonnegative matrix factorization, and 3) the most important keywords generated by word-based TextRank algorithm. Compared with traditional event discovery techniques, the experimental results show that this system can effectively discover important patterns from tweets and unveil 3Ws of an event (i.e. what happens, when it happens, what its effect is), which provides good reference on understanding behavior changes and making strageties. Furthermore, the system was applied to analyze people's sentiment changes towards the two candidates during the 2016 U.S. presidential election. It can also be applied in other scenarios where people's attitude plays an important role like the brand influence marketing and financial investment markets.

CCS CONCEPTS

• **Applied computing → Sociology**; **Document analysis**;

KEYWORDS

Event Discovery, Sentiment Classification, Text Summarization, Topic Modeling

ACM Reference format:
Lili Zhang, Ying Xie, and Guoliang Liu. 2017. A Sentiment-Change-Driven Event Discovery System. In *Proceedings of WI '17, Leipzig, Germany, August 23-26, 2017,* 7 pages.
DOI: 10.1145/3106426.3109038

1 INTRODUCTION

In situations from the marketing campaign to the presidential election, crowds sentiment always plays a crucial role in affecting final consequences. Therefore, it is beneficial for strategy makers to understand the 3Ws related to people's sentiment changes, namely, what happens, when it happens, and what its effect is. In this regard, public opinions need to be collected for the study of sentiment changes along the time. The traditional way is to conduct surveys, but this approach may not be able to produce comprehensive information beyond the predesigned questionaire on a timely manner. Fortunately, the emergence and prevalence of social sites (Twitter, Facebook, Google+, LinkedIn, etc.) provides public platforms for people to actively exchange information and express opinions, which makes a huge amount of data available for study. Among them, Twitter is an ideal platform for deriving insights on people's sentiment changes over time since it offers a micro-blog service platform for people to post what happens and share their attitudes on a timely manner.

Many researches and applications have been conducted to leverage Twitter data (i.e. tweets) from different perspectives including sentiment analysis, event detection, and event summarization. Among these studies, both supervised machine learning and unsupervised machine learning methods are used. We integrate and extend methods from those individual perspectives into a system that is able to automatically detect people's sentiment changes and effectively discover important events. There are four major components in this system, namely, Tweets Sampling, Sentiment Sensor, Sentiment Filter, and Event Discovery. This system was applied to discover significant sentiment change patterns towards Trump and Clinton during the 2016 U.S. presidential election, as well as important events behind those sentiment changes.

Moreover, based on the experimental results of the system performance evaluation, we have two major findings. One is related to tweets data itself. As known, the performance of machine learning models or systems depends on both the data and algorithms. However, for all machine learning methods, a big challenge with tweets is their messy characteristics because users have diverse backgrounds and different word usage practices. These cause lots of noises in the tweets, making it harder to discover informative information from tweets text than formatted documents, e.g. articles. To eliminate the noise, we propose a novel way by utilizing the sentiment classifier as the filter.

The other finding is related to the technique used for event discovery in political context. In general, when performing event discovery, we have two options.

M1. Treat all tweets as a big document with each tweet as a paragraph and then perform text summarization to generate a gist.

M2. Treat all tweets as a document corpus with each tweet as an individual document and then perform topic modeling to generate multiple topics.

Which way can provide more precise information? Through the analysis on tweets during the 2016 U.S. presidential election, we

WI '17, Leipzig, Germany

DOI: 10.1145/3106426.3109038

find that in most cases the summarization generated by M1 is more meaningful than topics generated by M2.

2 RELATED WORK

Many researches have been performed to gain insights from tweets in various aspects like sentiment analysis, event detection, and event summarization. Go et al. proposed a novel sentiment classification method trained by tweets using distant supervision, which achieved high accuracy [3]. Weng et al. built an event detection system with the application of wavelet analysis on words frequency, which was used to analyze discussions on Twitter during Singapore General Election 2011 [10]. Chakrabarti et al. formalized the tweets event summarization problem and provided a solution via Hidden Markov Models [1]. Considering tweets as text data, some general text summarization and topic modeling methods can also be applied and extended to tweets analysis. Mihalcea et al. proposed the TextRank algorithm to select top sentences or words based on the order to represent a document [7]. Non-negative matrix factorization (i.e. NMF) has been applied successfully to extract topics from a document corpus and generated results competing with Latent Dirichlet Allocation (i.e. LDA) in recent years [2].

While these methods are studied individually, some researchers and practitioners combine different techniques together. Wang et al. developed a real-time system of sentiment analysis towards presidential candidates for the 2012 U.S. presidential election and associated the events with public opinions [9]. Their system, however, is mainly focused on the sentiment analysis without events being automatically detected and summarized. Some practitioners performed both sentiment analysis and topic modeling for the 2016 U.S presidential election. Jagtap analyzed 2000 tweets per day for 5 days in total [5]. Kummar worked on tweets from May 2016 to August 2016 [6]. And Stecanella utilized tweets from July 1, 2016 to January 1, 2017 [8].

In our system, we not only perform sentiment analysis and topic modeling, but also quantify the importance of events based on the sentiment changes they have led to. Moreover, the concept of using the sentiment classifier as the filter is introduced, along with comparisons of event discovery techniques, including tweet-based TextRank algorithm, word-based TextRank algorithm, nonnegative matrix factorization. Last but not least, as a use case, we applied the system to comprehensively study tweets from June 2015 to November 2016 during the 2016 U.S. presidential election and generated insights on people's sentiment change behaviors and events behind those sentiment changes.

3 SYSTEM DEVELOPMENT

3.1 Motivation and Architecture Design

With large populations' participation, social media sites have emerged as important platforms for people to post what happens and express opinions. How can we effectively discover a series of events along with their influence from messy and huge amount of information in a long-term process? In order to automatically discover 3Ws related to an event, we developed a sentiment-change-driven event discovery system. The core of the system is that people's sentiment changes are quantified as the importance measurements of the significance of events behind them and then used to detect the occurrence of important events. Furthermore, the event description is derived from the most informative tweets to represent what happens. The system architecture, which is shown in Figure 1, includes four major components.

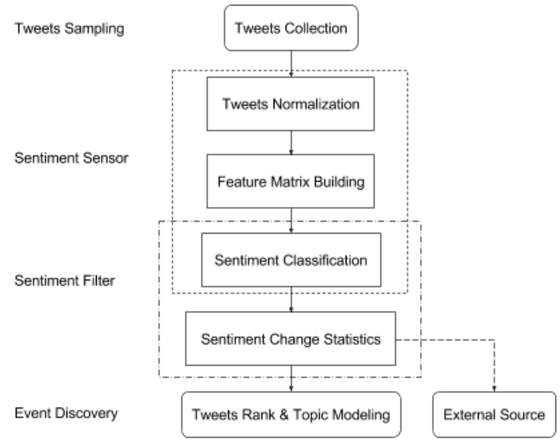

Figure 1: System Architecture

1) Tweets Sampling: Collect tweets sample for the target to satisfy query requirements.
2) Sentiment Sensor: Measure people's daily sentiment change towards the target. It contains tweets normalization, feature matrix building, and sentiment classification.
3) Sentiment Filter: Filter tweets for further event discovery based on the sentiment change direction. If positive ratio decreases or negative ratio increases, we will only analyze tweets labeled with negative sentiment. Otherwise we will only analyze tweets labeled with positive sentiment.
4) Event Discovery: Discover what happens at sentiment change time points through text summarization and topic modeling techniques. Moreover, a problem-dependent module, namely, External Source, can be integrated into system to provide APIs of other related information sources beyond tweets and help eliminate any potential bias of the collected tweets.

This system can be applied to study any particular target like a company, a product, and a person in a wide range of scenarios. The 2016 U.S. presidential election was used as the case study, where the targets were the candidates Trump and Clinton respectively. Figure 2 shows the web interface demo indicating how to use the system to perform the event discovery for Trump. First, we specify parameters *Keywords*, *Start Date*, *End Date*, and *Maximum Daily Tweets to Retrieve* to collect right tweets. These tweets will be classified by the sentiment classifier with a label either positive, or negative, or neutral. The parameter for the *Type of Sentiment Change* is also required. When it is specified as positive, positive sentiment changes will be computed to generate a time series plot of daily positive sentiment ratio change. Then we can query the event on a specific day by specifying parameters *Query Top-ranked Tweets/Keywords/Topics for Date* and *Number of Tweets to Display*. Then *Top-ranked Tweets*, *Top-ranked Keywords*, and *Topics* will be

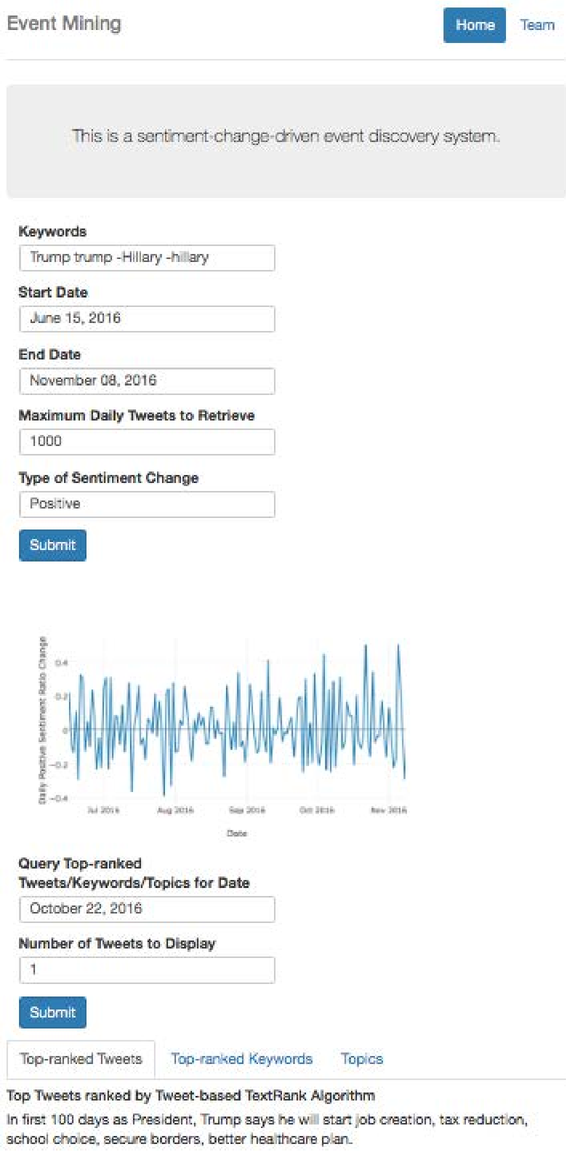

Figure 2: Web Interface Demo for Case Study Example

generated. In the following sections, each component of the system is illustrated in details based on its application on the 2016 U.S. presidential election.

3.2 Tweets Collection

In order to collect right tweets in an application, we first formulate the application scenario and then derive query parameters from it.

Taking the 2016 U.S. presidential election as an example, we feel that the answers towards the following questions can help us to gain insights on people's voting decisions.

1) When people talk about a candidate, are their words positive, negative, or neutral?

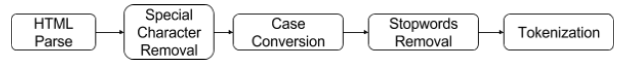

Figure 3: Tweets Normalization Flow

2) How have people's sentiments changed towards a candidate over time?
3) What have driven those significant sentiment changes?

In this application, the targets are the candidates Trump and Clinton respectively. For each day from June 16, 2015 to November 08, 2016, we collected 1000 tweets mentioning Trump but not mentioning Hillary and 1000 tweets mentioning Hillary but not mentioning Trump, respectively. Totally $1,020,672$ tweets were retrieved. By retrieving tweets in this way, we aimed to find out whether a candidate was talked about more positively or more negatively by people when they mentioned that candidate. The starting date June 16, 2015 was chosen for the purpose of comparison, because Trump officially announced his candidacy on that day (Hillary did that on April 12, 2015).

3.3 Tweets Normalization

The collected tweets are normalized by a general text normalization process, as shown in Figure 3. We first clean HTML and special characters in tweets, which are followed by case conversion and stopwords removal. Last, tweets are tokenized into a list of words.

3.4 Feature Matrix Building

Following the normalization, the system vectorizes each tweet by extracting features from it and transform it to a numeric vector. After vectorization, all tweets form a matrix with each column as a feature, called feature matrix. The feature matrix will be used in the sentiment classification and topic modeling. There are different ways of extracting features and building feature matrix from text data like bag-of-word and tf-idf document-term. The tf-idf document-term matrix is adopted in our system.

3.5 Sentiment Classification

One role of the sentiment classifier in this system is to act as a sentiment-detection sensor. With it, each tweet is classified to be positive, negative, or neutral, and then daily sentiment ratio for the target is calculated. The sentiment classification on tweets is done with Sentiment140, a sentiment analysis API developed by Stanford University [4]. More specifically, let's assume the number of positive, negative, and total tweets mentioning the target on a specific day to be n_{pos}, n_{neg}, and n_{tot}, respectively. The daily positive ratio for the target is then defined as $\frac{n_{pos}}{n_{tot}}$, while the daily negative ratio for the target is then defined as $\frac{n_{neg}}{n_{tot}}$. These ratios are time series, namely sentiment time series.

Figure 4 shows the daily positive ratio for Trump and Clinton since candidacy, which is from June 16, 2015 to November 08, 2016. Figure 5 shows the daily positive ratio for Trump and Clinton since primaries, which is from June 15, 2016 to November 08, 2016 when Trump and Clinton only campaigned against each other. From Figure 4 and Figure 5, we find that Trump's daily positive ratio is overall higher than Clinton's during the 2016 U.S. presidential

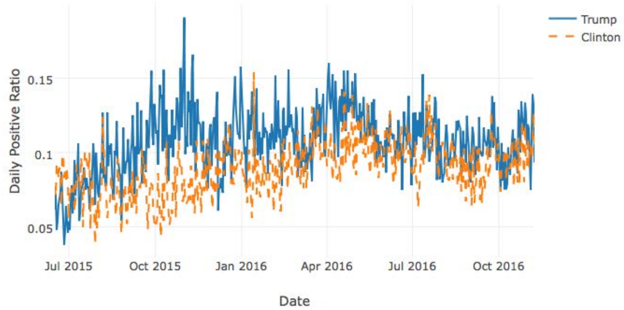

Figure 4: Daily Positive Ratio since Candidacy

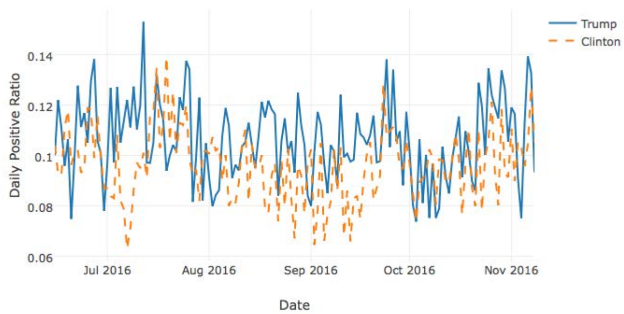

Figure 5: Daily Positive Ratio since Primaries

election. To avoid the underlying influence from other candidates before primaries, we will only focus on studying the sentiment changes towards Trump and Clinton since primaries in the next phases.

3.6 Sentiment Change Statistics

In this phase, the daily sentiment ratio change is calculated. Suppose the positive sentiment ratios on the current day and previous day are p_1 and p_2 respectively, and then the positive sentiment ratio change on the current day is $\frac{(p_1 - p_2)}{p_2}$. The negative sentiment ratio change is calculated in the same way. Then ratio changes are sorted and the tweets on days with highest ratio changes will be analyzed.

For the 2016 U.S. presidential election, Figure 6 shows the daily positive ratio change for Trump and Clinton since primaries, which is from June 15, 2016 to November 08, 2016. Table 1 lists the days with top 8 percentile positive ratio increase and top 8 percentile positive ratio decrease for Trump. Table 2 lists the days with top 8 percentile positive ratio increase and top 8 percentile positive ratio decrease for Clinton. In the next phase, we will discover the event associated with these sentiment changes.

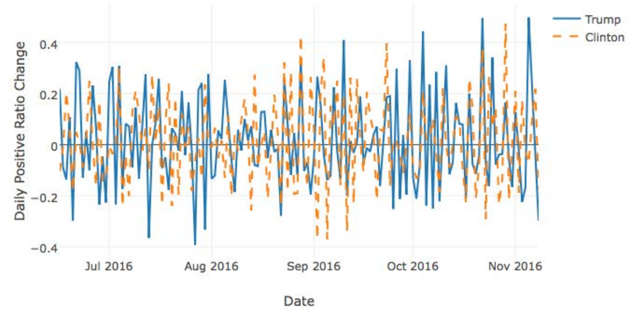

Figure 6: Daily Positive Ratio Change since Primaries

Table 1: Top 8 Percentile Positive Ratio Change for Trump

Effect	Date	Positive Ratio Change
Positive Increase	Oct. 22, 2016	50%
	Nov. 05, 2016	50%
	Oct. 04, 2016	44%
	Sept. 10, 2016	41%
	Oct. 25, 2016	34%
	Aug. 28, 2016	34%
	Sept. 30, 2016	33%
	Jun. 21, 2016	32%
	Oct. 11, 2016	31%
	Jul. 04, 2016	31%
	Jul. 02, 2016	30%
	Sept. 26, 2016	30%
Positive Decrease	Jul. 27, 2016	-39%
	Jul. 13, 2016	-36%
	Jul. 30, 2016	-33%
	Nov. 08, 2016	-30%
	Jun. 20, 2016	-30%
	Aug. 22, 2016	-28%
	Sept. 25, 2016	-25%
	Oct. 07, 2016	-25%
	Oct. 05, 2016	-24%
	Jun. 28, 2016	-23%
	Jul. 03, 2016	-23%
	Jun. 30, 2016	-22%

3.7 Tweets Rank and Topic Modeling

In order to discover what events are behind sentiment changes, the most straightforward way is to read through all related tweets and figure out what events caused the changes. But the number of tweets is usually huge, leading to high cost of time and labor for manual analyses. To extract the gist of the events from a huge number of tweets more efficiently, we utilize text summarization and topic modeling techniques. When performing text summarization,

Table 2: Top 8 Percentile Positive Ratio Change for Clinton

Effect	Date	Positive Ratio Change
Positive Increase	Oct. 29, 2016	47%
	Aug. 28, 2016	43%
	Sept. 23, 2016	40%
	Oct. 22, 2016	38%
	Sept. 04, 2016	34%
	Aug. 23, 2016	32%
	Jul. 04, 2016	30%
	Aug. 25, 2016	28%
	Jul. 14, 2016	27%
	Aug. 14, 2016	27%
	Oct. 24, 2016	27%
	Aug. 31, 2016	26%
Positive Decrease	Sept. 02, 2016	-37%
	Sept. 05, 2016	-37%
	Sept. 11, 2016	-34%
	Oct. 23, 2016	-29%
	Aug. 13, 2016	-26%
	Jul. 05, 2016	-24%
	Jul. 20, 2016	-24%
	Aug. 22, 2016	-24%
	Jul. 17, 2016	-24%
	Oct. 27, 2016	-23%
	Sept. 13, 2016	-23%
	Oct. 17, 2016	-22%

Table 3: Algorithms Used in Event Discovery

Algorithm	Name	Output
TS1	Tweet-graph TextRank	Top ranked tweets
TS2	Word-graph TextRank	Top ranked keywords
TT1	NMF	Multiple topics with each topic represented by keywords

we treat all tweets as a big document with each tweet as a paragraph. And when performing topic modeling, we treat all tweets as a document corpus with each tweet as an individual document. Table 3 lists three algorithms used in this context. The details of each algorithm in Table 3 can be found below.

TS1: Construct tweets as nodes and tweets similarity as edge weights. Then run PageRank algorithm to get top ranked tweets.

TS2: Construct words in tweets as nodes and add edges between two words if they co-occur within the window of size 2. The initial value on each node is set to be 1, and then run PageRank algorithm.

TT1: First build tf-idf document-term matrix on tweets and then apply NMF to generate topics.

Table 4: Events Driving Sentiment Changes

Candidate	Date	Event
Trump	Oct. 22, 2016	***Top 1 Tweet***: "In first 100 days as President, Trump says he will start job creation, tax reduction, school choice, secure borders, better healthcare plan." ***Effect***: Positive Increase
Trump	Jul. 27, 2016	***External Source***: "Trump calls on Russia to find Clinton's missing emails." ***Effect***: Positive Decrease
Clinton	Oct. 29, 2016	***External Source***: "FBI reviews emails related to Clinton's case." ***Effect***: Positive Increase
Clinton	Sept. 02, 2016	***Top 1 Tweet***: "BREAKING FBI NEWS: Hillary Clinton Lost Laptop With Classified Data." ***Effect***: Positive Decrease

Besides applying the text mining algorithms on the collected tweets, a problem-dependent module, called External, can be added into the system to facilitate the event discovery. It is used to access to other related information sources beyond tweets through APIs, which may help eliminate the potential bias that may be caused by the collected tweets.

For the 2016 U.S. presidential election, based on Table 1 and Table 2, we have the following observations:

1) The largest positive ratio increase for Trump occured on October 22, 2016.
2) The largest positive ratio decrease for Trump occured on July 27, 2016.
3) The largest positive ratio increase for Clinton occured on October 29, 2016.
4) The largest positive ratio decrease for Clinton occured on September 02, 2016.

What events were behind sentiment changes listed above? Table 4 shows the events on each date for each candidate, discovered by the top 1 tweet ranked by algorithm TS1 and the external source (i.e. Google news related to a candidate on a specific day). For Trump, his first 100-day plan won the favor for him, while people didn't like his call on Russia to find Clinton's missing emails. For Clinton, more people were driven to be positive towards her on the next day when FBI started to review emails related to her case. However, Clinton's email incident brought more people to be negatitve at the early stage of the incident, as shown by the top 1 tweet related to her on September 02, 2016. Furthermore, based on the outputs of keywords generated by TS2 and topics generated by TT1, as shown in Table 5, words including "email" and "fbi" are ranked high, which is consistent with the top 1 tweet representing what people were talking about Clinon on September 22, 2016. We further find that the email incident has long-lasting influence. When analyzing important keywords generated by TS2 on all days listed in Table 2,

Table 5: Keywords and Topics for Clinton on Sept. 02, 2016

Date	TS2 Keywords	TT1 Topic 1	TT1 Topic 2
Sept. 02, 2016	hillary fbi Clintonsmemory emails email	lost team actual email laptop full mail clinton hillary	remember recall time could damage brain say fbi

one observation is that the email-related keywords are ranked high on six out of twelve days (i.e. September 02, 2016; August 13, 2016; July 05, 2016; July 20, 2016; August 22, 2016; October 17, 2016).

3.8 Performance Evaluation

To justify the techniques introduced in the system, we set up two experiments. The data used in both experiments is from tweets collected towards Trump on October 22, 2016.

To validate the performance of sentiment classifier as the filter, we use two sets of tweets as the input of the algorithm TS1 respectively in the Experiment 1,

1) Tweets Set 1: All tweets.
2) Tweets Set 2: Only tweets labeled as "positive". Note that there is significant positive sentiment increase on October 22, 2016, as shown in Table 1.

Top 1 tweet ranked from these two sets of tweets are as follows.

1) Tweets Set 1: "Keefe Undercover Video Reveals Dems Planning to Use Women as Secret Weapon Against Trump Fans."
2) Tweets Set 2: "In first 100 days as President, Trump says he will start job creation, tax reduction, school choice, secure borders, better healthcare plan."

As shown, when ranking Tweets Set 1, we get misleading result. But when ranking Tweets Set 2, a more reasonable result is obtained, consistent with what has brought up positive sentiment increase for Trump. By using Tweets Set 2, the running time is also reduced considerably, considering the reduced size of tweets input, because Tweets Set 2 is only a portion of Tweets Set 1.

To validate the performance of NMF in this context, we compare the performance of topic modeling with NMF and LDA in the Experiment 2, where both of them are used to generate two topics from Tweets Set 2. As shown in Table 6, NMF is able to pick "gettysburg" and "speech" which are keywords for the event on that day, while LDA is not able to. We further find that Topic 1 generated by either NMF or LDA is consistent with what happened on October 22, 2016 related to Trump's speech about his first 100-day plan in Gettysburg, PA, which is also also indicated by results of tweet-based TextRank algorithm and word-based TextRank algorithm. Topic 2 does not provide very good interpretable information.

4 CONCLUSIONS

By using the sentiment classifier as a sensor, we can successfully detect events when they happen and measure their importance based on people's sentiment changes. Also as the filter, the sentiment classifier effectively eliminates the noise from tweets. Tweet-based TextRank algorithm along with NMF and the word-based TextRank can be combined to provide comprehensive overviews of events. We also find that NMF performs better than LDA in terms of important keywords provided in topics in the context of the presidential election. And only one topic generated by topic modeling algorithms is consistent with the results of Tweet-based Text-Rank algorthms, while other topics do not contribtue much to the interpretation of the event. By applying this system to the 2016 U.S. presidential election, we have discovered some events that have driven people's sentiment changes towards the two candidates.

Table 6: Topic Modeling Results for Trump on Oct. 22, 2016

Date	NMF Topic1	NMF Topic2	LDA Topic1	LDA Topic2
Oct. 22, 2016	day 100 plan first speech video gettysburg lay via trump	latina spanish shirt eric hilarious wear trick photo Bowl cuz	trump vote realdonaldtrump day voting first get good 100 go	van jean damme claude interview need tell great like say

5 FUTURE WORK

Currently, the system displays the outputs of three algorithms to represent event descriptions, namely, top tweets from Tweet-based Text-Rank algorithm, top keywords from word-based TextRank, and topics generated by nonnegative matrix factorization. In the future, we plan to develop a new topic modeling algorithm that can automatically combine the advantages of all three methods and generate better results than individual methods. For that, we will apply a generic measurement to evaluate different event discovery techniques comprehensively. Moreover, we plan to apply this system to financial markets, such as digging the reasons for sentiment changes on stock investments. Based on the specific application scenario of the system, more statistical analysis will be conducted to help mine the sentiment change patterns. For example, Z score can be used to quantify the statistical significance of a sentiment change, which can serve as a reference for its practical significance.

REFERENCES

[1] Deepayan Chakrabarti and Kunal Punera. 2011. Event Summarization Using Tweets. *ICWSM* 11 (2011), 66–73.
[2] Jaegul Choo, Changhyun Lee, Chandan K Reddy, and Haesun Park. 2013. Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 1992–2001.
[3] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1, 12 (2009).
[4] Alec Go, Richa Bhayani, and Lei Huang. 2012. Sentiment140. http://help.sentiment140.com/api/. (2012). Accessed: 2016-09-15.
[5] Kunal Jagtap. 2016. 2016 USA Presidential Election Twitter Sentiment Analysis & Topic Modeling. https://www.linkedin.com/pulse/2016-usa-presidential-election-twitter-sentiment-analysis-jagtap. (2016). Accessed: 2016-11-15.
[6] Praveen Kumar. 2016. Campaigns, Polls and People: Sentiment Analysis of tweets. https://www.linkedin.com/pulse/

campaigns-polls-people-sentiment-analysis-tweets-praveen-kumar. (2016). Accessed: 2016-11-15.

[7] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into texts. Association for Computational Linguistics.

[8] Bruno Stecanella. 2016. Donald Trump vs Hillary Clinton sentiment analysis on Twitter mentions. https://blog.monkeylearn.com/donald-trump-vs-hillary-clinton-sentiment-analysis-twitter-mentions. (2016). Accessed: 2016-11-15.

[9] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 115–120.

[10] Jianshu Weng and Bu-Sung Lee. 2011. Event detection in twitter. *ICWSM* 11 (2011), 401–408.