Tracking Climate Change Opinions from Twitter Data

Xiaoran An
Dept. of Mech. & Indus. Eng.
Northeastern University
Boston, MA 02115
an.xi@husky.neu.edu

Auroop R. Ganguly
Dept. of Civil & Env. Eng.
Northeastern University
Boston, MA 02115
a.ganguly@neu.edu

Yi Fang
Dept. of Comp. Eng.
Santa Clara University
Stanta Clara, CA 95053
yfang@scu.edu

Steven B. Scyphers
Marine Science Center
Northeastern University
Nahant, MA 01908
s.scyphers@neu.edu

Ann M. Hunter
Behavioral Neuroscience
Northeastern University
Boston, MA 02115
hunter.ann@husky.neu.edu

Jennifer G. Dy
Dept. of Elec. & Comp. Eng.
Northeastern University
Boston, MA 02115
jdy@ece.neu.edu

ABSTRACT

The Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) released in 2014 states that warming of the climate system is "unequivocal," and it is "extremely likely" that human influence has been the dominant cause. However, public perceptions of anthropogenic climate change have varied widely, and indeed may have been significantly influenced by a disproportionate set of non-scientific opinion makers. While the statistics of extremes such as heat waves and heavy rainfall have been scientifically attributed to climate change, such attributions are not possible for single extreme events. Nevertheless, articles in social science and climate journals, including Nature Climate Change, have suggested that exposure to extreme weather events can directly influence opinions about climate change. Greenhouse-gas reduction policies, resilience to natural hazards, and adaptation decisions ultimately rely significantly on having adequate public support, but conducting real-time surveys of public perceptions can be difficult, expensive, and occasionally even impossible. The role of the micro-blogging site Twitter (http://twitter.com) has turned the Web into a major repository of topical comments, and hence a potential source of information for social science research. This paper attempts to understand whether Twitter data mining can complement and supplement insights about climate change perceptions, especially how such perceptions may change over time upon exposure to climate related hazards. A combination of techniques drawn from text mining, hierarchical sentiment analysis and time series methods is employed for this purpose. Future research is motivated in these areas, while potential pitfalls are discussed.

1. INTRODUCTION

Despite scientific consensus [14], climate change remains

a politically polarizing topic. A recent Op-Ed article [1] in the New York Times claimed: "Here's a scary fact about America: We're much more likely to believe that there are signs that aliens have visited Earth (77 percent) than that humans are causing climate change (44 percent)."

The IPCC's Special Report on Extremes [4] and the Fifth Assessment Report [14] attempted to relate the statistics of weather or hydrological extremes to human induced global warming. The current state of the science does not usually permit relating individual extremes to climate change. However, as indicated in Nature Climate Change [13], public opinions are known to be disproportionately influenced by exposure to extreme events. In addition, national and international policies, news events, and even email leaks appear to influence public perceptions.

Public perceptions of climate change have been tracked through carefully designed manual and localized surveys[13]. Here we ask the question whether mining the social media may provide an alternate source of tracking public opinions on climate change. While manual surveys will likely remain indispensable, automated or semi-automated surveys of the social media may offer complementary and supplementary benefits. The findings from social media surveys may help sharpen manual surveys, while the latter may lead to more focused surveys on the social media. In situations where manual surveys may be difficult, social media surveys may offer a first-order assessment of public sentiment.

The role of the micro-blogging site *Twitter* has turned the Web into a major repository of comments on many topics and a potential source of information for social science research. Twitter's core function allows users to post short messages, or *tweets*, which are up to 140 characters long and allows several ways for users to communicate with each other or express their opinions about a specific object. There is a significant prior literature which have attempted to relate Twitter sentiment insights to particular events [15, 9, 1]. The proof-of-concept results presented here is a step toward understanding the value of social media mining for tracking climate change opinions.

Several effort have been placed on detecting public perception on climate change [13, 7, 2, 3]. However, none of the previous work has utilized the widely available comment in-
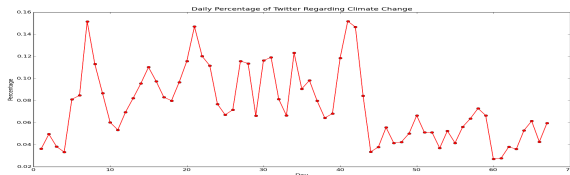
Figure 1: Daily Percentage of Tweets Regarding Climate Change

formation from social network and microblogging sites. Conducting studies based on surveys are limited as they can only collect a limited number of participants and may also be subject to survey bias. In this paper, we apply machine learning and data mining techniques to detect public sentiment on climate change, taking advantage of the freely and richly available text and opinion data from Twitter.

2. TWITTER DATA

The entire collection of data consists of $7,195,828$ Twitter messages posted by users between October, 3rd 2013 and December, 12th 2013 (excluding November, 21, 22, 23 and 24). The data is continuously collected using a Twitter Streaming API [2] with a Java application called Twitter Data Analytics [3] [6]. We then build the indexing and searching application with Java Package, Lucene [4] [10] to extract Twitter messages that are written in English and that also mention climate change. This querying method results in a highly significant set of tweets, but it is only a random subsample of all the posted content which are written in English and are related to climate change. Twitter allows users to re-tweet the tweets that they think might be interesting, this includes articles, news, and opinions. Note that our analysis is based on this climate Twitter collection excluding re-tweeted tweets. The reason for excluding re-tweets is because we assume that we cannot detect the sentiment of users by re-tweeting tweets of other users. However, to monitor the ratio of tweets regarding climate change to the daily collection of tweets, we included re-tweets in our count. A discussion is provided in the first subsection.

Climate Change Related Twitter with Re-Tweet.
The intention of this extraction is to try to explore the percentage of climate change related tweets per day. There are a total of $494,097$ tweets related to climate change in our collection, with $7,375$ climate change tweets daily on average. A plot of the percentage of tweets regarding climate change recorded daily is displayed in Figure 1. We observe that the percentages show high variability, and several major fluctuations are also detected. For example, on Day 40, 41 (i.e., November 11 and 12), the percentage goes up extremely high but then goes down significantly in 3 days. The occurrence of super Hurricane Haiyan seems to connect to this observation. We also observe that there is a significant increase on Day 21 (i.e., October 23). On that day, there was a series of devastating bushfire outbursts in Australia. There were several debates that went on about the connection between climate change and bushfires.

Climate Change Related Twitter without Re-Tweet.

Our major sentiment analysis is based on the tweets excluding re-tweeted tweets. This is because we assume the sentiment behind re-tweeted tweets is hard to detect and analyze. We extracted a total of $285,026$ tweets posted in English that are not re-tweeted. Although Twitter is a big collection of comment information and text data, a major disadvantage of Twitter data is that tweets do not have labels. Previous work on sentiment analysis with Twitter data employed emoticons such as ":-)" and ":-(" to form a training set for sentiment classification [12]. In our work, we manually labeled the Twitter data and classified them into *subjective* and *objective* groups. Within the *subjective* group, we further distinguish them into *positive* and *negative* classes. Subjective tweets mean that the tweets express users' opinions or emotions regarding climate change; whereas, objective tweets are normally news regarding climate change or the title of an article. We label the tweets which believe in climate change, are concerned about it, or express the opinion that actions need to be taken to mitigate climate change as *positive* ones. In contrary, the *negative* tweets include opinions that do not believe in climate change, and think it is just a scam. We have five people who worked on labeling the Twitter data, and choose the label which have at least three people who agree (i.e., we use the majority label). Overall, we have collected 1050 *objective* tweets and 1500 *subjective* tweets, and within the 1500 *subjective* tweets, we have 1000 *positive* tweets and 500 *negative* tweets.

3. APPROACH

In this paper, we analyze the data hierarchically, where we first apply subjectivity detection to distinguish *subjective* tweets from the *objective* ones in the entire corpus and then perform sentiment analysis only within the *subjective* tweets.

We represent each tweet with a bag-of-words representation. Because each tweet is short, we use binary word indicators as our feature representation. We pre-process our data as follows: We lowercased all letters (strip casing off all words), tokenized (convert the string to a list of tokens based on whitespace and remove punctuation marks), removed rare words ([5] suggests that words occurring two or fewer times may be removed, since these words are unlikely to be present to aid in future classifications), removed stopwords and frequent words, and reduced each word to its stem (removing any prefixes and suffixes).
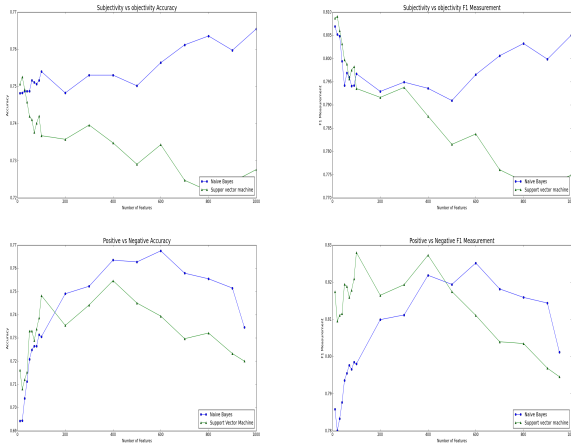
We explored two classification methods for sentiment text classification: Naive Bayes [11] and Support Vector Machines (SVMs) [16]. Naive Bayes is a generative classifier, whereas a support vector machine is an example of a discriminative classifier. We chose Naive Bayes and SVM in this study because both have been proven to perform well on text classification tasks. In addition to these two methods, we also performed feature selection on our Twitter data. We found that feature selection is important because each tweet is typically very short, where each message is not allowed to exceed 140 characters, making a bag-of-word feature representation (with dimensionality equal to the number of words in the Twitter dictionary) for each sample tweet to be very sparse.

Feature Selection. We initially have $D = 1300$ features (words). Not all of these features will be important for the classification task. Furthermore, our problem is quite sparse; even sparser than typical document classification tasks. Thus, feature selection will be helpful. Feature

---

[2]https://dev.twitter.com/docs/streaming-apis

[3]http://tweettracker.fulton.asu.edu/tda/

[4]http://lucene.apache.org/

Figure 2: 10-Fold Cross-Validation for Each Method, Varying the Number of Features

selection algorithms are defined by the criterion utilized for evaluating features and the search strategy. Searching all 2^D possible feature subsets is intractable. Here, we apply a simple search strategy by simply scoring each feature individually. There are numerous ways of evaluating or scoring features. [5] compares various feature selection metrics and their impact on the performance of classifiers. In our work, we use the chi-squared metric, which is a common statistical test that measures divergence from the distribution expected if one assumes the feature occurrence is actually independent of the class value. The formulation of the chi-squared measure is: $X^2(D, f, c) = \sum_{e_f \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_f e_c} - E_{e_f e_c})^2}{E_{e_f e_c}}$, where $e_f = 1$ means the document contains term f, and $e_f = 0$ means the document does not contain term f. $e_c = 1$ represents the document is in class c and $e_c = 0$ represents the document is not in class c. E is the expected frequency when the assumption that the presence of feature f and class c is independent is satisfied. Higher value of X^2 indicates that the hypothesis of independence is incorrect. We then rank order the features based on this score.

To determine the model order, meaning the number of features to keep, we measure the classification performance on a held-out validation set. We use both *macro* F1 measure and accuracy as performance measures. F1 measure is defined as $F_1 = \frac{2 \times recall \times precision}{(recall + precision)}$, where recall measures the ratio of the number of true positive cases to the number of all observed positive cases, and precision measures the ratio of the number of true positive cases to the number of all predicted positive cases by classifier. There are two methods for averaging the F-measure over a collection of 2-class classification problems. One is the *macro* averaged F-measure, which is the traditional arithmetic mean of the F-measure computed for each problem. Another is the *micro* averaged F-measure, which is an average weighted by the class distribution. Since we are interested in average performance across different classifications, so we focus on *macro* averaged F-measure.

4. SENTIMENT ANALYSIS

In this section, we report the results on our sentiment analysis on mining the Twitter climate change data. We first randomly select one-fifth of entire labeled tweets as valida-

Table 1: Candidate Models

|	Algorithms	No. features	Accuracy	F_1 measure
Subjectivity	Naive Bayes	1000	0.7654	0.8057
Subjectivity	SVM	20	0.7538	0.8090
Subjectivity	SVM	30	0.7596	0.8126
Polarity	Naive Bayes	400	0.7677	0.8278
Polarity	Naive Bayes	600	0.7645	0.8215
Polarity	SVM	100	0.7548	0.8382
Polarity	SVM	400	0.7581	0.8292

tion set. There are 210 objective and 310 subjective tweets and within the subjective tweets, there are 210 positive and 100 negative tweets. The rest of the four-fifth of entire labeled tweets become our training data set which consists of 840 objective and 1190 subjective tweets, and 790 positive and 400 negative tweets. We perform 10-fold cross-validation on the training data set to train our model and choose the best model by comparing the performance on the validation set. We use the default settings of parameters for Naive Bayes and Support Vector Machine in the scikit-learn Python package (http://scikit-learn.org/stable/).

Model Selection.

We perform feature selection for both SVM and Naive Bayes classifiers. We rank ordered the features based on the chi-squared scoring described in the previous section and evaluated the performance of these two classifiers for varying number of features and evaluated the classifiers' performance on both tasks, *subjective* vs *objective* and *positive* vs *negative*, based on accuracy and F1-measure using 10-fold cross-validation on the training set. The results are shown in Figure 2. The performances of both algorithms vary significantly with different number of features. As the feature size increases, both methods have serious over-fitting problem. The reason for this observation is that: firstly, tweets are relatively short compared to other documents leading to very sparse feature vectors in high dimension; secondly, the training data size is relatively limited for such high dimension. The result confirms our premise of the importance of feature selection. With small number of features, the two algorithms perform well in predicting sentiment. After careful comparison of the result, we selected a few set of candidate models to compare. We report the performance results of these candidate models on our held-out validation set in Table 4. For the subjectivity detection task, Naive Bayes has the highest accuracy performance using 1000 features compared to the SVM results. But the F1 measure of this Naive Bayes classifier is slightly less than that of SVM. Because, the performance of both classifiers are almost the same, we prefer to use SVM because it uses much fewer features to avoid the over-fitting problem. We chose SVM with 30 features compared to 20 features because it has better performance on both accuracy and F1 measure. Similarly, for the polarity sentiment task, we select SVM with 100 features. Because their performance are almost similar, we pick the model with fewer features.

Prediction and Event Detection.

With the selected subjectivity detection and sentiment polarity algorithm, we extract the *subjective* tweets from our entire climate change related tweets which have been divided into subgroups based on day. We, then, predict the sentiment *polarity* on the subjective tweets as reported daily to calculate the percentage of *positive* and *negative* sentiments. The daily percentage of subjective and objective tweets are shown in Figure 3 (top), and the percentage of *positive* and
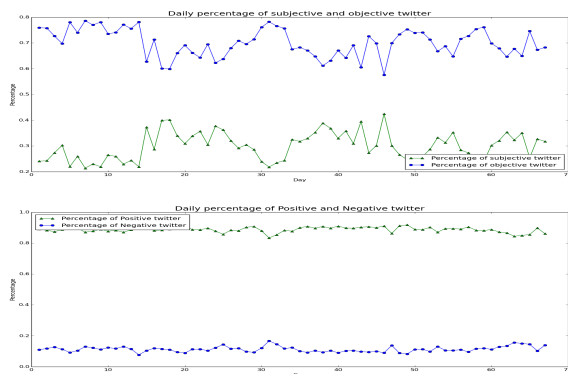
Figure 3: Subjectivity Detection and Sentiment Polarity Prediction

negative tweets are shown in Figure 3 (bottom). The day presented in the graph is from October 3, 2013 to December 12, 2013 (excluding November, 21, 22, 23 and 24).

The *subjective* and *objective* percentages present large variability as we move along the time axis. This variability is influenced by many factors, such as the news, articles published on that specific day or the occurrence of any event. Because of these confounding factors, it is not easy to detect any major change or event using the *subjective* and *objective* percentages. It would be quite beneficial to climate sentiment studies if we can detect whether the sudden change in Twitter sentiment regarding climate change are related to major climate events or extreme weather conditions. We, thus, focus on the sentiment polarity percentages. We analyze the sentiment polarity percentage trend by tracking the mean and standard deviation calculated from a fixed-size sliding window for each time point, and plot the z-score normalization as a function of time [8]. In [8], they calculate the z-score normalization for each of the six normalized moods scores, which are in the range of $[0, 1]$, from POMS scoring (Profile of Mood States, a well established psychometric instrument). The z-score normalization can be calculated as follows: $m_z = \frac{m - \bar{x}(\theta[i, \pm k])}{\sigma(\theta[i, \pm k])}$, where $\bar{x}(\theta[i, \pm k])$ and $\sigma(\theta[i, \pm k])$ represent the mean and standard deviation of the time series within the local $[i, \pm k]$ m-neighborhood for a specific day. m is a normalized mood score.

In this work we have the negative percentage data expanded for 67 days, we consider it as the normalized negative mood score m in the above approach. For example, higher percent of negative sentiment for a specific day represents the higher mood score for the negative sentiment. Using a sliding window size of 7 days (3 days before and 3 days after), we can derive the z-score normalization to detect short-term fluctuations of public negative sentiment as a result of particular short-term events. The result is shown in Figure 4. We are interested in looking at the point which has a z-score close to or near ± 2. We observe that we can relate several climate change related events to the sudden fluctuation of negative sentiment. On Nov. 1st, because of President Obama's Executive order on climate change, the significant increase in negative sentiment about climate change (day 31) can be detected from the graph. Decrease in negative sentiment about climate change can be observed on Nov. 11th (day 40). This is probably because of the occurrence of the destructive Typhoon Haiyan in the Philippines.
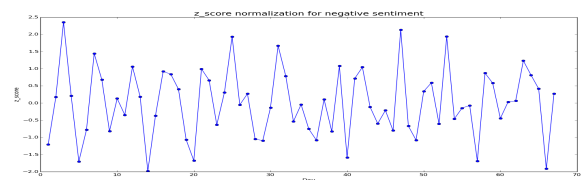
Figure 4: z-score Normalization for Negative Sentiment

The observation indicates that Twitter users express their concerns to the extreme weather. We also detect there is a decrease in negative sentiment on Oct. 29th (day 28), when it was the anniversary of super storm Sandy. On Oct. 22 (day 20), a statistically significant decrease in negative sentiment is observed. This may be related to the series of bushfire outbursts in Australia. Hence, the change in climate change related sentiment polarity is somehow related to major climate change events.

From the sentiment polarity analysis based on climate change related tweets, an average of more than 80% of tweets believe in climate change in our data collection, this can be observed from Figure 3. Compared to the majority *positive* sentiment tweets, there is only a small percentage of tweets that express doubt regarding climate change. This unintuitive result indicates that the majority of Twitter users in our collection think climate change is happening and believe that action is needed to mitigate it.

5. CONCLUSION

Traditionally, the attitudes, knowledge, and opinions of citizens and key decision-makers have been studied through relatively expensive and logistically challenging survey techniques, but more recently scientists and many other groups have begun to exploit the vast amounts of information available in social media platforms. This paper presents proof of concept results to suggest that mining social media data, exemplified here through Twitter accounts, can be a valuable way to yield insights on climate change opinions and societal response to extreme events.

Our work points to new opportunities for leveraging and developing knowledge discovery methods, such as opinion mining and time series techniques using social media platforms, for social science research and informing urgent societal priorities. Specifically, we used classical sentiment analysis algorithms in detecting and tracking opinions regarding climate change from Twitter feeds. In addition to measuring overall patterns and trends in climate-related sentiments, we detected a connection between short-term fluctuations in negative sentiments and major climate events. We found that major climate events can have a result in sudden change in sentiment polarity, but considering the variation in sentiment polarity shows that there is still significant uncertainty in overall sentiment. We used Twitter data to illustrate how the opinions of Twitter users can change over time and in the aftermath of specific events, but similar approaches may be extended to other publicly available information and social media platforms. While Twitter users may not represent all social groups, its large and rapidly growing popularity supports information exchange among roughly 50 million U.S. citizens (̃250 M users globally), government agencies, political leaders, activist organizations, and other influential opinion-makers.

6. REFERENCES

[1] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, March' 2011.

[2] S. Brody, S. Zahran, A. Vedlitz, and H. Grover. Examing the relationship between physical vulnerability and public perceptions of global climate change in the united states. *Environment and behavior*, 40(1):72–95, 2008.

[3] L. Feldman, E. Maibach, C. Roser-Renouf, and A. Leiserowitz. Climate on cable: The nature and impact of global warming coverage on fox news cnn and msnbc. *International Journal of Press/Politics*, 17(1):3–31, 2011.

[4] Field, C. B., D. Q. V. Barros, T.F. Stocker, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G. k. Plattner, S. Allen, M. Tignor, and P. M. (eds.). *IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2012.

[5] G. Forman. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 3:1289–1305, March 2003.

[6] S. Kumar, F. Morstatter, and H. Liu. *Twitter data Analytics*. Springer, 2013.

[7] I. Lorenzoni and N. Pidgeon. Public views on climate change: European and usa perspectives. *Climate Change*, pages 73–95, 2006.

[8] H. Mao, A. Pepe, and J. Bollen. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, 2011.

[9] A. Marcus, M. Bernstein, O. Badar, D. Karger, S. Madden, and R. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 227–236, 2011.

[10] M. McCandless, E. Hatcher, and O. Gespodnetic. *Lucene in Action, Second Edition*. Manning, 2010.

[11] K. Nigam, A. McCallum, S. Thrun, and T. M. Mitchell. Learning to classify text from labeled and unlabeled documents. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference (AAAI/IAAI)*, pages 792–799, 1998.

[12] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, May 2010.

[13] A. Spence, W. Poortinga, C. Butler, and P. N. F. Perceptions of climate change and willingness to save energy related to flood experience. *Nature Climate Change*, 1:46–49, 2011.

[14] Stocker, T. F., D. Qin, G. K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. (eds.). *IPCC, 2013: Climate Change 2013: The Pyhsical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[15] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment in twitter events. *American Society for Information Science and Technology*, 62(2):406–418, February 2011.

[16] V. Vapnik. *The nature of statistical learning theory*. Springer, New York, 1995.

Acknowledgments