Behavioral Deviations in Social Media Caused by Emergency Events

Aruna Jammalamadaka[1], Jiejun Xu[1], and Tsai-Ching Lu[1]

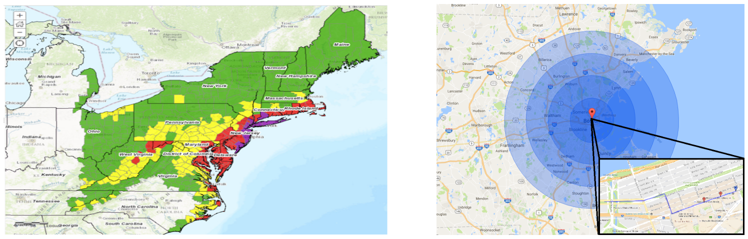
HRL Laboratories, Malibu CA, USA
{ajammalamadaka,jxu,tlu}@hrl.com

Abstract. We perform an exploratory study to estimate the social impact caused by national emergencies such as hurricanes and bombings and the evolution of that impact over time and space. We extract various features derived from Twitter data as measures of normal social media posting behavior, and use these to determine which behaviors became more or less common due to the event. A Bayesian structural time-series (BSTS) model is used to predict what would have happened in the absence of the event. We conduct a temporal analysis to determine affected time periods, and a large-scale spatial analysis using FEMA declarations of economic impact for Hurricane Sandy and radial distance from the Boston Marathon bombing. Our results show that the moderately (not severely) affected regions show the strongest behavioral deviation, and we hope that such results can eventually be used to estimate spatio-temporal resource allocation needs for protest events, terrorist strikes, and natural disasters.

Keywords: causal impact · Bayesian structural time-series · social media · disaster response

1 Introduction

Social media, especially the Twitter platform, is increasingly used during natural disasters and states of emergency to both gauge public response, and disseminate real-time information regarding the event. While numerous correlative studies have been performed using such data, causal effect (or impact) has so far been difficult to estimate. In particular, many works attempt to study the psychological, fiscal, and political impact of events such as the Boston Marathon Bombing and Hurricane Sandy [3, 9]. Our exploratory research attempts to answer the following questions: 1. When Hurricane Sandy or the Boston bombing occurred, did they cause nearby populations to exhibit deviations in their Twitter posting behavior? 2. Can quantification of the spatio-temporal extent of that impact aid in emergency resource allocation efforts for organizations like FEMA? We make use of the recently developed Bayesian structural time-series (BSTS) model [1] to determine the impact of these two events on the posting patterns of users within the affected areas by constructing a synthetic "counterfactual" i.e. a prediction of what would have happened in the absence of the event. We believe that analysis of multiple emergency events of the same type (i.e. hurricanes, bombings) could be used to train a baseline system for understanding and even predicting behavioral deviations and other impacts resulting from protest events, terrorist strikes, or natural disasters.



(a) (b)

Fig. 1. (a) Area of the northeastern US affected by Hurricane Sandy, ranging from Low impact (green), Moderate (Yellow), High (Red), and Very High (Purple) [4], (b) 5, 10, 15 and 20 mile radial distances from Boston Marathon Bombing.

2 Methodology

Dataset We utilize a 10% real time tweet stream (Gnip Decahose[1]) from April 1st, 2012 to January 1st, 2014. We obtain a geocoding of user locations by using principles of homophily, as detailed in [2]. Using these locations, we determine affected (Figure 1) and unaffected populations as detailed below. We attempt to match the affected and unaffected geographic regions by attributes that may confound the causal effect on Twitter posting behavior, such as age, ethnicity and income. We can only perform a population-level propensity score matching based on US census data, since individual demographic data (besides location) is not readily available. We believe this matching minimizes the effects of possible confounders, as does our large sample dataset of 10,000 users per geographic region. Hurricane Sandy and the Boston Marathon bombing are ideal datasets to test this methodology because we hypothesize their spatio-temporal impact to be quite different. The hurricane was not well resolved in time, since the affected population had the benefit of hearing forecasts up to 5 days before it made landfall, whereas the bombing was highly resolved in time of impact since it took the country by surprise, and was contained quickly. In addition, we analyze the Boston Marathon bombing at the city-level population, whereas Hurricane Sandy is analyzed using state-level populations. Affected areas and their corresponding level of economic impact from Hurricane Sandy were graciously provided by Dr. Gene Longenecker's team at FEMA [11]. Unaffected states were matched loosely by 2012 US Census data for population level age, ethnicity, and income. This resulted in 10,000 randomly sampled users from each of the Very High, High, Moderate, and Low areas shown in Figure 1a, as well as 10,000 users randomly sampled using shapefiles [5] from each of the 4 unaffected states: California (CA), Georgia (GA), Illinois (IL), and Texas (TX). The 40,000 unaffected and 40,000 affected (based on 4 different levels of affectation) define our total population of 80k users for the Hurricane Sandy Dataset. Since we could not obtain delineations of impact of the Boston Marathon bombing, we hypothesize that the impact might decay with distance from the location of the two bomb blasts that occurred. Therefore, we extract 40,000 randomly

[1] https://gnip.com/sources/twitter/

Table 1. Dirichlet Multinomial Mixture Model Topics

Example Words	Theme
lol, like, shit, fuck, get, got, nigga, bitch, lmao, ass	Profanity
love, like, lol, happy, birthday, one, know, good, thanks, day	Appreciation
get, day, like, good, school, going, time, lol, today, tomorrow	Daily Greetings
que, por, con, los, para, una, las, como, pero, del	Spanish
game, like, win, get, team, good, tonight, play, one, time	Sports
follow, new, followers, please, retweet, back, photo, love, one, teamfollowback	Twitter Related
like, get, want, lol, got, love, need, hair, one, right	Needs
people, like, love, know, never, life, get, someone, want, one	Opinions
new, video, music, check, https, youtube, party, get, love, tonight	Multimedia
new, today, people, one, obama, boston, get, news, like, time	News
like, people, know, get, lol, shit, fuck, hate, really, want	Opinions2

sampled users from increasingly larger rings around the blast sites. The affected user set includes 10,000 users from within each of the following 4 radial ranges: 0-5 miles of the blasts, 5-10 miles, 10-15 miles, and finally 15-20 miles, as shown in Figure 1b. The latitude and longitudinal locations of the bomb blasts were used from [6]. We randomly sample 10,000 users from each of 4 unaffected cities, which were chosen based on loosely matching 2013 US Census data for population level age, ethnicity, and income. The sampling of users was possible due to publicly available ESRI shape files for Chicago (CHI), District of Columbia (DC), Los Angeles (LA), and New York (NY). The 40,000 unaffected and 40,000 affected (based on 4 different rings of affectation) define our total population of 80k users for the Boston Marathon Bombing Dataset.

Feature Extraction For each of the above 160,000 users, we extract all of their available tweets and retweets from 6 months before the event to 6 months after the event. This results in a total of 15,880,442 tweets for the Hurricane Sandy dataset, and 14,167,952 for the Boston Marathon bombing. The features described below are extracted from each tweet and aggregated over 1 day periods to provide 16 different time-series (one per 10,000 regional user population) for each feature. We first extract simple daily counts of retweets, @mentions (one user mentioning another's handle within the tweet), hashtags, and URL links. Sentiment values for each tweet were calculated by averaging unigram values (over tweets and then over days) from the NRC Hashtag Sentiment Lexicon (version 0.1). This list of common twitter words, hashtags, and emoticons and their corresponding sentiment values comes from [8]. We also extracted 1 topic per tweet, by using a Dirichlet Multinomial topic model, and pre-processing text as described in [10] ($\alpha = 0.01, \beta = 0.1$). The paramters α and β were determined empirically, and resulting topics for both events are shown in Table 1. The final feature set that was analyzed for causal impact consisted of the following 18 features: 10 topics per dataset (see Table 1), Number of Tweets Per user, Total Tweet Count, Total User Count, Hashtag Count, URL count, Mention count, Retweet count, Average sentiment.

Causal Impact Model Our BSTS model includes only a local level trend, and contemporaneous covariates from unaffected populations with temporally static regression coefficients, therefore the observation and transition equations can be simplified as $y_t = \mu_t + \beta^T \mathbf{x}_t + \eta_t$. The local level state component assumes the trend is a random walk, i.e. $\mu_t \sim N(\mu_{t-1}, 0.01)$. The details of default priors and model inference can be found in [12, 1]. The BSTS regression component $\beta^T \mathbf{x}_t$ allows us to obtain counterfactual predictions by constructing a synthetic control based on aggregation of unaffected user behaviors. Since these regression covariates come from a randomized group of unaffected Twitter users, they already account for any additional local linear trends and seasonal variation in the response variable.

3 Results

Temporal analysis Features are extracted from tweets, and aggregated over 1 day periods within 6 months before and after the hurricane. Using less than 6 months of pre-event data did not result in a good model fit, and caused the statistical significance of the causal impact to rise artificially. We varied the start date for the affected time period and visually inspected the resulting causal impact plots and models for each of these values. An example causal impact plot is shown in Figure 2. Hurricane Sandy was first reported as heading towards the Northeast US on October 24, 2012, hit landfall October 29th, and dissipated November 2. Based on the prediction intervals described below, the impact on Twitter was best captured by a 1 week period from midnight on October 27th to November 2nd. For the Boston bombing, the causally impacted period was much more clearly defined. The affected period started on April 15 when the bombs exploded and ended slightly less than 1 week later on April 21st, as shown in Figure 2. Both impact periods were visually determined by the deviation between predicted (dashed blue line) and observed (solid black line) time-series during the affected period. We can see that outside of the affected time period, the predicted (counterfactual) tracks the observed feature values and stays within the 95% prediction envelope. Interestingly, the affected period for both Hurricane Sandy and the Boston bombing deviated the most mid-week and reconvened approximately 1 week after the start date of the intervention. The periodicity in Figure 2 is due to the fact that people post less on weekends than during the work-week. As mentioned above, there is no known ground truth temporal window of impact for either of these events, however the techniques elucidated above can be used for discovering robust estimation methods. Further statistical studies must be performed in order to model the exact relationship between the prediction intervals, p-values, and actual temporal social or economic impact.

Spatial analysis Once the temporal windows of pre-treatment, treatment, and post-treatment have been determined, we run two tests of statistical significance on each of the 18 extracted features. The first test determines impacted features based on the p-values that result from summing each feature inside the affected time period, and comparing the actual observed sum to the distribution of counterfactual predicted sums. The second test checks to make sure that there is an absence of effect on the unaffected populations, as described in "Analysis 3" of [1]. Only features which displayed statistical significance at the 0.05 level in the first test and not the second were considered causally impacted. These features were further analyzed using their relative effects and 95% posterior prediction interval, however the full results table cannot be shown here for brevity. We write $(X, [r1, r2])$ to indicate that the feature increased/decreased by $X\%$ and

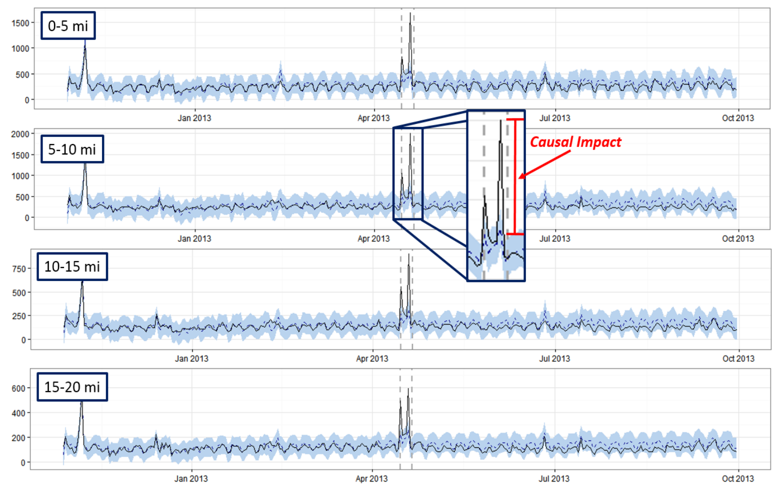


Fig. 2. Causal impact plots for "News" topic tweet count within varying radii of Boston Marathon Bombing. From the y-axis values we can see that the 5-10 mile population exhibited the largest increase in posting news-related tweets due to the bombing.

we are 95% confident that the increase/decrease falls between $r1$ and $r2$. The top 3 most severely impacted features from Hurricane Sandy all came from the "moderate" affected population: Hashtag Count ($17.82[22.35, 13.16]$), Retweet count ($14.14[18.75, 9.46]$), and tweets about multimedia (Topic 1 in Table 1, ($13.42[20.93, 6.5]$). The top 3 most severely impacted features from Boston were Twitter related for the 0-5 miles population ($-28.95[-17.36, -40.59]$) and the 15-20 miles population ($-29.47[-17.45, -41.26]$), and the "news" topic for the 5-10 miles population ($27.62[47.52, 9.33]$), shown in Figure 2. We did not see the expected trend showing that the impact tapers off as radial distance from the bombing is increased, or post-event assessed economic impact is decreased. Instead we see that the moderately affected areas ("high" and "moderate" in the case of Hurricane Sandy, and "5-10 miles" or "10-15 miles" in the case of the Boston bombing) show the most deviation in their social media behavior. We hypothesize that this is because the most severely affected populations are dealing with the emergency first-hand as opposed to tweeting about it, and the farthest or least impacted according to FEMA could be outside of the causally affected area. If emergency resource teams would like to use Twitter in the future for prediction of temporal and spatial extent of disasters, the moderately affected populations might be the strongest available signal. Interestingly, the raw tweet count showed an increase, or no impact in all of the spatial regions analyzed here, indicating that mobile networks were reasonably intact and power outages did not significantly prevent people from posting on Twitter.

4 Discussion

We expected a difference in impacted topics between the two disasters, i.e. that the bombing would result in more sympathetic, fearful posts immediately following the event [7], whereas the hurricane would results in more warning/anticipation posts prior to the event followed by a total drop in tweets during, with recovery related posts following both events. We found instead that the topics we arrived at are more general than "fear" or "sympathy" and that the only expected result is the spike corresponding to the "News" topic for Boston during the affected time period. We were able to determine that if emergency resource teams would like to use Twitter in the future for prediction of temporal and spatial extent of disasters, moderately affected populations may provide the strongest indicators. We have yet to further explore the link between finding causally impacted features, and consequently using those features for the reverse problem: to predict the time period and spatial impact of an "emergency event", or disaster. For this to be possible, future work should include multiple disasters of the same type, i.e. multiple hurricanes, so that the re-usability of these features for prediction can be explored. If we can reliably estimate the varying impact with respect to location, or possibly a network distance on social media, impacted features can be used to predict recovery time period for a new event to aid in resource allocation for organizations like FEMA. Evaluation of economic and psychological impact is often performed years after an event, therefore the utilization of social media data as a real-time sensor is appealing.

References

1. Brodersen, K.H., Gallusser, F., Koehler, J., Remy, N., Scott, S.L., et al.: Inferring causal impact using bayesian structural time-series models. The Annals of Applied Statistics **9**(1), 247–274 (2015)
2. Compton, R., Jurgens, D., Allen, D.: Geotagging one hundred million twitter accounts with total variation minimization. In: Big Data (Big Data), 2014 IEEE International Conference on. pp. 393–401. IEEE (2014)
3. DHS: Lessons learned: Social media and hurricane sandy, virtual social media working group and dhs first responders group. `https://www.hsdl.org/?abstract&did=739332` (2013)
4. FEMA.MOTF: Fema modeling task force (motf) - hurricane sandy impact analysis. `http://fema.maps.arcgis.com/home/item.html?id=307dd522499d4a44a33d7296a5da5ea0/` (2015)
5. Fitzpatrick, C.: U.s. 50 states plus dc shapefile. `https://www.arcgis.com/home/item.html?id=f7f805eb65eb4ab787a0a3e1116ca7e5` (2012)
6. Google: Boston marathon bomb blasts. `https://www.google.com/maps/d/viewer?mid=1LEkpLUOHooonTe1JDULLwM8mQyQ&hl=en_US` (2013)
7. Lin, Y.R., Margolin, D.: The ripple of fear, sympathy and solidarity during the boston bombings. EPJ Data Science **3**(1), 31 (2014)
8. Mohammad, S.M., Kiritchenko, S., Zhu, X.: Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In: Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013). Atlanta, Georgia, USA (June 2013)
9. Nazer, T.H., Xue, G., Ji, Y., Liu, H.: Intelligent disaster response via social media analysis a survey. ACM SIGKDD Explorations Newsletter **19**(1), 46–59 (2017)
10. Nguyen, D.Q.: jLDADMM: A Java package for the LDA and DMM topic models. `http://jldadmm.sourceforge.net/` (2015)
11. Rozelle, J., Longenecker, G.: Fema modeling task force. Personal Communication via Email (2016)
12. Scott, S.L., Varian, H.R.: Predicting the present with bayesian structural time series. International Journal of Mathematical Modelling and Numerical Optimisation **5**(1-2), 4–23 (2014)