Understanding Social Media's Take on Climate Change through Large-Scale Analysis of Targeted Opinions and Emotions

Neetu Pathak[1], **Michael Henry**[1], and **Svitlana Volkova**[2]
[1]Visual Analytics Group, [2]Data Sciences and Analytics Group
Pacific Northwest National Laboratory
902 Battelle Blvd, Richland, WA 99354
neetu.pathak, michael.j.henry, svitlana.volkova@pnnl.gov

Abstract

Social media provides a powerful data source for researchers interested in understanding population-level behavior, having been successfully leveraged in a number of different application areas including disease prediction models, detecting civil unrest, and measuring public sentiment towards a given topic of interest within the public discourse. In this work, we present a study of a large collection of Twitter data centered on the social conversation around global climate change during the 2015 UN Climate Change Conference, held in Paris, France in December 2015. We first developed a mechanism for distinguishing between personal and non-personal accounts. We then analyzed demographics, emotion, and opinion dynamics over time and location in order to understand how the different user types converse around meaningful topics on social media. This methodology offers in-depth insight into the behavior and opinions around a topic where multiple distinct narratives are present involved, and lays the groundwork for future work in studying narratives in social media.

Introduction

The analysis of social media data can provide real-time insights into world events by providing a platform for users to express their opinions, report on newsworthy events, or generally discuss topics of interest. For this study, we focused on Twitter,[1] which is a open platform microblogging social networking site. Many researchers have successfully incorporated Twitter data in a number of research areas including flu prediction (Achrekar et al. 2011), political events (Wang et al. 2012a), and civil unrest (Ramakrishnan et al. 2014).

Building upon prior research on the analytics of Twitter data, we employed sentiment and demographic analysis in order to understand the social conversation around a complex, global topic: global climate change. Whereas previous social media analysis research has focused primarily on predicting behavior of interest (Achrekar et al. 2011), analyzing a single characteristic of the data (Wang et al. 2012a), or incorporating social media features into a larger analytic system (Sadilek et al. 2013), our work is focused on describing a wholestic characterization of the complimentary and diversionary narratives present in the social conversation surrounding global climate change.

The main goal of this study is to perform a large-scale analysis of opinion and emotion variations expressed in social media around topics related to COP21 This work seeked to answer the following research questions:

- Can we identify climate change relevant discourse in social media?
- What are the most influential non-personal accounts that communicate topics about climate in social media?
- Do personal vs. non-personal users express different opinions and emotions toward climate change on Twitter?
- How do user demographics e.g., gender, age, and income influence opinions and emotions they express about climate change on Twitter?

Related Work

Social media platforms—such as Twitter, Facebook, Google+, Instagram, and Flickr—provide researchers with large amounts of data that is timely, multilingual and diverse. Most of the research in analysis of social media analytics falls into one of three main categories: opinion mining with respect to understanding human behavior, grouping user communities based a criteria such as their inferred demographics, and sentiment and emotion classification. There is also research that utilizes social media data to describe a specific use case, one of which is global climate change.

Opinion Mining in Social Media Social media has been used to understand human behavior across many domains including healthcare[2] (Achrekar et al. 2011; De Choudhury, Counts, and Horvitz 2013; Sadilek et al. 2013), politics (Wang et al. 2012a; Bakliwal et al. 2013), and crisis management (Asur and Huberman 2010; Rui, Liu, and Whinston 2013; Nichols, Mahmud, and Drews 2012; Yu and Wang 2015; Bollen, Mao, and Zeng 2011).

Signals drawn from social media have been used for sentiment analysis of political discourse (Wang et al. 2012a; Bakliwal et al. 2013), sports events (Yu and Wang 2015), and real-time crisis management (MacEachren et al. 2011). Twitter data has also been applied to predicting stock mar-

[1]https://www.twitter.com

[2]https://www.google.org/flutrends

ket (Bollen, Mao, and Zeng 2011) and movie box-office returns (Asur and Huberman 2010).

User Demographic Prediction Many approaches have been developed to classify Twitter accounts into different groups to understand how these groups react toward a specific topic of interest. For example, researchers classified Twitter accounts as those that belong to organizations vs. people to improve event recognition models (De Silva and Riloff 2014). Others focused on developing a model to distinguish between personal vs. non-personal accounts, and found the latter are more connected in a social network (McCorriston, Jurgens, and Ruths 2015).

Prior work has been conducted in developing models to predict user income based on language and communication behavior in social media (Preoţiuc-Pietro et al. 2015). Others focused on analyzing correlations between user demographics and emotions on Twitter (Volkova and Bachrach 2015). They showed that many perceived demographic traits correlate with the emotional contrast between users and their neighbors in social media, and that user emotions can be effectively used to improve demographic prediction (Volkova and Bachrach 2016). Earlier work presented language-independent models for predicting gender of Twitter users. Their approach relied on tweets and other meta-data e.g., screen-name, full-name, and biography field (Burger et al. 2011). In addition to using language features, other approaches explored combining user and neighbor features to predict three attributes – gender, age, and political orientation of Twitter users (Zamal, Liu, and Ruths 2012; Volkova, Coppersmith, and Van Durme 2014). Researchers relied on user interests to classify a variety of latent user properties (Volkova, Bachrach, and Van Durme 2016), as well as user language to predict latent sociodemographic traits including age, gender, income, education, relationship status, optimism, and life satisfaction (Volkova et al. 2015a).

Emotions and Sentiment Classification Sentiment classification in social media has been extensively studied (Pang, Lee, and Vaithyanathan 2002; Pang and Lee 2008; Pak and Paroubek 2010; Nakov et al. 2013; Zhu, Kiritchenko, and Mohammad 2014). Emotion analysis has been successfully applied to informal and short communications including emails, blogs (Kosinski, Stillwell, and Graepel 2013), and news headlines (Strapparava and Mihalcea 2007). However, fine-grained emotions in social media, including Twitter, have only been investigated recently (Volkova et al. 2015b).

To detect emotions and opinions in social media researchers have used supervised models trained on text, emoticons, topics, and lexicons (Wang et al. 2012b; Roberts et al. 2012; Qadir and Riloff 2013). Most of this line of work focused on capturing six high-level emotions proposed by Ekman–joy, anger, sadness, fear, disgust, and surprise (Ekman 1992). Other papers studied moods, including tension, depression, fatigue, and issues such as politeness, rudeness, embarrassment, and formality (De Choudhury, Gamon, and Counts 2012; Esmin, de Oliveira, and Matwin 2012).

Climate Change in Social Media There has been an emerging interest in analyzing human behavior toward climate change through social media. In 2014, the United Nations Global Pulse[3] collected tweets related to climate change between April and December 2014 and categorized them into nine topics of interest (General, Politics/Opinion, Weather, Economy, Risk/Disaster, Energy, Agriculture/Forestry, Artic, and Oceans/Water.) Related research focused on classifying tweets related to climate change as subjective or objective, and tracked sentiment dynamics over time (An et al. 2014). More recent work analyzed social network structure around climate change communications and identified the most influential accounts that share information about climate change on Twitter.[4]

While these recent studies analyzed factors related to climate change using social media, there is no work that focused on studying variations in fine-grained emotion and opinion dynamics toward specific climate change topics nor any that measured the influence of user demographics on expressing targeted opinions, or contrasted affects expressed by personal vs. non-personal accounts.

This work extends prior research on the polarity of tweets relevant to climate change by focusing on the differences in fine-grained emotions expressed toward topics on climate change, and analyzing the influence of user demographics on affects expressed toward climate topics on Twitter.

Data and Methodology

In this section we discuss the data used in this study, as well as the methodology by which we were able to infer topics of interest related to climate change, account type classification, and demographic and affect classification.

Twitter Dataset The 21st Conference of the Parties of the United Nations Climate Change Conference (also referred to as COP21) took place from 30 November, 2015 through 12 December, 2015 at Le Bourget in Paris, France. It was during this conference that the 197 Parties to the United Nations Framework Convention on Climate Change met to negotiate the Paris Agreement—the globally-negotiated agreement which outlined a plan for limiting carbon emissions with the goal of capping global warming increases to a less than 2°C.

We acquired our data by filtering from Twitter all the tweets that matched the terms, hashtags, and user accounts specified in Table 1. The terms, hashtags, and users selected as filters were determined by an empirical analysis of a sample of Twitter data regarding the topic of climate change. In addition to terms, hashtags, and user accounts, all data that was geotagged as originating from Paris, France during that time period were acquired as well. This curated dataset contains over 8M tweets, over 4.5M of which are in English.

Climate Change Topic Identification To analyze targeted opinions on specific topics relevant to climate change, we categorized Twitter discourse into nine categories. Five of those nine categories were drawn from UN Global Pulse[5]

[3] http://unglobalpulse.net/climate/

[4] https://www.carbonbrief.org/mapped-the-climate-change-conversation-on-twitter

[5] http://unglobalpulse.net/climate/

Table 1: Terms, hashtags, and account names used to query Twitter to acquire the data (in addition to tweets geotagged as originating in Paris, France).

Terms
cap trade, global climate change, emissions trading, carbon tax, alarmism, global warming, climate change, emissions, environmental defense, EPA, climate, COP21, le bourget, CFACT, warmism, paris pledge for action
Hashtags
#COP21 , #alarmism, #climat, #cities4climate, #climateaction, #climatechange , #gmo, #renewableenergy, #earthtweet, #climatedeniers, #foodjustice, #green, #climatehustle, #sustainable, #4change, #cleantech, #climatedenier, #windenergy, #windenergy, #globalwarming, #warmism, #parispledgeforaction
Accounts
@cop21, @CFACT, @ccnucc

– Energy, Weather, Economy, Agriculture/Forestry, and Water. We added four more categories—Security, Climate Denial, Air Issues, and Animals—to cover a broader range of climate-related topics than the other four used by UN Global Pulse, based on an empirical examination of our data set.

To categorize social media discourse into climate-related topics we empirically determined a list of seed words for every category as shown in italics in Table 2. We then take advantage of compositional distributional semantics and discover new terms and hashtags[6] that appear in similar contexts with the seed keywords using Word2Vec model (Mikolov et al. 2013) implemented in gensim (Řehůřek and Sojka 2010).

Account Type Classification To better understand human behavior through social media it has been shown (McCorriston, Jurgens, and Ruths 2015) to be benefitical to distinguish between personal and non-personal accounts (such as communities, groups, organizations, and influential users).

Following recent practices on account type prediction (McCorriston, Jurgens, and Ruths 2015), our approach relies on a follower-to-friend ratio (F2F), the number of tweets per day (TPD), the number of listed count (LC) (the number of times the account has been listed by others[7], and if the account is a verified account. We take the majority voting from the statistics above (giving higher weights to verified and listed accounts) and add several constraints, e.g., F2F $\geq$ 200, TPD $\geq$ 70, LC $\geq$ 150. After manual inspection of the data we discovered that accounts that have higher follower-to-friend ratio and the number of tweets per day are typically non-personal accounts (McCorriston, Jurgens, and Ruths 2015). Overall, we have 13,439 (less than 1%) non-personal accounts with 185,930 tweets and 1,865,743 personal accounts (99.28%) with 8,781,470 tweets. 12,198 (90.76%) out of 13,439 non-personal accounts have URLs attached to their profiles, and 9,643 (71.75%) have the geolocation or timezone listed.

[6]The list of hashtags relevant to climate change is available upon request.

[7]https://gigaom.com/2009/11/02/twitters-listed-stat-is-it-a-measure-of-influence/

Table 2: Climate change topics (sorted by popularity) with seed keywords (highlighted in italic) and a sample of discovered hashtags in our dataset.

Weather (955,577)
heatwaves, risingseas, flood, globalwarming, bashfire, climatemigrant, overheating, heatwave
Energy (719,023)
energy, solarenergy, renewableenergy, energyfuture, energyhour, solarenergy, solar, protectdefendrenew, ener, ene
Air Issues (207,445)
carbon, carbonprice, ozone, priceoncarbon, carbonfarming, hfcs, montrealprotocol, slcps, emis, cleanair
Security (160,852)
nationalsecurity, terrorism, gunsense, guncontrol, gunviolence, natsec, security, incomeinequality, publichealth, dod
Economy (152,737)
money, economy, capital, economy, followthemoney, cleaneconomy, lowcarboneconomy, circulareconomy
Food (Food/Agriculture/Livestock) (116,883)
food, gmo, livestock, agriculture, safefood, lessmeatlessheat, livestock, gmos, genetically, gmosalmon, glyphosphate
Animals (63,873)
polarbears, habitat, hummingbird, protectpolarbears, tuna, dolphins, whales, lizards, turtles, corals, endangeredspecies
Water (Water/Ocean/Arctic) (37,995)
water, ocean, arctic, sea, oceanlife, oceans, oceansday, oceancop, oceanoptimism, oceanacidification, oceanforclimate
Climate change denial (24,780)
climatedeniers, climatehustle, climatedeniers, globalwarminghoax, globalwarmingisahoax, climatechangefraud

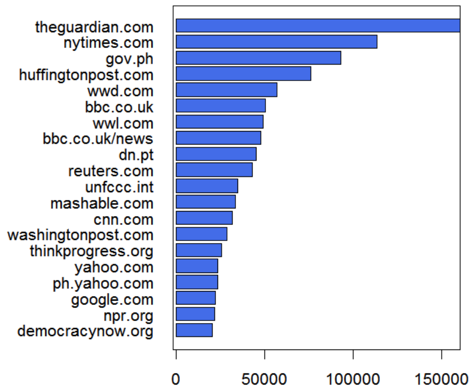
Demographic and Affect Classification Unlike Facebook (Bachrach et al. 2012; Kosinski, Stillwell, and Graepel 2013), Twitter profiles do not require personal demographic information to be attached to the profile. Thus, to infer demographics for personal accounts in our dataset we relied on pre-trained models learned from 5,000 user profiles annotated via crowdsourcing[8] released by (Volkova and Bachrach 2015). We annotated 1,354,918 user profiles with three sociodemographic attributes—gender, age, and income. We only used a subset of models trained on annotations with high or moderate inter-annotator agreement.

For annotating all tweets (personal and non-personal) with emotions and sentiments we used publicly available affect models.[8] These models rely on lexical (word ngrams), syntactic, and stylistic (e.g., elongations, positive and negative emoticons, hashtags, punctuation, and negation) features. We used their sentiment and emotion classifiers to label 5,681,301 tweets with one of three sentiment categories—positive, negative, or neutral, and one of six emotion categories—joy, fear, sadness, surprise, anger, and disgust.

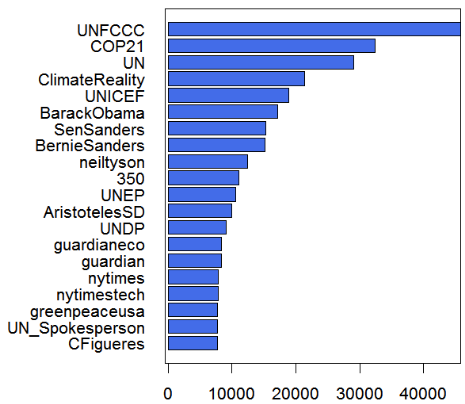
Analysis and Results

This section reports our main findings on estimating account influence, opinion and emotion dynamics toward targeted topics, the differences in affects expressed toward climate

[8]Data collection and perceived attribute annotation details are discussed in (Volkova and Bachrach 2015).



(a) Shared URLs



(b) Retweets

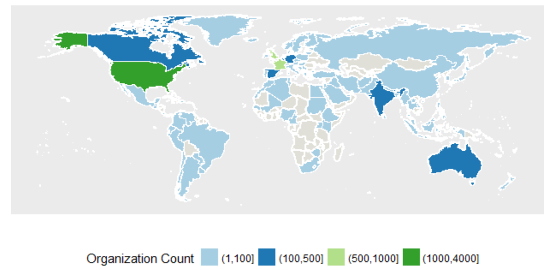
Figure 1: Estimating influence on non-personal Twitter accounts (top 20 accounts shown) using shared URLs and retweets on climate change topics(whole Dataset).

change issues by personal and non-personal accounts, and emotion and opinion variations regarding climate change given user demographics.

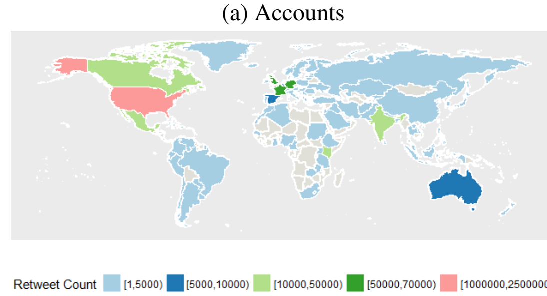
Non-Personal Account Influence In our dataset we labeled 13,439 (0.71%) Twitter accounts as influential accounts. The example influential accounts include *bbc, nytimestech, and BarackObama*. We grouped influential accounts using the URL domain names extracted from the profile information e.g., *nytimetech, nytimesbusiness, nyopinion* are all representative of the *nytimes* domain.

We estimate non-personal account influence when it comes to communicating climate change topics using shared URLs and retweets. Figure 1a presents the frequency of the top mentioned shared URLs (aggregated by domain names) in our data sample.

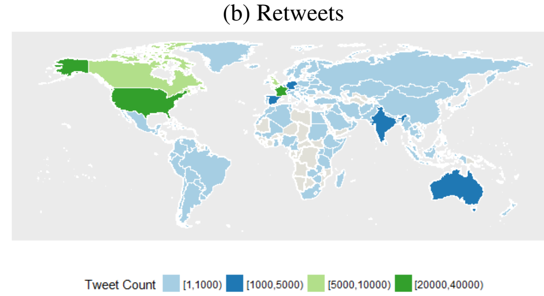
In addition, we aggregate non-personal accounts by geo-location and report the number of accounts per location, as well the number of tweets and retweets about climate change per geo-location as shown below. Figure 2a shows the distribution of the number of non-personal accounts per country. We observe that the majority of organizational accounts are located in USA, France, UK, Canada, Australia, Germany, Spain, Belgium, India, and Netherlands.



(a) Accounts



(b) Retweets



(c) Tweets

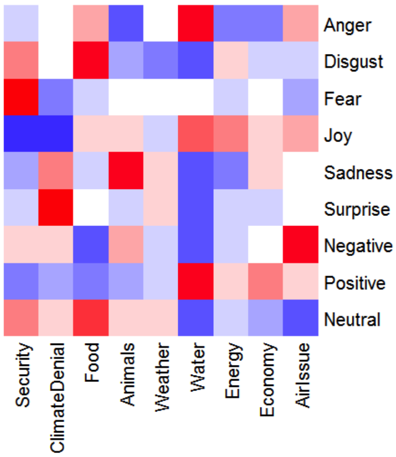
Figure 2: The distribution of the number of non-personal accounts, retweets, and tweets produced by non-personal accounts per geo-location (full dataset).

Figure 2b presents the distribution of retweeted non-personal accounts per country. We found that the top 10 countries with the highest number of retweeted accounts are USA, Germany, France, UK, Lebanon, Canada, Mexico, India, Kenya, and Belgium. Its interesting to note that Lebanon, Mexico, and Kenya replace Australia, Spain and Netherlands from Figure 2a.

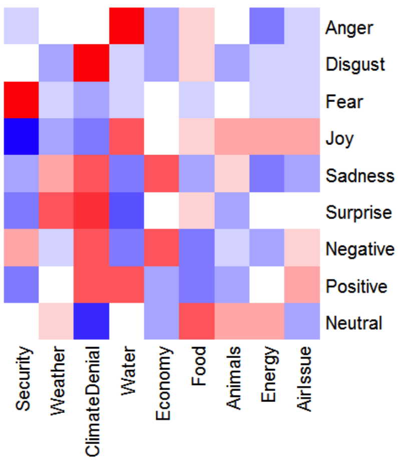
In Figure 2c we show the distribution of the number of tweets on climate change produced by non-personal accounts per country. We observe that the top countries with the highest number of tweets on climate change include USA, France, UK, Canada, Germany, Belgium, Australia, India, and Spain.

Opinion and Emotion Dynamics Over Time Figure 5 presents opinion and emotion dynamics over 31 day period in December 2015. We observe several spikes in affects dynamics, for example more:

- Anger tweets on Dec 3rd and 4th about *Water*.
- Disgust tweets on *Climate Denial* on Dec 23rd and 24th.
- Joy tweets toward *Water* on Dec 25th – 27th.
- Sadness tweets toward *Economy* on Dec 4th.
- Surprise tweets toward *Weather* on Dec 15th.



(a) Non-Personal



(b) Personal

Figure 3: Non-personal vs. personal account emotion and sentiment variations toward climate change topics.

Non-Personal Account Affect Variations toward Climate Topics Figure 3a describes emotion and opinion proportions expressed by non-personal accounts toward climate change topics. We observe that non-personal accounts e.g., organizations and influential users express the most anger and joy tweets toward *Water*, the most disgust tweets toward *Food*, the most fear tweets toward *Security*, the most sadness tweets toward *Animal*, the most surprise tweets toward *Climate Denial*, the most negative opinions toward *Air Issues*, the most positive opinions toward *Water* and the most neutral opinions toward *Food*. Similarly, we found that non-personal accounts generate the least anger tweets toward *Animals*, the least disgust tweets toward *Water*, the least fear tweets toward *Climate Denial*, the least joy tweets toward *Climate Denial* and *Security*, the least sadness and surprise tweets toward *Water*, the least negative opinions toward *Food* and *Water*, the least positive opinions toward *Food* and *Security*, and least neutral opinions toward *Water* and *Air Issue*.

Personal Account Affect Variations toward Climate Topics Figure 3b describes emotion and opinion proportions expressed by users toward climate change topics. We observe that people express more anger and joy tweets toward *Water*, more disgust and surprise tweets toward *Climate Denial*, more fear tweets toward *Security*, more sadness tweets towards *Economy* and *Climate Denial*, more negative opinions toward *Economy* and *Climate Denial*, more positive opinions toward *Climate Denial* and *Water*, more neutral opinions toward *Food*. Similarly, personal accounts generate the least anger tweets toward *Energy*, the least disgust tweets toward *Animals*, *Economy* and *Weather*, the least fear tweets toward *Climate Denial*, the least joy tweets toward *Security*, the least sadness towards *Energy*, the least surprise tweets toward *Water*, the least negative opinions towards *Food* and *Water*, the least positive opinions toward *Food* and *Security*, and the least neutral opinions toward *Climate Denial*.

The Influence of User Demographics on Opinions and Emotions toward Climate Topics Figure 4 demonstrates variations in the amount of communications about different climate change topics given user demographics. We found that *Weather* is most discussed by older accounts, *Climate Denial* is most discussed by higher income people, *Water, Animals, and Food* are most discussed by females, *Air Issue ,* *Economy,* and *Energy* is most discussed by younger, male, and lower income people. *Security* is most discussed by lower income people. We also found that *Weather* is least discussed by younger, male, and lower income people, *Climate Denial* is least discussed by females, *Water, Animals, Food, Energy,* and *Economy* are least discussed by older people, *Air issue* is least discussed by females, and *Security* is least discussed by older people.

The difference in emotion intensity shown by person and non-person accounts (only English tweets) Figure 6 describes difference in emotion and opinion proportions expressed by person and non person users toward climate change topics. We observe that difference is prominent while expressing anger toward *Air Issue*, disgust *Energyl* and *Economy*, fear toward *Weather* and *Water*, joy towards *Weather* and *Economy*, sadness towards *Food*, *Air Issue*, *Animals*, *Security*,*energy* and *Water*, surprise towards *Climate Denial*, more negative toward *Animal* and *Air Issue*, more positive toward *Economy* and *Water*, more neutral opinions toward *Climate Denial*. Similarly, least difference in emo-

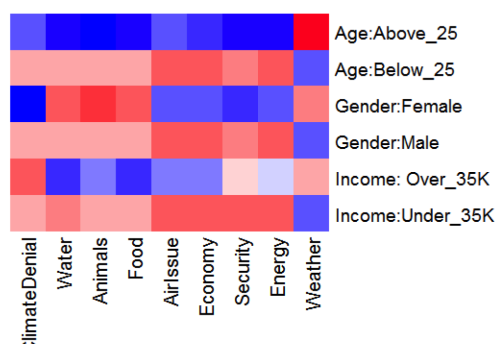


Figure 4: The influence of user demographics on opinions and emotions toward climate topics (only English tweets).

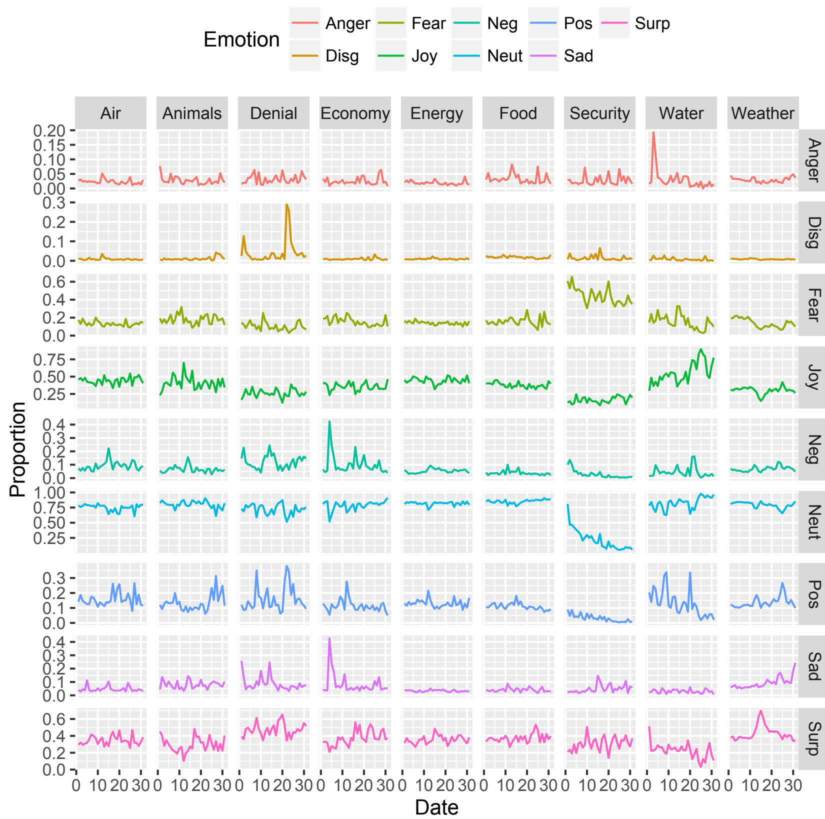


Figure 5: Sentiment and emotion dynamics toward climate change topics in December 2015 (only English tweets).

tion proportion which mean they show similar emotion is anger tweets toward *Animal*, disgust tweets toward *Climat Denial*, fear tweets toward *Security*, joy tweets toward *Climate Denial*, sadness towards *Economy*, surprise tweets toward *Weather*, the least negative opinions towards *Economy*, positive opinions toward *Climate Denial*, and the least neutral opinions toward *Water*.

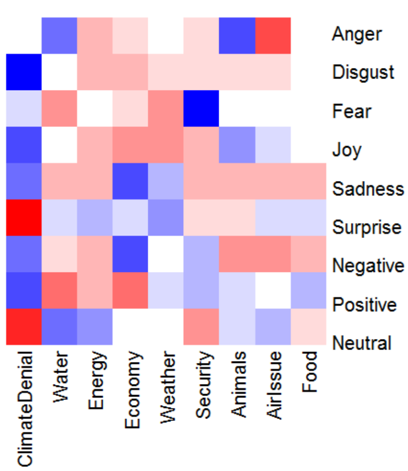


Figure 6: The difference in sentiment and emotion dynamics shown by person and non-person user toward specific climate change topics in December 2015 (only English tweets).

Summary

In this work we presented the results of a study of over 8M tweets collected from December 2015 centered around the topic of global climate change and the 21st Conference of the Parties of the United Nations Climate Change Conference (COP21). From this data, we were able to determine whether an account was a personal account, or a non-personal account by leveraging metadata regarding the message and the account. Following that, we were able to track the opinion and emotion dynamics over time, as well as the affect variations between personal and non-personal accounts towards climate-related topics. Furthermore, we were able to track the influence of user demographics on opinion and emotion expression towards climate topics, and the difference in emotion intensity shown by each account type. Further research will explore the influence of the images used and shared within a social network and between personal and non-personal accounts to better understand the diversity of opinions within a social conversation.

Acknowledgments

This research was performed at PNNL, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy.

References

Achrekar, H.; Gandhe, A.; Lazarus, R.; Yu, S.-H.; and Liu, B. 2011. Predicting flu trends using Twitter data. In *Proceedings of INFOCOM*, 702–707.

An, X.; Ganguly, A. R.; Fang, Y.; Scyphers, S. B.; Hunter, A. M.; and Dy, J. G. 2014. Tracking climate change opinions from Twitter data. In *Workshop on Data Science for Social Good*.

Asur, S., and Huberman, B. A. 2010. Predicting the future with social media. In *Proceedings of WI-IAT*, 492–499.

Bachrach, Y.; Kosinski, M.; Graepel, T.; Kohli, P.; and Stillwell, D. 2012. Personality and patterns of Facebook usage. In *Proceedings of ACM WebSci*, 24–32.

Bakliwal, A.; Foster, J.; van der Puil, J.; O'Brien, R.; Tounsi, L.; and Hughes, M. 2013. Sentiment analysis of political tweets: Towards an accurate classifier. In *Proceedings of the NAACL Workshop on Language Analysis in Social Media*.

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

Burger, J. D.; Henderson, J.; Kim, G.; and Zarrella, G. 2011. Discriminating gender on Twitter. In *Proceedings of EMNLP*, 1301–1309.

De Choudhury, M.; Counts, S.; and Horvitz, E. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of WebSci*, 47–56.

De Choudhury, M.; Gamon, M.; and Counts, S. 2012. Happy, nervous or surprised? Classification of human affective states in social media. In *Proceedings of ICWSM*.

De Silva, L., and Riloff, E. 2014. User type classification of tweets with implications for event recognition. *Proceedings of ACL*.

Ekman, P. 1992. An argument for basic emotions. *Cognition & Emotion* 6(3-4):169–200.

Esmin, A. A. A.; de Oliveira, R.; and Matwin, S. 2012. Hierarchical classification approach to emotion recognition in Twitter. In *Proceedings of ICMLA*, 381–385.

Kosinski, M.; Stillwell, D.; and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*.

MacEachren, A. M.; Robinson, A. C.; Jaiswal, A.; Pezanowski, S.; Savelyev, A.; Blanford, J.; and Mitra, P. 2011. Geo-twitter analytics: Applications in crisis management. In *Proceedings of Cartographic Conference*, 3–8.

McCorriston, J.; Jurgens, D.; and Ruths, D. 2015. Organizations are users too: Characterizing and detecting the presence of organizations on Twitter. In *Proceedings of ICWSM*.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nakov, P.; Rosenthal, S.; Kozareva, Z.; Stoyanov, V.; Ritter, A.; and Wilson, T. 2013. Semeval-2013 task 2: Sentiment analysis in Twitter. In *Proceedings of SemEval*, 312–320.

Nichols, J.; Mahmud, J.; and Drews, C. 2012. Summarizing sporting events using Twitter. In *Proceedings of IUI*, 189–198.

Pak, A., and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of LREC*.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations of Trends in IR* 2(1-2):1–135.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, 79–86.

Preoţiuc-Pietro, D.; Volkova, S.; Lampos, V.; Bachrach, Y.; and Aletras, N. 2015. Studying user income through language, behaviour and affect in social media. *PloS one* 10(9):e0138717.

Qadir, A., and Riloff, E. 2013. Bootstrapped learning of emotion hashtags# hashtags4you. *Proceedings of WASSA*.

Ramakrishnan, N.; Butler, P.; Muthiah, S.; Self, N.; Khandpur, R.; Saraf, P.; Wang, W.; Cadena, J.; Vullikanti, A.; Korkmaz, G.; et al. 2014. 'beating the news' with embers: forecasting civil unrest using open source indicators. In *Proceedings of ACM SIGKDD*, 1799–1808. ACM.

Řehůřek, R., and Sojka, P. 2010. Software Framework for Topic Modeling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.

Roberts, K.; Roach, M. A.; Johnson, J.; Guthrie, J.; and Harabagiu, S. M. 2012. Empatweet: Annotating and detecting emotions on Twitter. In *Proceedings of LREC*.

Rui, H.; Liu, Y.; and Whinston, A. 2013. Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems* 55(4):863–870.

Sadilek, A.; Brennan, S.; Kautz, H.; and Silenzio, V. 2013. nemesis: which restaurants should you avoid today? In *Proceedings of Human Computation and Crowdsourcing*.

Strapparava, C., and Mihalcea, R. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of SemEval*, 70–74.

Volkova, S., and Bachrach, Y. 2015. On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. *Cyberpsychology, Behavior, and Social Networking* 18(12):726–736.

Volkova, S., and Bachrach, Y. 2016. Inferring perceived demographics from user emotional tone and user-environment emotional contrast. In *Proceedings of ACL*.

Volkova, S.; Bachrach, Y.; and Van Durme, B. 2016. Mining user interests to predict perceived psycho-demographic traits on Twitter. In *Proceedings of BigDataService*, 36–43.

Volkova, S.; Bachrach, Y.; Armstrong, M.; and Sharma, V. 2015a. Inferring latent user properties from texts published in social media. In *AAAI*, 4296–4297.

Volkova, S.; Bachrach, Y.; Armstrong, M.; and Sharma, V. 2015b. Inferring latent user properties from texts published in social media (demo). In *Proceedings of AAAI*.

Volkova, S.; Coppersmith, G.; and Van Durme, B. 2014. Inferring user political preferences from streaming communications. In *Proceedings of ACL*, 186–196.

Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; and Narayanan, S. 2012a. A system for real-time Twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, 115–120.

Wang, W.; Chen, L.; Thirunarayan, K.; and Sheth, A. P. 2012b. Harnessing Twitter "big data" for automatic emotion identification. In *Proceedings of SocialCom*, 587–592.

Yu, Y., and Wang, X. 2015. World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans tweets. *Computers in Human Behavior* 48:392–400.

Zamal, F. A.; Liu, W.; and Ruths, D. 2012. Homophily and latent attribute inference: Inferring latent attributes of Twitter users from neighbors. In *Proceedings of ICWSM*.

Zhu, X.; Kiritchenko, S.; and Mohammad, S. M. 2014. NRC-Canada-2014: Recent improvements in the sentiment analysis of tweets. *Proceedings of SemEval*.