



ELSEVIER

Contents lists available at [ScienceDirect](#)

# Information Processing and Management

journal homepage: [www.elsevier.com/locate/infoproman](http://www.elsevier.com/locate/infoproman)

## Sentiment, emotion, purpose, and style in electoral tweets

Saif M. Mohammad\*, Xiaodan Zhu, Svetlana Kiritchenko, Joel Martin

National Research Council Canada, Ottawa, Ontario K1A 0R6, Canada

### ARTICLE INFO

#### Article history:

Received 15 August 2013

Received in revised form 6 June 2014

Accepted 18 September 2014

Available online xxx

#### Keywords:

Sentiment analysis

Tweets

Elections

Emotions

Purpose

Crowdsourcing

### ABSTRACT

Social media is playing a growing role in elections world-wide. Thus, automatically analyzing electoral tweets has applications in understanding how public sentiment is shaped, tracking public sentiment and polarization with respect to candidates and issues, understanding the impact of tweets from various entities, etc. Here, for the first time, we automatically annotate a set of 2012 US presidential election tweets for a number of attributes pertaining to sentiment, emotion, purpose, and style by crowdsourcing. Overall, more than 100,000 crowdsourced responses were obtained for 13 questions on emotions, style, and purpose. Additionally, we show through an analysis of these annotations that purpose, even though correlated with emotions, is significantly different. Finally, we describe how we developed automatic classifiers, using features from state-of-the-art sentiment analysis systems, to predict emotion and purpose labels, respectively, in new unseen tweets. These experiments establish baseline results for automatic systems on this new data.

Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

### 1. Introduction

Elections, the cornerstone of democratic process, occur across the globe and often involve tens of millions of potential voters. Social media platforms such as Twitter give new opportunities to the electorate, the politicians, news corporations, and other participants to make their voice directly accessible to a large audience. However, the number of posts pertaining to a single event or topic such as a national election can grow to the hundreds of millions. The large number of tweets negates the possibility of a single person reading all of them to gain an overall global perspective. Thus, automatically analyzing electoral tweets, and specifically, analyzing sentiment and emotions in electoral tweets, can be beneficial for a number of downstream applications:

- *Understanding the role of target entities:* A number of entities tweet during elections, for example, the politicians, the voters, the disenfranchised, news corporations, non-governmental organizations, special interest groups, etc. Analyzing the extent to which tweets from various entities help shape public sentiment will improve our understanding of how social media is used during elections. It is also of interest to identify which portions of the voting electorate tweet about politics during elections. For example, some studies have shown that the more partisan electorate tend to tweet more, as do members from minority groups (Lassen & Brown, 2011).

\* Corresponding author.

E-mail addresses: [saif.mohammad@nrc-cnrc.gc.ca](mailto:saif.mohammad@nrc-cnrc.gc.ca) (S.M. Mohammad), [xiaodan.zhu@nrc-cnrc.gc.ca](mailto:xiaodan.zhu@nrc-cnrc.gc.ca) (X. Zhu), [svetlana.kiritchenko@nrc-cnrc.gc.ca](mailto:svetlana.kiritchenko@nrc-cnrc.gc.ca) (S. Kiritchenko), [joel.martin@nrc-cnrc.gc.ca](mailto:joel.martin@nrc-cnrc.gc.ca) (J. Martin).<http://dx.doi.org/10.1016/j.ipm.2014.09.003>

0306-4573/Crown Copyright © 2014 Published by Elsevier Ltd. All rights reserved.

- *Determining how public sentiment is shaped*: Some tweets (or some sets of tweets) have more impact in shaping public opinion than others. Determining characteristics of influential tweets is particularly useful.
- *Nowcasting and forecasting*: Tweet streams have been shown to help identify current public opinion towards the candidates in an election (nowcasting) (Conover, Gonçalves, Ratkiewicz, Flammini, & Menczer, 2011; Golbeck & Hansen, 2011). Some research has also shown the predictive power of analyzing electoral tweets to determine the number of votes a candidate will get (forecasting) (Birmingham & Smeaton, 2011; Lampos, Preotiuc-Pietro, & Cohn, 2013; Tumasjan, Sprenger, Sandner, & Welpe, 2010a), however, other research expresses skepticism at the extent to which this is possible (Avello, 2012).
- *Identifying key electoral issues*: Electoral tweets can be analyzed to determine the extent to which voters are concerned about particular issues. For example, does the electorate value economic development much more than environment protection, and to what extent? Other related problems include identifying contentious issues (Maynard & Funk, 2011) and detecting voter polarization (Conover et al., 2011).
- *Impact of fake tweets*: Often during elections there is an increase of artificially generated tweets from twitterbots, botnets, and sock-puppets. Understanding the impact of these tweets on public sentiment and automatic methods to filter out such tweets are both important research problems.

**Contributions of this work:** Traditional information retrieval systems usually identify facts such as what a person is doing, at what time, in what location, etc. In this paper we analyze electoral tweets for more subtly expressed information such as sentiment (positive or negative), the emotion (joy, sadness, anger, etc.), the purpose or intent behind the tweet (to point out a mistake, to support, to ridicule, etc.), and the style of the tweet (simple statement, sarcasm, hyperbole, etc.). To our knowledge, this is the first tweets dataset annotated for all of these phenomena. We also developed two automatic statistical systems that use the annotated data for training and predict emotion and purpose labels in new unseen tweets. These experiments establish baseline results for automatic systems on this new data.

*Data Annotation:* We designed two detailed online questionnaires and annotated the tweets by crowdsourcing to Amazon's Mechanical Turk.<sup>1</sup> We obtained over 100,000 responses from about 3000 annotators. We present an extensive analysis of the annotations which lend support to interesting conclusions such as electoral tweets almost always express the emotion of the tweeter as opposed to somebody else's, the predominant emotion in these tweets is disgust followed by trust, electoral tweets convey negative emotions twice as often as positive emotions, and that different intents of tweeting may be associated with the same emotion. All the data created as part of this project: about 100,000 responses to questions about emotions, purpose, and style in electoral tweets are made available:

<http://www.purl.org/net/PoliticalTweets2012>.

*Automatic Classifiers:* We developed a classifier for emotion detection that obtains an accuracy of 56.84%. We show how the stimulus identification task can be framed as a classification task that circumvents more complicated problems of detecting entity mentions and coreferences. On this stimulus classification task, our supervised classifier obtains an F-score of 58.30.

We show that emotion detection alone can fail to distinguish between several different types of purpose. For example, the same emotion of disgust can be associated with many different kinds of purpose such as 'to criticize', 'to vent', and 'to ridicule'. Thus, detecting purpose provides information that is not obtained simply by detecting sentiment or emotion. We developed a preliminary system that automatically classifies electoral tweets as per their purpose, using various features that have traditionally been used in tweet classification, such as word ngrams and elongated words, as well as features pertaining to eight basic emotions. We show that resources developed for emotion detection are also helpful for detecting purpose. We then add to this system features pertaining to hundreds of fine emotion categories. We show that these features lead to significant improvements in accuracy above and beyond those obtained by the competitive preliminary system. The system obtains an accuracy of 44.58% on a 11-class task and an accuracy of 73.91% on a 3-class task. The various emotion lexicons are made freely available.<sup>2</sup>

The rest of the paper is organized as follows. In Section 2, we present related work. Section 3 presents the data annotation step and also a detailed analysis of the annotations obtained. In Section 4, we describe an automatic classifier for detecting emotions (Section 4.1), an experiment showing that emotion detection although related to purpose detection is in fact a different task (Section 4.2), and finally a classifier for detecting purpose (Section 4.3). Section 5 presents conclusions and directions for future work.

## 2. Related work

Related work is organized into two sub-sections: (1) on annotating text for sentiment, emotion, style, and categories such as purpose, and (2) on automatic classifiers for detecting these categories.

<sup>1</sup> <https://www.mturk.com/mturk/welcome>.

<sup>2</sup> <http://www.purl.org/net/NRCEmotionLexicon>.

Download English Version:

<https://daneshyari.com/en/article/10355182>

Download Persian Version:

<https://daneshyari.com/article/10355182>

[Daneshyari.com](https://daneshyari.com)