Exploring Government Uses of Social Media through Twitter Sentiment Analysis

Hsuanwei Michelle Chen[1], Patricia C. Franks[2]
San Jose State University
United States
[1]hsuanwei.chen@sjsu.edu
[2]patricia.franks@sjsu.edu

Lois Evans
University of British Columbia
Canada
levans18@mail.ubc.ca

Journal of Digital Information Management

ABSTRACT: *As social media becomes an important platform for organizations to use to interact with users, the ability to understand user opinions in social media communications has gained increased attention. One of the most popular approaches for exploring user opinions is sentiment analysis, which employs natural language processing, statistics, or machine learning to extract the sentiment of a text unit in terms of positive or negative attitudes. However, the effectiveness, interpretation, and accuracy of sentiment analysis rely heavily on the context in which it is conducted. In this paper, we investigate three sentiment analysis techniques for Twitter use by governments with their citizens, including a lexicon-based approach, a machine learning-based approach, and a hybrid approach. Our results reveal that, while each technique is developed based upon different rationales, the results are statistically robust and comparable. The study provides new insights into sentiment analysis in the context of government uses of social media.*

Categories and Subject Descriptors
H.3.5 [Online Information Services]: Web-based services; **I.2.7 [Natural Language Processing]:** Text analysis; **H.2.8 [Database Applications]:** Data mining

General Terms: Sentiment analysis, social media

Keywords: Sentiment Analysis, Opinion Mining, Lexicon-Based, Twitter, Social Media, Social Media Mining, User Engagement, Electronic Government

Received: 1 May 2016, Revised 4 June 2016, Accepted 10 June 2016

DOI: 10.6025/jdim/2016/14/5/290-301

1. Introduction

With both growing popularity and prevalence, social media is considered a platform on which human opinions, comments, thoughts, and attitudes are expressed, shared, exchanged, or even influenced. For example, Twitter users build social relationships with friends and strangers by sharing short messages of interests and activities. This user-generated content on social media has become valuable assets to organizations and businesses, as they often contain significant information for better strategies and decision-making. Many businesses, cultural organizations, and social institutions are leveraging social media to achieve their own strategic goals. According to research that has assessed the social media activity of the top 100 most valuable global brands, the brands that were the most socially active saw an 18% increase in their revenue for the previous year, while the least active experienced a 6% revenue decrease during the same period [1].

One of the most effective approaches for exploring and understanding these opinions is sentiment analysis.

Sentiment analysis is a technique that uses natural language processing, statistics, or machine learning methods to extract, identify, or characterize the sentiment content of a specific text unit ([2], [3]) in terms of feelings, attitudes, emotions, and opinions. Sentiment analysis has been widely applied in a variety of disciplines, ranging from business, politics, law or policy-making, and sociology and psychology to better understand online user sentiments and provide appropriate and timely responses ([4], [5]).

The effect and accuracy of sentiment analysis, however, relies heavily on the context in which it is conducted. Both local and global contextual information affects sentiment analysis and the approaches to modelling complex linguistic structures in sentences often can result in a failure to interpret sentiment through capturing of contextual cues [6]. Therefore, how different sentiment analysis techniques perform in different contexts is an important research issue with both academic and practical impacts. In this study, we conduct an investigation of sentiment analysis techniques for the government uses of Twitter. In particular, we examine and compare three main types of sentiment analysis approaches through the lens of how citizens respond to government posted messages on Twitter, using a lexicon-based approach, a machine learning-based approach, and a hybrid approach called SentiStrength [7].

The application of these techniques to the selected, specific context considered two concepts. First, local, state and federal governments use Twitter for different purposes that range from crime prevention and police assistance, emergency alerts and severe weather updates, activities and class registration, to public service announcements [8]. How citizens respond to these messages can significantly determine how effective these government social media efforts are, and how these efforts may potentially affect the on-going relationship between government and its citizens. Sentiment analysis is one of the first methods used to address this important issue by exploring and better understanding citizen attitudes, opinions, and thoughts toward government posted messages. Second, the selected three techniques cover the broad spectrum of sentiment analysis methods to provide a fair, representative comparison of the three different sentiment analysis techniques for the selected context.

The rest of the paper is organized as follows. Prior research is presented in the Related Work section. Next, we discuss data collection and its retrieval process in the Data and Methodology section, including main discussions of the sentiment analysis techniques used. The results are given in the Results and Findings section. The paper ends with a section titled Conclusions and Limitations.

2. Related Work

Today, government officials and public institutions are using social media sites like Facebook, Twitter, and YouTube to connect with citizens. The main reason why government agencies are increasingly adopting social media is it can play an important role in influencing and growing the government-citizen relationship. For example, Bertot et al.[9] state that the interactive and instant capabilities of social media make it a promising tool for increasing democratic participation. Song and Lee [10] agree and state that social media works as a complementary communication and participation channel of government. The authors further state that social media has created a new type of citizen-government interaction, wherein author content can play a huge role in increasing citizen trust in government. In their research, Graham and Avery [11] argue that social media can not only help government interact and engage with citizens but also help them build effective relationships with citizens and ultimately meet citizens' expectations for transparency in government. However, care about message truthfulness is needed, as indoctrinated citizens or the spread of false information are common issues on social media platforms [12].

Considering the importance of social media for influencing public trust in government, governments of countries around the world have initiated several programs to direct government officials on how to use social media to communicate with their citizens. Several studies have examined the role of government in its use of social media for the government-citizen relationship. For example, Nam [13] studied American citizens' attitudes on the adoption of social media use by government. The author found that the use of social media by government contributed to positive attitudes toward government. Song and Lee [10] also studied the new types of citizen-government interactions that are enabled by social media and found that the use of social media services by government significantly increases trust in government. In another study, Hong [14]examined the experiences of 2,000 American citizens with government social media usage and their perception of the government-public relationship. The author found that experiences with informational online services and social media were associated with greater trust in government at the local and state levels. In spite of all the benefits that social media provides government, social media remains highly underutilized by government agencies [11]. In fact, Lee and Kwak [15]note that several social media-based public engagement initiatives launched by U.S. federal agencies do not deliver their intended outcomes because of certain organizational, technological, and financial challenges. Moore [16] suggests that governments should focus on enhancing the two-way interactions between government and citizens using the features of social media.

One of the most popular and effective approaches for facilitating these "two-way" interactions on social media is gaining a better understanding of user opinions and attitudes. The technique of mining opinions, also commonly known as sentiment analysis, refers to an an automated method of extracting, identifying, or characterizing attitudes, opinions, and emotions from text, speech and database sources into categories like "positive," "negative," or "neutral" using natural language processing, machine learning, and statistical methods [2]. This process of sentiment analysis can be divided into three stages [17]. First, the input text is divided into smaller units, such as words. Next, these words are analyzed either through lexicon matching or machine learning classification to detect their sentiment polarity or semantic orientation [2]. Finally, the overall sentiment of a text unit is extracted [18]. To complete this three-stage process, there are two main approaches that have been commonly used: the lexicon-based approach and the machine learning-based approach. A lexicon-based approach uses a lexicon (or a dictionary) that contains already pre-classified "positive" and "negative" words for matching with the data and identifying the sentiments ([19],[20], [21], [22]). A sentiment score is usually calculated based on the statistical distribution of positive and negative words matched in a text unit, leading to a classification of a positive, negative, or neutral sentiment. A machine learning-based method, on the other hand, develops a classification model using training data with pre-labelled sentiments. The machine learning algorithms are then used to identify the general features associated with positive and negative sentiments, where these features are a subset of the words in the text unit or n-grams (e.g., [23],[24], [25],[26]). The model is further applied to classify future data into pre-defined categories, such as positive or negative. There are also more advanced, hybrid techniques that integrate methods from lexicon-based and machine learning-based approaches, with linguistic knowledge then added. For example, SentiStrength [7] employs novel methods to simultaneously extract positive and negative sentiment strength from short informal electronic text. This technique uses a dictionary of sentiment words with associated strength measures and a range of recognized non-standard spellings and other common textual methods for expressing sentiment.

City Name	Twitter Account	Date Joined	# of days' Presence as of 8/25/14	# of posts Between 1/1/13 & 8/25/14	# of Followers as of 8/25/14	# of Citizen Responses Between 1/1/13 & 2/51/14
U.S.						
Atlanta, GA	@cityofatlanta	2/19/09	2,013	319	44,600	10,064
Austin, TX	@austintexasgov	5/18/09	1,925	3,637	43,400	27,816
Boston, MA	@notifyboston	3/19/10	1,620	5,941	77,900	35,643
Honolulu, HI	@honolulugov	10/7/10	1,418	4,198	9,772	1,255
Kansas City, MO	@kcmo	5/21/09	1,922	6,040	28,500	25,747
Mesa, AZ	@mesaazgov	7/29/08	2,218	2,228	4,422	1,925
New York City, NY	@nycgov	2/11/11	1,291	7,311	191,000	69,497
Raleigh, NC	@raleighgov	1/3/09	2,050	1,125	16,200	7,053
Riverside, NC	@riversidecagov	1/20/09	2,043	4,230	7,401	5,679
Seattle, WA	@cityofseattle	1/14/09	2,049	159	22,100	7,350
Canada						
Calgary	@cityofcalgary	8/21/08	2,195	9,697	104,000	53,441
Edmonton	@cityofedmonton	2/5/09	2,027	5,096	68,700	64,837
Halifax	@hfxgov	6/4/10	1,543	2,340	11,800	13,659
Montreal	@mtl-ville	6/17/11	1,165	1,038	11,100	11,502
Ottawa	@ottawacity	12/5/08	2,089	5,119	42,700	48,615
Regina	@cityfregina	9/18/09	1,802	477	24,100	18,939
Surrey	@cityofsurrey	9/27/10	1,428	3,686	9,689	21,942
Toronto	@tornotocomms	1/22/09	2,041	1,368	56,100	18,969
Vancouver	@cityofvancouver	7/9/09	1,873	4,906	48,400	42,748
Winnipeg	@cityofwinnipeg	10/5/09	1,785	4,807	15,700	19,521

Table 1. Descriptive Summary of 20 City Twitter Accounts

3. Data and Methodology

For this study, we collected Twitter data from 20 city government Twitter accounts. The collection period was from January 1, 2013 to August 25, 2014. The 20 cities included 10 from the U.S. and 10 from Canada, chosen with the objective of diversity in both geographic location and population. All re-tweets were considered as normal tweets for this analysis. Table 1 presents a descriptive summary of the collected data set for the 20 city accounts.

The data for the 20 Twitter accounts were retrieved through Twitter Python API's (get_user_timeline) and included both tweets and re-tweets made as responses to the government accounts. The data collected were saved in the JSON format, done in Python, to retrieve the list of tweets and save them in a tabular format. The tabular data was used for sentiment analysis of the content field, which contained the actual tweet text. Finally, the retrieved data were cleansed by removing symbols, punctuation, special characters, URLs, and numbers for a precise sentiment analysis.

Figure 1 depicts the overall methodology and the flow of each analysis step used for this study.

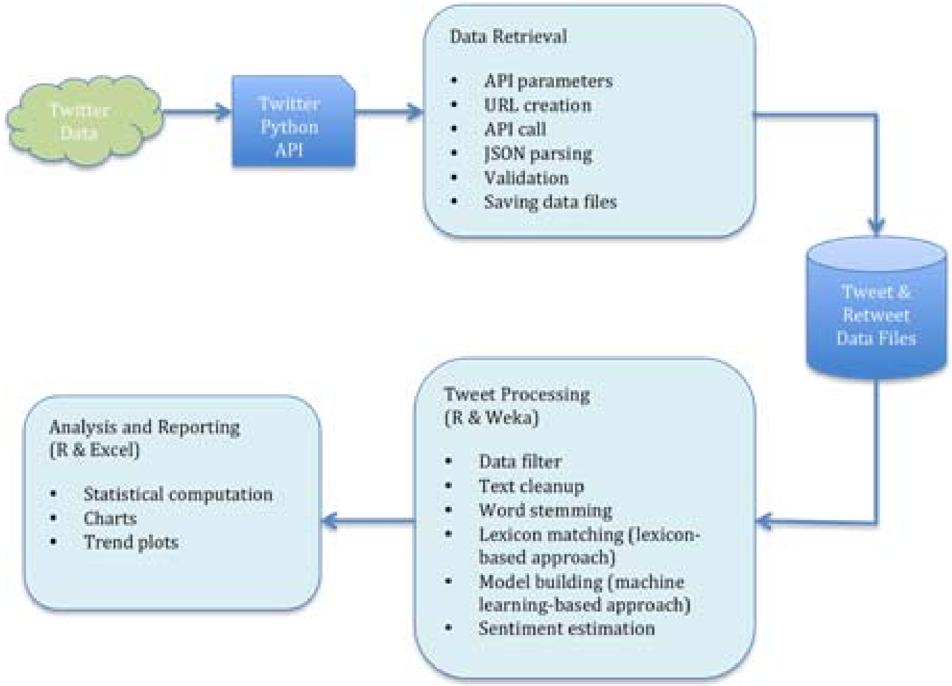

Figure 1. The Flow of Step-by-Step Sentiment Analysis Process

3.1 Sentiment Analysis: Lexicon-Based Techniques

To investigate lexicon-based techniques for sentiment analysis, we adopted a dictionary-matching approach. This type of approach uses dictionaries of words annotated with their semantic orientation, or sentiment, and matches the text that needs to be analyzed with the dictionary to determine the text's sentiment label: positive, negative, or neutral. In other words, the dictionary is used for the process of assigning a positive, negative, or neutral label to a text to capture the text's opinion, sentiment, or attitude within its context, and in this case, the government use of Twitter. While this method is relatively less involved with a machine learning or full linguistic analysis, it is considered a well-performed, robust, and effective approach [27].

To implement a rigorous lexicon-based approach, the first step is to choose a dictionary thatconsists of acomprehensive list of words with their semanticorientation annotated as positive, negative, or neutral.

To achieve this goal, in this study, we adopted a combined-lexicon approach, where three lexicons were used and weighted for sentiment matching and calculation. This approach has the benefits of generating higher accuracy and higher confidence in the sentiment analysis results. The three adopted lexicons include:

1. The dictionary developed by Taboada et al., which has been carefully designed and used in the work published by *Computational Linguistics* and has been widely cited [27]. In this dictionary, a comprehensive list of individual words has been provided with both their sentiment polarity and strength. To be more specific, the dictionary consists of a list of 2,827 positive and negative adjectives, such as *priceless* (positive), *awesome* (positive), *humiliating* (negative), and *vicious* (negative), a list of 876 positive and negative adverbs, such as *flawlessly* (positive), *perfectly* (positive), *woefully* (negative), and *bitterly* (negative), a list of 219 positive and negativeinterjections, such as*tremendous* (positive), *incredible* (positive),*barely* (negative), and *arguably*(negative), a list of 1,550 positive and negative nouns, such as *beauty* (positive), *pride* (positive), *violence* (negative), and *curse* (negative), and a list of 1,142 positive and negative verbs, such as *succeed* (positive), *amuse* (positive), *moan* (negative), and *hinder* (negative).

2. The Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon, which is specifically attuned to sentiment analysis for social media text [28]. With this lexicon, the positive, negative, or neutral sentiment of each word is weighted based on its semantic meaning, its relationship with nearby texts, whether it is capitalized, and with which punctuation it is associated. These

"heuristics" are carefully developed based on linguistic rules, making it an effective and appropriate lexicon for social media text analysis.

3. The National Research Council (NRC) Emotion Lexicon, which consists of a list of 14,182 unigrams (words) and totals around 25,000 senses that are associated with eight basic emotions, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, and with two sentiments, including positive and negative [29].

To provide a comparable analysis with the machine learning-based and hybrid techniques, when usingthese three lexicons, we only adopted the "polarity" of the words for sentiment analysis, i.e., positive or negative, and did not consider the "strength" of the words. In addition, for those words with polarity on the borderline, i.e., very weak negative and very weak positive, we treated them as "neutral" sentiments. The lexicons have also been extended,as needed,with bi-grams and tri-grams, in which we take into account the words that negate the meaning. For example, "not good" is considered negative, despite the fact that it contains the word "good".

Wealso pre-processed our data (tweets and re-tweets)based on some natural language processing rulesto provide a meaningful and accurate comparison with the three lexicons. The text pre-processing steps include:

• Using TweetTokenizer for tokenization, which includes removing Twitter mentions, treating hashtags as separate tokens, and shortening words that contain repeated symbols [30];

• Using a regular expression tool to remove non-alphanumeric characters [31];

• Splitting each tweet/re-tweet into a list of words;

• Applying the Natural Language Toolkit (NLTK) for stop-words removal and lemmatization [30];

• Using Porter Stemmer for stemming [32].

The lexicon-based analysis involves a comparison between the pre-processed tweets/re-tweets and the three lexicons respectively. Each pre-processed tweet/re-tweet corresponding to a certain government account was matched against each lexicon to classify each word into positive, negative, or neutral. For each tweet/re-tweet, a sentiment score was then calculated based on the distribution of positive, negative, and neutral words found in each lexicon.

3.2 Sentiment Analysis: Machine Learning-Based Techniques

To examine the robustness of the sentiment analysis results from the lexicon-based technique and further understand the citizens' sentiments, we developed a machine learning-based model for sentiment prediction and classification. We used the data mining software, Weka, to conduct sentiment analysis on the collected Twitter data [33]. Weka is an open-sourced platform that provides tools for various machine learning algorithms. It has become a widely adopted, standard tool in the data mining and machine learning community. Our sentiment analysis task was based on the tools provided by Weka using the following processes and configurations.

***Training data*:** An essential first step for building a predictive model is to prepare a training data set. In our study, we adopted the corpus provided by Sentiment140 [34], which has already been used in several prior studies and publications (e.g., [35], [36]). This corpus consists of 1.6M tweets, is balanced, and also captures emotion icons.

***Text pre-processing*:** To prepare our collected Twitter data for the machine learning task, we conducted text pre-processing, including word parsing and tokenization, stop-words removal, and lemmatization and stemming. This process helps the transformation of each textual unit into a vector form, in which each document is further represented by the presence (or frequency) of the terms declared important. Term selection and feature extraction were further performed to filter the terms with poor prediction ability or strongly correlated to other terms.

***Weka configuration*:** To perform pre-processing in Weka, we used the StringToWordVector filter from the package weka.filters.unsupervised.attribute and configured the tokeniser, specified a stop-words list, and chose a stemmer [37].

***Classifier selection*:** We chose three different algorithms to build our predictive model, i.e., NaïveBayes, K-Nearest Neighbors, and Random Forests. These three methods are briefly explained as follows.

• **Naïve Bayes:** The Naïve Bayes method is a probabilistic classifier that is based on Bayes' theorem with an assumption of independence between features. This classifier uses a maximum likelihood principle to assign each unlabelled instance a class and represent features using vectors [38].

• **K-Nearest Neighbors:**The K-Nearest Neighbors method is a non-parametric algorithm that assigns an instance to a class by a majority vote of its neighbors, i.e., the instance is assigned to the class most common among its *k* nearest neighbors.We chose *k* to be an odd number, 3, so that a majority class always exists [40].

• **Random Forests:** The Random Forests method uses multiple learning algorithms to obtain better predictive results, including classification, regression, and other tasks [39]. With Random Forests, a multitude of decision trees are constructed with training data, and the resulting class is either the mode of the classes (using a "classification" algorithm) or the mean prediction (using a "regression" algorithm) of individual trees.

For the three selected classifiers,we considered their different requirements on bias and variance for training data sets to avoid biases from training data selection[40]. We then applied the three classifiers to the training data with 10-fold cross-validation and evaluated the different classifiers with standard accuracy features, including a true positive rate and a false positive rate [40].

3.3 Sentiment Analysis: Hybrid Techniques

To provide a fair and comprehensive comparison of our sentiment analysis techniques, we further expanded this study by including a third method, SentiStrength [7], which has been described and evaluated in academic articles (e.g., [41], [5]). We consider it a hybrid technique. SentiStrength provides estimates of positive and negative sentiments in short or even informal texts. A unique feature of SentiStrength is that it also reports single scale (-4 to +4) results, which complements our previous methods in which only binary sentiments were identified.

Figure 2 provides an architectural view of the three sentiment analysis techniques that were adopted in this study.

4. Results and Findings

In this section, we offer a comparison of sentiment analysis results,using the three sentiment analysis techniques. These results include an overall comparison of Twitter posts for all cities, followed by a case study of one chosen city to further examinethe three techniques.

To understand the overall sentiment analysis for all Twitter messages collected using the three techniques and statistically examine the distribution of these sentiments, we first coded the sentiments using the following scheme:

0: neutral sentiment
+1: positive sentiment
-1: negative sentiment

The sentiment means and the standard deviations from these three techniques, respectively, were then calculated. Table 2 presents the percentages of positive, negative, and neutral sentiments from all city accounts, followed by the means and standard deviations of these sentiments given below in Table 3.

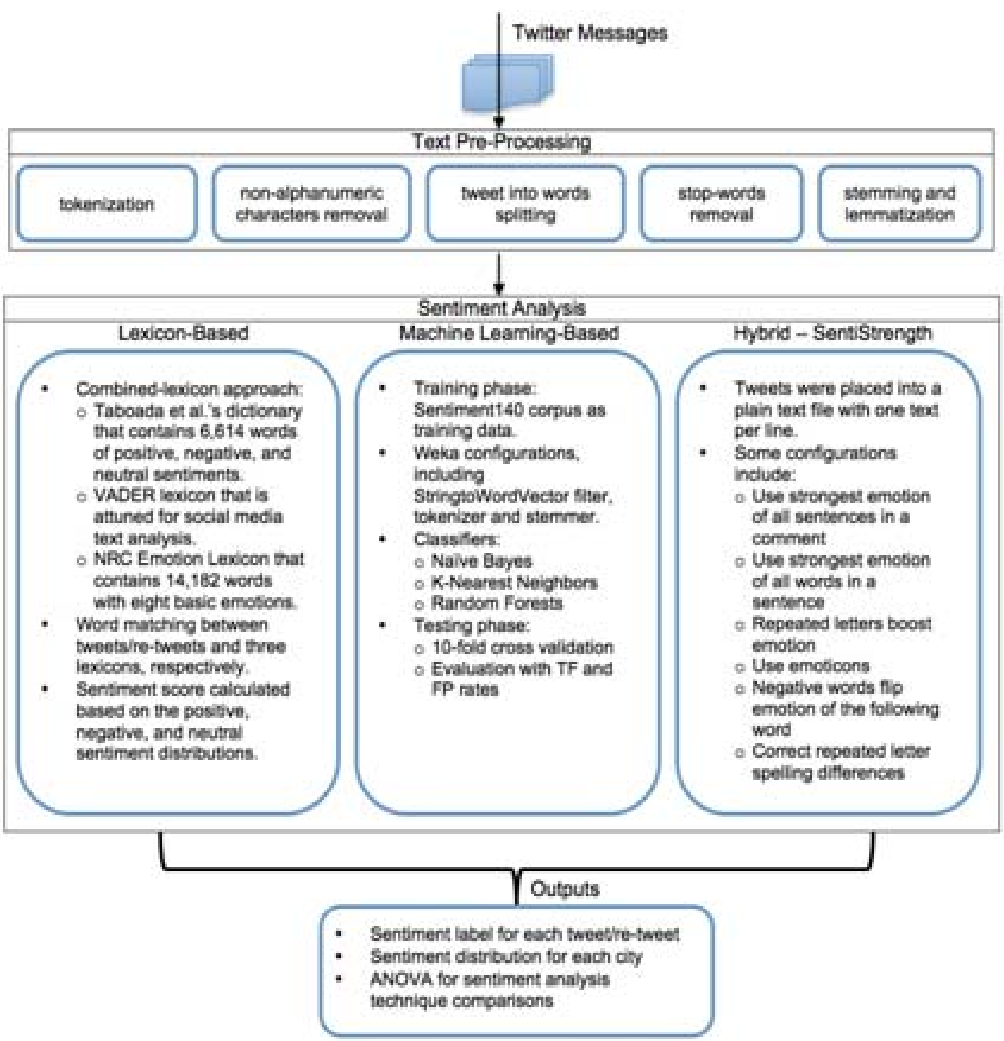

Figure 2. An Architectural View of Sentiment Analysis Techniques

Twitter Account	Lexicon - Based Approach			Machine Learning - Based Approach			Senti Strength 0,-1,1 : neutral 2, 3, 4 : positive -2,-3,-4 : negative		
Sentiment Percentage (%)									
	Pos.	Neg.	Neutral	Pos.	Neg.	Neutral	Pos.	Neg.	Neutral
@cityofatlanta	23.3	9.0	67.7	18.4	6.4	75.2	20.0	7.0	73.0
@austintexasgav	17.6	8.9	73.5	24.8	5.0	70.2	19.1	3.0	77.9
@notifyboston	15.0	14.5	70.3	25.8	6.8	67.4	20.7	4.0	75.3
@honolulugov	13.8	11.9	75.0	24.3	7.8	67.9	22.5	12.0	65.5
@kcmo	22.7	6.8	70.5	24.3	11.6	64.1	29.0	3.0	68.0
@mesaazgov	20.9	5.7	72.4	31.8	6.5	61.7	19.5	7.0	73.5
@nycgov	15.3	7.9	76.7	30.0	5.0	65.0	18.6	4.0	77.4
@raleighgov	21.0	6.7	72.3	27.0	6.9	66.1	21.5	3.0	69.9
@riversidecagov	25.9	4.7	69.5	17.8	12.5	69.7	27.1	3.0	69.9
@cityofseattle	22.5	9.5	68.0	21.4	8.8	69.8	26.3	12.0	61.7
@cityofcalgary	19.3	10.5	70.2	22.3	7.0	70.7	23.7	8.0	68.3
@cityofedmonton	19.4	10.3	70.3	27.4	4.5	68.1	18.6	12.0	69.4
@hfxgov	15.2	11.0	73.8	25.0	2.3	72.7	11.5	5.0	83.5
@mtl-ville	3.5	3.3	93.2	8.4	2.3	89.3	16.8	10.0	61.5
@ottawacity	16.8	8.2	75.0	16.6	12.7	70.7	28.5	10.0	61.5
@cityofregina	16.7	13.7	69.6	23.9	5.1	71.0	18.1	3.0	78.9
@cityofsurrey	26.4	7.9	65.7	23.0	6.7	70.3	21.1	3.0	75.9
@torontocomms	13.5	9.1	77.5	24.2	3.4	72.4	15.2	4.0	80.8
@cityfvancouver	22.3	9.2	68.5	30.0	2.7	67.3	11.6	5.0	83.4
@cityofwinnipeg	14.4	15.7	69.9	17.3	6.5	76.2	21.3	8.0	70.7

Table 2. Percentages of Positive, Negative, and Neutral Sentiments Using 3 Techniques for 20 City Accounts

To statistically investigate whether the results of the three sentiment analysis techniques differed significantly or not, we performed an ANOVA test on the sentiments. These results are given in Table 4 and Figure 3.

The ANOVA test shows that, at an aggregate level, the three sentiment analysis techniques, while functioning based on different rationales and algorithms, also provide a statistically consistent and robust result.

5. Case Study: The City of Austin, Texas

To further explore how these three sentiment analysis techniques perform at a finer level, we chose to focus and present our analysis for the City of Austin, Texas. Austin is a mid-sized city of about 800,000 people and is the capital city of the state of Texas. Austin is known for its independent spirit, with “Keep Austin Weird” a prominent slogan, along with “The Live Music Capital of the World.” Austin has the stated goal of being the “best managed city” in the United States. The city launched Facebook, Twitter, and YouTube accounts in 2009.

We first randomly selected 10 Twitter messages in response to the selected “@austintexasgov” city account. Table 5 presents these findings, in which the actual message and the estimated sentiments from all three techniques are also given.

The results show that for these randomly selected 10 messages, the sentiment predictions using the lexicon-based approach and the machine-learning approach were identical. There were some slight differences in sentiment predictions between SentiStrength and the other two approaches, specifically for Tweets #1 and #6. If we take a closer look at these tweet contents, we can conclude

TwitterAccount	Lexicon-Based Approach	Machine Learning-Based Approach	SentiStrength
	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)
@cityofatlanta	0.14 (0.57)	0.12 (0.48)	0.13 (0.40)
@austintexasgov	0.09 (0.52)	0.20 (0.51)	0.16 (0.39)
@notifyboston	0.01 (0.55)	0.19 (0.54)	0.17 (0.41)
@honolulugov	0.02 (0.54)	0.17 (0.54)	0.10 (0.42)
@kcmo	0.16 (0.53)	0.13 (0.59)	0.26 (0.46)
@mesaazgov	0.15 (0.51)	0.25 (0.56)	0.13 (0.40)
@nycgov	0.07 (0.50)	0.25 (0.54)	0.15 (0.39)
@raleighgov	0.14 (0.53)	0.20 (0.55)	0.14 (0.41)
@riversidecagov	0.21 (0.54)	0.05 (0.55)	0.24 (0.45)
@cityofseattle	0.13 (0.57)	0.13 (0.53)	0.14 (0.44)
@cityofcalgary	0.09 (0.55)	0.15 (0.52)	0.16 (0.43)
@cityofedmonton	0.09 (0.56)	0.23 (0.52)	0.07 (0.39)
@hfxgov	0.04 (0.53)	0.23 (0.47)	0.06 (0.32)
@mtl_ville	0.0 (0.31)	0.06 (0.32)	0.07 (0.37)
@ottawacity	0.09 (0.52)	0.04 (0.54)	0.19 (0.45)
@cityofregina	0.03 (0.57)	0.19 (0.51)	0.15 (0.40)
@cityofsurrey	0.19 (0.57)	0.16 (0.52)	0.18 (0.41)
@torontocomms	0.04 (0.54)	0.21 (0.48)	0.11 (0.36)
@cityofvancouver	0.13 (0.57)	0.27 (0.50)	0.07 (0.32)
@cityofwinnipeg	-0.01 (0.56)	0.11 (0.48)	0.13 (0.41)

Table 3. Sentiment Means and Standard Deviations Using 3 Techniques for 20 City Accounts

Source	df	SS	MS	F	P-value
SA	2	0.06	0.03	**7.7841**	**0.001**
Error	57	0.22	0.004		
Total	59	0.28			

Table 4. One-Way ANOVA Test for the Sentiment Analysis (SA) Techniques

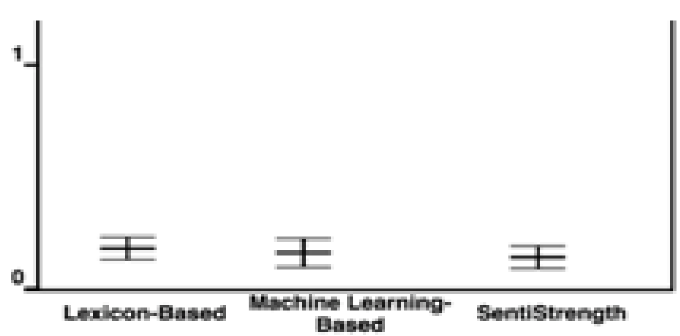

Figure 3. The Mean for Each Sentiment Analysis
Technique and a Vertical Error Bar Containing Values Within One Standard Deviation of the Mean

#	Tweets/Re-tweets	[A]	[B]	[C]
1	It's beautiful out at Austin's New Year! It's not too late to get down to Auditorium Shores for fireworks, Del Casti ...	+	+	n
2	Traffic signals not working at: Koenig at Shoal Creek, Koenig at Marilyn, 290 at Berkman. Plan ur commute. #wind #ATX	-	-	-
3	I'm at Lady Bird Lake Trail - @austintexasgov (Austin, TX) http://t.co/I3IIRrJUhy	n	n	n
4	@TheaGood @JohnCornyn @google @austintexasgov Thea, broadband via Google Fiber will be free. https://t.co/FRGEZigyrx	+	+	+
5	Thank You! Thank You! @WellsFargo @RepLloydDoggett @UT_DDCE @CapMetroATX @austintexasgov Austin Revitalization Authority for your support!	+	+	+
6	@austintexasgov: Do you buy local? Today is your last day to be vocal! Tell the City how you feel about locally grown foods here:	n	n	-
7	HA! Love this city. RT @austintexasgov #48 ‰ÛÒ NANANANANA, BAT FEST! This Aug. 24 fee-paid event just got its permit wings. #ATXcouncil	+	+	+
8	I'm at Austin, TX - @austintexasgov (Austin, TX) w/ 4 others http://t.co/Xnh6EiXbBo	n	n	n
9	@EddieforTexas: @austintexasgov Thank you to City Council for putting $65 million affordable housing bond package on Nov. ballot. http:/‰Û_	n	n	n
10	Austin, TX wins the '2013 Best of the Web' Award for government sites. Way-2Go @AustinTexasGov http://t.co/0Cw64p8pke #Austin	+	+	+

[A]: Sentiment prediction using the lexicon-based approach
[B]: Sentiment prediction using the machine learning-based approach
[C]: Sentiment prediction using SentiStrength
+: positive sentiment
-: negative sentiment
n: neutral sentiment

Table 5. 10 Randomly Selected Twitter Posts for "@austintexasgov" and Their Sentiment Predictions Using 3Techniques

conclude that it is largely because SentiStrength predicts sentiments as more than a binary classification and reports sentiments on a wider (-4 to +4) scale.

We also conducted a sentiment analysis to better understand the trends and patterns for how citizens responded to governments' use of social media—in this specific case, Twitter. To achieve this goal, we created two visual displays based on the sentiment analysis results for each city account, namely, the Twitter Sentiment Trends and the Comparison Word Cloud. The Twitter Sentiment Trends graph can be used to explore the changes in citizen sentiments over time, which may correspond to unique events, new policies, and important government announcements. The Comparison Word Cloud can be a powerful tool to use to understand the discussion interests of citizens on Twitter within a given period of time. We chose Austin, Texas (Twitter account:@austintexasgov) as an example to discuss these two graphs further.

Figure 4 presents the Twitter sentiment trends for @austintexasgov by showing the percentages of positive, negative, and neutral tweets per month, respectively, for the research period January 1, 2013 to August 25, 2014. The peaks and valleys in these trends may reveal how citizen sentiment changed in line with significant city events, announcements, and activities. For example, we noticed a spike in positive sentiments in February of 2014. We found that February was the month in which the Austin city government was promoting the upcoming world-famous SXSW (South by Southwest) festival, along with several other cultural and art events (e.g., "*We're now accepting applications for #ATX Creative Ambassadors*"; "*City of Austin announces new public art opportunity at Montopolis Neighborhood Center*"). On the other hand, we noticed a spike in negative sentiments in March of 2013, which might have resulted from arguments and discussions about the panellists who were selected for the redistricting commission (e.g., "*There were actually more women in the pool than men. Very few racial minorities to choose from, though*"; "… *Hopefully the applicant pool for the commission will be more diverse*"). These observations indicate how citizen sentiments can be driven by events, and that the government should value citizens' social media responses when making its decisions and designing its policies.

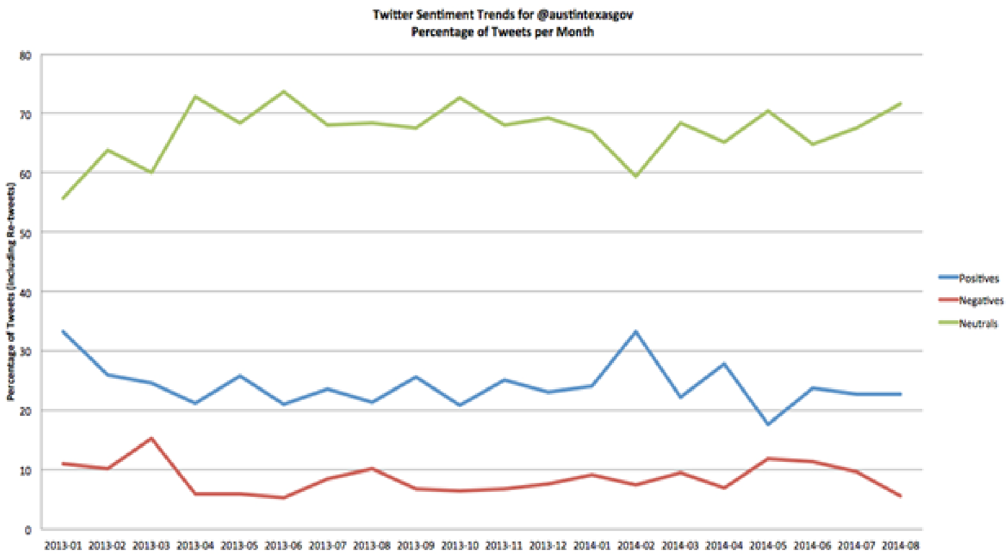

Figure 4. Twitter Sentiment Trends between 1/1/2013 and 8/25/2014 (@austintexasgov)

Figure 5.Word Cloud between 1/1/2013 and 8/25/2014 (@austintexasgov)

Finally, Figure 5 represents the word cloud of all tweets and re-tweets for @austintexasgov between January 1, 2013 and August 25, 2014. The cloudserves as an informative snapshot to understand what topics citizens cared about and became interested in within a given timeframe.

6. Conclusions and Limitations

In this study, we examined three sentiment analysis techniques performed on Twitter data in the specific context of citizens' responses to governments' Twitter posts. These three techniques include a lexicon-based approach, a machine learning-based approach, and a hybrid approach called "SentiStrength,"covering a wide spectrum of possible sentiment analysis techniques. The selected data provide one of the first attempts to examining and comparing these sentiment analysis techniques in the context of government use of social media. Our study contributes to the understanding of how sentiment analysis techniques can perform similarly, or differently, in the given context of government uses of social media. The different interpretations of governmental social media data and of citizen engagement on social media can greatly affect policy-making, government-citizen relationships, and public trust.

Our study also suggests how sentiment analysis results can be used to identify the trends and patterns of citizens' sentiments driven by events. This finding has two implications. First, citizens' sentiments can indeed be influenced by city events, activities, announcements, and many more, so it is critical for governments to take into account citizens' opinions via social media for making decisions and policies. Second, sentiment analysis has been shown again to be an effective tool for both identifying current sentiments and predicting future sentiments.This technique should be integrated into an open government's system thatencourages public trust, transparency, and public participation.

There are several limitations to this study. First, while we tried to examine a wide range of sentiment analysis techniques using the three representative types, there are still many other choices available that provide numerous variations in algorithms, features, effectiveness, and accuracy. How to develop a benchmark to provide a meaningful comparison will be an important issue. Second, the sentiment analysis techniques examined in this paper did not take into account the extensive uses of emotional icons and especially irony on social media. A sophisticated natural language processing algorithm that is context-aware is needed in order to capture the meanings of emotional icons and interpret irony more accurately. Finally, the sentiment analysis results can be greatly biased by citizen activities on social media. These activities highly depend on how governments use, manage, manipulate, and operate on social media. The future directions of our study include a deeper investigation of the social media initiatives, policies, and administrative operations undertaken in these city governments.

Acknowledgement

This research was conducted under the InterPARES Trust (I-Trust) project, which is funded in part by a Social Sciences and Humanities Research Council of Canada Partnership Grant (SSHRC Grant No. 895-2013-1004). We thank our colleagues from the I-Trust project who provided insight and feedback that greatly enhanced the research, although the interpretations and conclusions of this paper are strictly those of the authors.

References

[1] Factiva, D. J. (2009). Direct correlation established between social media engagement and strong financial performance. *PR News*, 65(29), p. 3.

[2] Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2 (1-2) 1-135.

[3] Vinodhini, G., Chandrasekaran, RM. (2012). Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2 (6) 282-292.

[4] Kale, A., Karandikar, A., Kolari, P., Java, A., Finin, T., Joshi, A. (2007). Modeling trust and influence in the blogosphere using link polarity. *ICWSM 2007*, Boulder, CO.

[5] Calderon, N. A., Fisher, B., Hemsley, J., Ceskavich, B., Jansen G., Marciano, R., Lemieux, V. L. (2015). Mixed-initiative social media analytics at the World Bank. *Proceedings of 2015 IEEE International Conference on Big Data*, Santa Clara, CA, 1678-1687.

[6] Yang, B., Cardie, C. (2014). Context-aware learning for sentiment-level sentiment analysis with posterior regularization. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD.

[7] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61 (12) 2544-2558.

[8] CivicPlus. (2016). 7 ways local government can use social media. CivicPlus, January 4, 2016. Retrieved February 20, 2016 from: http://www.civicplus.com/blog/seven-ways-local-government-can-use-social-media

[9] Bertot, J. C., Jaeger, P. T., Munson, S., Glaisyer, T. (2010). Engaging the public in open government: Social media technology and policy for government transparency. Retrieved January 25, 2016 from http://tmsp.umd.edu/TMSPreports_files/6.IEEE-Computer-TMSP-Government-Bertot-100817pdf.pdf

[10] Song, C., Lee, J. (2013). Can social media restore citizen trust in government? *Proceedings of the Public Management Research Conference*, Madison, WI.

[11] Graham, M., Avery, E. J. (2013). Government public relations and social media: An analysis of the perceptions and trends of social media use at the local government level. *Public Relations Journal*, 7 (4) 1-21.

[12] Castillo, C., Mendoza, M., Poblete, B. (2011). Information credibility on Twitter. *Proceedings of WWW 2011*, Hyderabad, India, 675-684.

[13] Nam, T. (2012). Citizens' attitudes toward Open Government and Government 2.0. *International Review of Administrative Sciences*, 78 (2) 346-368.

[14] Hong, H. (2013). Government websites and social media's influence on government-public relationships. *Public Relations Review*, 39 (4) 346-356.

[15] Lee, G., & Kwak, Y. H. (2012). An open government maturity model for social media-based public engagement. *Government Information Quarterly*, 29 (4) 492-503.

[16] Moore, A. (2013). *Looking beyond likes: Increasing citizen engagement with government Facebook pages*. University of North Carolina, NC. Retrieved February 20, 2016 from http://www.mpa.unc/sites/www.mpa.unc.edu/files/Allison%20Moore.pdf

[17] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Goot, E. v. d., Halkia, M., et al. (2010). Sentiment analysis in the news. *Proceedings of the Seventh Conference on International Language Resources and Evaluation*. Retrieved January 25, 2016 from: http://www.lrec-conf.org/proceedings/lrec2010/pdf/909_Paper.pdf.

[18] Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.

(2005). Pulse: Mining customer opinions from free text. *Lecture Notes in Computer Science*, *3646*, 121-132.

[19] Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: The MIT Press.

[20] Strapparava, C., Valitutti, A. (2004). Wordnet-affect: An affective extension of Wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (p. 1083-1086). Lisbon.

[21] Esuli, A., Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. *Proceedings of Language Resources and Evaluation (LREC) 2006*, Retrieved January 25, 2016 from: http://hnk.ffzg.hr/bibl/lrec2006/pdf/384_pdf.pdf

[22] Agerri, R., García-Serrano, A. (2010). Q-WordNet: Extracting polarity from WordNet senses. *Proceedings of the Seventh conference on International Language Resources and Evaluation*, Retrieved January 25, 2016 from: http://www.lrec-conf.org/proceedings/lrec2010/pdf/695_Paper.pdf

[23] Abbasi, A., Chen, H., Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, 26 (3) 12:11-12.34.

[24] Ng, V., Dasgupta, S., Arifin, S. M. N. (2006). Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. *Proceedings of the COLING/ACL 2006 Main Conference*, 611-618.

[25] Tang, H., Tan, S., Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications: An International Journal,* 36 (7) 10760-10773.

[26] Koto, F. Adriani, M. (2015). A comparative study on Twitter sentiment analysis: Which features are good? *Natural Language Processing and Information Systems*, 9103, 453-457.

[27] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37 (2) 267-307.

[28] Hutto, C. J., Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. AAAI 2014, Quebec City, Canada.

[29] Mohammad, S., Turney, P. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29 (3) 436-465.

[30] Natural Language Toolkit: http://www.nltk.org/

[31] Python Software Foundation: 7.2 re – Regular Expression Operations: https://docs.python.org/2/library/re.html

[32] Porter, M. (2006). Porter Stemming Algorithm: https://tartarus.org/martin/PorterStemmer/

[33] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. (2009). The Weka data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11 (1) 10-18.

[34] Go, A., Bhayani, R., Huang, L. (2009). Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project.

[35] Friedrich, N., Bowman, T. D., Stock, W. G., Haustein, S. (2015). Adapting sentiment analysis for tweets linking to scientific papers. Retrieved February 20, 2016 from http://arxiv.org/abs/1507.01967

[36] Kiritchenko, S., Zhu, X., Mohammad, S. F. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723-762.

[37] Scerra, S. (2014). A data mining experiment: movie review classification using Weka. Retrieved on January 25, 2016 from: http://www.stefanoscerra.it/movie-reviews-classification-weka-data-mining/

[38] Hand, D. J., Yu, K. (2001). Idiot's Bayes – not so stupid after all? *International Statistical Review*, 69 (3) 385-399.

[39] Ho, T. K. (1995). Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition. Montreal, QC, 14-16 August 1995, p. 278-282.

[40] Zou, H., Chen, H. M., Dey, S. (2015). Exploring user engagement strategies and their impacts with social media mining: The case of public libraries. *Journal of Management Analytics*, 2 (4) 295-313.

[41] Thelwall, M., Buckley, K. (2013). Topic-based sentiment analysis for the Social Web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*, 64 (8) 1608–1617.