



Predicting gasoline shortage during disasters using social media

Abhinav Khare¹ · Qing He^{1,2} · Rajan Batta¹

Received: 5 November 2018 / Accepted: 19 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Shortage of gasoline is a common phenomenon during onset of forecasted disasters like hurricanes. Prediction of future gasoline shortage can guide agencies in pushing supplies to the correct regions and mitigating the shortage. We demonstrate how to incorporate social media data into gasoline supply decision making. We develop a systematic approach to examine social media posts like tweets and sense future gasoline shortage. We build a four-stage shortage prediction methodology. In the first stage, we filter out tweets related to gasoline. In the second stage, we use an SVM-based tweet classifier to classify tweets about the gasoline shortage, using unigrams and topics identified using topic modeling techniques as our features. In the third stage, we predict the number of future tweets about gasoline shortage using a hybrid loss function, which is built to combine ARIMA and Poisson regression methods. In the fourth stage, we employ Poisson regression to predict shortage using the number of tweets predicted in the third stage. To validate the methodology, we develop a case study that predicts the shortage of gasoline, using tweets generated in Florida during the onset and post landfall of Hurricane Irma. We compare the predictions to the ground truth about gasoline shortage during Irma, and the results are very accurate based on commonly used error estimates.

Keywords Social media analytics · Gasoline shortage prediction modeling · Disaster management · Hybrid loss function · Hurricane Irma

✉ Rajan Batta
batta@buffalo.edu

Abhinav Khare
abhinavk@buffalo.edu

Qing He
qinghe@buffalo.edu

¹ Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA

² Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY 14260, USA

1 Introduction

Shortage of essential supplies, like food, water and fuel, is a common problem in disasters like Hurricanes. There is a surge in demand as people panic-buy and hoard supplies in preparation for evacuation to safer areas or for staying indoors for long periods (Flood 2017). Price gouging of commodities is also observed (Fessenden 2017). These shortages continue on for a few days beyond the disaster. It is imperative that the demand of essential commodities is satisfied in a timely fashion, both pre- and post-disaster to mitigate losses. An accurate prediction of the surges in demand and subsequent shortages could help the authorities and the first responders plan better. It would give time to arrange for the additional supplies and infrastructure for pre-positioning, re-positioning and directing supplies in the impacted area. For instance, in the case of Hurricane Irma in Florida, the gasoline demand surged by 150 percent and distribution became the main limitation for fuel deliveries and shortages (Fdot 2017). Florida had sufficient fuel available at the ports but did not have enough carriers and drivers to transport the extra fuel from the ports to the gas stations. Additional drivers and carriers were brought in later from Arizona. If the shortage in the affected areas could have been predicted more drivers and carries could have been planned for in advance (Fdot 2017). To solve this prediction problem, in our work, we turn to a non-traditional prediction approach with social media data.

Recently, social media is transforming the way people communicate not only in daily lives, but also during disasters. There is a surge in usage of social media during an emergency in the affected regions. Nowadays, many people are willing to share the disaster information through social media. One evidence is given in Table 1 which contains top five disasters (by type) discussed in tweets. We collected these tweets from four major metropolitan areas during the year 2015. More than 180,000 tweets are found to discuss top five types of disasters, collected from four major US cities in 2015 only. Public uses social media to communicate, seek information, raise concerns and express sentiments, and responders use it to plan and communicate important messages to the public (Lachlan et al. 2014; Liu et al. 2016; Panagiotopoulos et al.

Table 1 Top five disasters by tweet distribution in five metropolitan areas in 2015

Disaster type	Num. of tweets	Percentage (%)	Related keywords
Earthquake	114,428	53.91	haiti, nepal, america, ene, california, julian california, wnw, japan, ssw, united states, earthquake, ese, magnitude, italy
Hurricane	29,098	13.71	wind, united states, atlantic ocean, alcohol, europe, work, rain, hit, school, america, people, storm
Drought	17,114	8.06	california, louisiana, lips, sacrifice, last year, COP21, poor people, lush, boreal forests, syria, ethiopia, maryland
Tornado	14,398	6.78	western, working, stay, phone, california, area warning, basement, work, county, canada, shelter, rotation

2016; van Gorp et al. 2015). As a result, there is a keen interest in employing social media for disaster management. For example, social media has been used to build a mass communication channel, in order to inform large numbers of stakeholders at once (Ki and Nekmat 2014; Stříteský et al. 2015; Utz et al. 2013). Social media can also aid in decision support systems and emergency management processes by utilizing the enormous amounts of real-time data it generates (Gaynor et al. 2005; Boulos et al. 2011). Consequently, multiple social media data analysis techniques have been developed in the context of a disaster, ranging from tools for event detection, prediction and warning; impact assessment; situation awareness; disaster tracking; and response planning.

However, there is scant literature exploiting social media data for detection and prediction of demand and shortage of essential commodities. We aim to fill this research gap by focusing on using social media data to predict the shortage of gasoline 1 day in advance (everyday) during the onset and post landfall of foreseen disasters like hurricanes. Our motivation comes from the fact that people have been found to use Twitter to tweet about shortages and needs during a disaster (Stowe et al. 2016; Tien Nguyen et al. 2016). For instance, during gasoline shortage in Florida in the onset of Irma, the following kinds of tweets were observed:

“The shelters are full, there is no gas. Tornados could happen, and storm surge is predicted. So what are people supposed to do? Irma”

“Insane..95 percent of Florida trying to leave at one time. Roads r slammed. No gas. No hotels available. Scared to see my neighborhood after irma”

“Gas stations out of gas, water shelves empty, stores and airports closed. Stocked up on food and wine, waiting on irma”

The natural question that arises in such a scenario is if Twitter posts be used to sense current shortage and forecast future shortages? There are two main challenges related to this question:

- *Challenge C1—how to identify tweets about shortage* Social media data, especially from twitter, is difficult to process and classify as it is unstructured, noisy and contains a plethora of information (large number of tweets). Also, a single tweet contains a maximum of 140 characters, is informal and contains abbreviations and spelling mistakes. Interpreting the semantics of such a short message and classifying it is a hard problem. There are methods in the literature that have classified tweets generated during crisis into caution/advise, information source, people, casualties and damage (Imran et al. 2013), pre-disaster or post-disaster (tweet4act) (Chowdhury et al. 2013), tweets reporting casualty or damage (Tweedr) (Ashktorab et al. 2014), information, preparation and movement (Stowe et al. 2016). However, classifying tweets for a specific problem like identifying gasoline shortage has never been done. Identifying important features for this classification task is a novel and unique question. If these issues are resolved, then one can identify tweets about gasoline shortage and treat them as sensors for shortage.
- *Challenge C2—forecasting the spatiotemporal shortage from tweets* The spatiotemporal distribution of tweets about shortage is not equivalent to the spa-

tiotemporal shortage distribution. Spatial and temporal lag between the origin of the shortage and the tweet about shortage is an uncertain quantity. This makes forecasting shortages using tweets a challenging problem.

To address challenge C1, we developed a tweet classifier that uses unigrams and latent topics (identified by topic modeling techniques) as features to identify tweets about gasoline shortage. Analysis of the identified tweets shows us that the number of tweets about shortage (in a day, in a city) predicts the number of stations out of gasoline. A detailed analysis of the spatiotemporal dynamics of the arrival of shortage tweets shows us that the arrival of tweets in a city follows a Poisson distribution. Using these two insights, we tackle challenge C2, as follows. We develop a regression model with a unique hybrid loss function (HLF that combines the properties of Poisson regression and time-series-based ARIMA models) to predict the number of future tweets in a city. A separate Poisson regression model is used to predict the amount of shortage from the predicted number of tweets.

Our contributions can be summarized as follows:

- Building of a classifier that identifies tweets about gasoline shortage from the corpus of all the tweets generated in the affected area.
- Discovering that the arrival of tweets about gasoline shortage follows a Poisson distribution.
- Developing a hybrid loss function method (HLF) that forecasts the number of tweets about gasoline shortage.
- Developing a four-stage gasoline shortage prediction methodology which takes tweets generated on a day in an affected city as input and generates the number of stations that will be out of gas on the next day as the output.
- Model validation with a case study based on Hurricane Irma, which contains around 1 million tweets, hurricane path data and ground truth about gasoline shortage.

The paper is organized as follows. Section 2 describes the related works, which is followed by the presentation of the details of our methodology in Sect. 3. Section 4 explains the application of our methodology in the case of gasoline shortage in Florida during Hurricane Irma in 2017 and presents our numerical results. Section 5 provides our conclusions and our future suggestions for improvement.

2 Related work

There is a surge in the use of social media during crisis as stated in Sect. 1. Social media Web sites like Facebook and Twitter have started playing a major role in disaster management such as post Japan Tsunami in 2011 (Kaigo 2012) and US Hurricane Sandy in 2012 (Hughes et al. 2014). Other Internet-based social applications like Waze (Waze 2017) and GasBuddy (Gasbuddy 2017b) have also set up special-purpose services to allow individuals to participate and report the availability of various resources (e.g., gas stations) via the Web or smartphones. These services were used by a large population after Hurricane Irma and Sandy. Apart from these cases of direct applications of social media during disasters, surveys by Imran et al. (2013) and Nazer et al. (2017) provide evidence that there has also been an uptick in research interest in the development of

social media data analysis techniques in the context of disaster management. These techniques can be categorized into: (1) data extraction and filtering, (2) event detection and impact assessment and (3) response planning and relief delivery (Nazer et al. 2017). Some papers address multiple issues that cross these boundaries. However, for ease of reading, we have categorized them as indicated above.

Data extraction and filtering Social media data are gigantic, noisy and unreliable. To obtain the posts that contain relevant information, posts are either extracted on the basis of important keywords (Imran et al. 2013; Starbird and Stamberger 2010; Olteanu et al. 2014) or using geo-location (Morstatter et al. 2014; Cheng et al. 2010; Han et al. 2013; Schulz et al. 2013). The posts extracted using the aforementioned techniques often contain rumors and spam and cannot be trusted as pointed out by Mendoza et al. (2010). Although rumor and spam detection is a hard problem, few methods have been successful in particular cases in recent times (Gupta et al. 2013; Sampson et al. 2015).

Event detection and impact assessment Twitter has been shown to have a potential for earthquake detection and act as an early warning system by using tweets as sensors. (Sakaki et al. 2010; Faulkner et al. 2011). Apart from this, there are some event detection and impact assessment methods that include sentiment analysis methods (Beigi et al. 2016; Caragea et al. 2014) and language change methods (Atefah and Khreich 2015; Cordeiro and Gama 2016).

Response planning and relief delivery Response planning requires situation awareness for which there are classifiers that classify posts into caution/advise, information source, people, casualties and damage (Imran et al. 2013), pre-disaster or post-disaster (tweet4act) (Chowdhury et al. 2013), tweets reporting casualty or damage (Tweedr) (Ashktorab et al. 2014), information, preparation, movement etc. (Stowe et al. 2016), into user-defined categories (AIDR) (Imran et al. 2014). For response delivery, there are tools crowdsourcing communities like Digital Volunteers (translation of posts, geotagging, building maps of damaged region) and OpenStreetMap [OSM, for volunteer to build maps for response used effectively in Haiti earthquake (Zook et al. 2010)]. When it comes to data-driven relief delivery tools, there is Ushahidi (2017) which is a platform that maps information from different sources like Twitter, RSS feed, SMSs, manual comma-separated files to a singular map of the affected area. AIDR (Imran et al. 2014) is an end-to-end data pipeline that extracts and classifies tweets for responders to assess and respond to the situation on the ground. TweetTracker is a system that tracks, analyzes and understands tweets related to specific topics. It has many functionalities and can use data from multiple social media Web site. However, it has a special module for disaster relief. It detects request for help tweets using a classifier based on n-grams and tweet meta-data and shows geo-location of tweet on the map if available (Kumar et al. 2011).

Aside from the papers cited above, we identified a paper by Gu et al. (2014) which is closest to our work, and hence presented separately. Their paper develops a methodology for sensing demands of essential commodities like food, water and gasoline using data extrapolation in participatory sensing applications. Participatory sensing technologies include sources that measure the state of the point of interest and report it at a later time (e.g., on getting access to WiFi). Their paper argues that data extrapolation algorithms that rely predominantly on spatial correlations or predominantly on

temporal correlations tend not to work consistently well, as the relative importance weights of temporal versus spatial correlations change significantly between periods of calm and periods of change post a disaster. Therefore, they develop a hybrid predictions algorithm combining spatial and temporal prediction methods which predicts the status of point-of-interest (POI) sites, when collected data are incomplete. Their methodology combines spatial and temporal extrapolations method for shortage prediction. We tackle this issue by combining temporal extrapolations with predictions using other factors related to the disaster (like hurricane path, days from arrival) to improve the accuracy of shortage prediction. For this, we fuse the ARIMA method with Poisson regression method using a hybrid loss function. As far as application is concerned, their methodology does a fine grain prediction at the level of POI and our methodology is suitable for making predictions at a city level.

As evident, the literature on social media data-driven disaster management is plentiful. However, there is limited literature that proposes to use social media assessing the demand and shortage of essential commodities in the affected population during a disaster. Our work addresses this research gap. Our methodology provides a means to assess shortage of commodities and can be used to prepare, preposition and redirect supplies before a disaster.

3 Data description

Our dataset had roughly one million tweets from Florida during the period September 6–15, 2017. The data covered a data frame in R with 1,048,575 rows and 41 columns that include *TWEET ID*, *TWEET TEXT*, *USER ID*, *DATE*, *HASHTAG*, *LATITUDE* and *LONGITUDE*. Summary statistics of the tweet data is given in Table 2.

Apart from the Twitter data, we also collected ground-truth data about gas shortage from Gasbuddy Application (Gasbuddy 2017a) and details and predictions of the Hurricane path from the National Hurricane Center Web site (National Hurricane Centre 2017). Table 3 represents a small sample of the data we collected and tabulated for each major city. We collected these data for eight cities, namely Gainesville, Jacksonville, Miami, Orlando, Tallahassee, Tampa, Naples and West Palm Beach, for the period of

Table 2 Summary statistics of tweet data

Summary statistic	Values
Number of tweets collected	1,048,575
Number of unique twitter users	111,801
Period of data collection	September 6–15, 2017
Date of Irma landfall in Florida	September 9, 2017
Number of tweets prior to Irma landfall in Florida	456,530
Number of tweets during Irma in Florida	151,792
Number of tweets post Irma in Florida	440,253
Number of gas-related tweets before landfall	2,805

Table 3 Gas shortage and hurricane prediction for different cities in Florida

City	Date	Proportion of gas stations without gas	On hurricane path	Inside 3-day cone	Inside 5-day cone	Days to arrival	Watch/warning	Wind speeds (mph)
Gainesville	09/07/17	0.58	y	n	y	4	n	175
Jacksonville	09/08/17	0.31	n	y	y	3	n	155
Miami	09/07/17	0.42	y	y	y	3	Watch	175
Orlando	09/08/17	0.35	y	y	y	3	Watch	155
Tallahassee	09/08/17	0.46	n	n	y	3	n	155
Tampa	09/06/17	0.3	n	n	y	5	n	185
Naples	09/07/17	0.54	n	y	y	3	Watch	175

September 6–15, 2017 (dates when shortage was observed). These cities were selected because they were the ones that experienced significant gas shortages during the onset of Hurricane Irma (Gasbuddy 2017a). For each date and city, we determined whether the city was predicted to be on the hurricane path, whether it was inside the hurricane 3-day or 5-day cone, the number of days to arrival of the hurricane, whether there were any hurricane/thunderstorm warning and watches from National Hurricane Center in the city on that date, the maximum sustained wind speed, and population of the city, the number of gas stations in the city and proportion of gasoline stations without gasoline. The idea behind collection of these attributes is that these variables also drive panic-buying behavior causing shortage and also influence the tweeting behavior of the people. Therefore, they are potential predictors of gasoline shortage and tweeting behavior of people in the models.

4 Methodology

As stated in Sect. 3, people tweeted extensively during the onset and landfall of Hurricane Irma in Florida. Figure 1 illustrates our four-stage methodology for the task of going from tweets generated (on a day in a city) to prediction of the number of stations out of gas (on the subsequent day). In stage 1, we filter out tweets related to gasoline by using keywords and regular expressions, and remove space, stopwords, stemwords from the noisy tweets. The tweets that remain after this stage are labeled “gasoline-related” tweets. In stage 2, we classify the gasoline-related tweets into “gasoline shortage” tweets and “non-gasoline shortage” tweets, using a support vector machine classifier that employs unigrams and latent topics as features. In stage 3, gasoline shortage tweets first are aggregated for each major city and then, along with other important features about the disaster, are input into a model with hybrid loss function (HLF) to generate the predicted number of future gasoline shortage tweets. In stage 4, the predicted number of future tweets, along with other features about the disaster, is input into a Poisson regression model to predict the number of stations without gasoline on the subsequent day. It is a rolling horizon methodology in which the gasoline shortage and the number of tweets about the actual gas shortage are used to

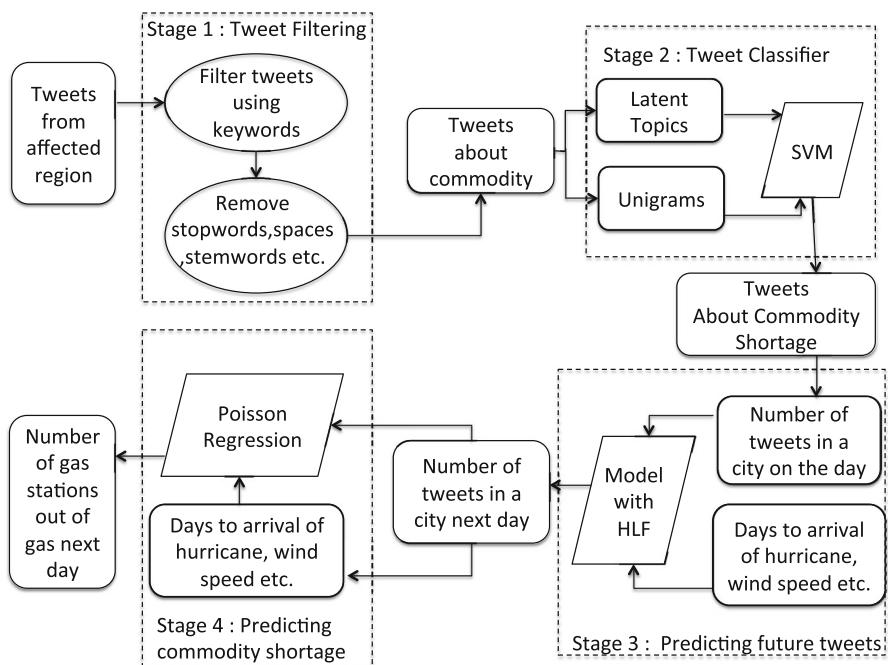


Fig. 1 Methodology

predict the number of tweets and shortage for the next day. The following subsections explain these stages in detail.

4.1 Stage 1: tweet filtering and creation of tweet corpus and document-term matrix

Stage 1 has two steps, tweet filtering to generate gasoline-related tweets and creation of a tweet corpus and a document-term matrix.

4.1.1 Tweet filtering

We filtered out “gasoline-related” tweets from the compendium of tweets generated in the affected area. We do this by keyword search in both the content and hashtags of each tweet. In the case of gasoline, any word which has the letters “gas” as part of the word is a possible keyword. We use regular expressions to identify these keywords. The regular expression `^gas` finds words starting with “gas” and also finds words that contain the string “gas” (e.g., the word “nogas”). For searching tweets, we use the regular expression `//bgas` to look for words in a sentence starting with “gas.” These regular expressions are the ones used with `grep()` function in R. Next, we identified the relevant keywords filtered through the regular expressions and retain the tweets containing those keywords. Finally, we combined the tweets curated from searching hashtags and tweets by using inner join, to account for duplicate tweets. Even after

Table 4 A small sample of the document–term matrix

Docs	Terms									
	can	gas	get	got	hurricaneirma	irma	just	line	station	water
1195	0	2	0	0	0	0	0	0	2	0
1433	0	2	0	0	0	1	0	0	0	0
267	0	2	1	1	0	0	0	1	0	0
272	1	1	0	0	0	0	0	0	0	1
298	0	0	0	0	0	0	0	0	0	0
408	0	1	0	1	0	0	1	0	0	0
443	0	1	0	1	0	0	1	0	0	0
556	0	1	0	0	0	1	0	0	0	1
680	0	1	0	0	0	0	2	0	1	0
901	0	1	0	0	0	1	1	0	0	1
Total	1	12	1	3	0	3	5	1	3	3

this filtering, we have many tweets that are very noisy and need to be cleaned so as to facilitate further processing. We achieved this cleaning by removing user names, links, punctuations, tabs and general whitespaces.

4.1.2 Corpus and document–term matrix generation

For any text mining application, there is a need for a framework of managing and manipulating heterogeneous text documents. (In our case, it is tweets.) The conceptual entity which provides this functionality is a text corpus which is a collection of the text documents being analyzed. According to Meyer et al., “It represents a collection of text documents and can be interpreted as a database for texts. Its elements are TextDocuments holding the actual texts and local meta-data” (Meyer et al. 2008). In our application, a text corpus (tweet corpus) is created using the tm package in R (Feinerer 2008). We note that, in our case, one tweet is equivalent to one text document of the corpus. From the corpus, stopwords (common words which have little or no value in classification, e.g., “the,” “and” and “a”), cursewords and numbers are removed. Words are reduced to their stemwords, e.g., words like “going” and “gone” are converted to “go.” Next, a term–document matrix is exported from the tweet corpus. Table 4 shows a sample from the term–document matrix of our case study. The document ID (Tweet ID) represent rows and terms/words represent the columns. The matrix elements are term frequencies. For instance, the term “gas” has been used twice in the tweet with Tweet Id = 1195.

4.2 Stage 2: classification to identify gasoline shortage tweets

To classify the tweets as “gasoline shortage,” first, they are manually annotated. Next, a SVM classifier is trained on a training set that classifies tweets using two kinds

of features: unigrams and abstract topics. In the following subsections, we describe unigrams and latent topics in depth.

4.2.1 Finding important unigrams

In the field of computational linguistics, an n -gram is a sequence of n items (phonemes, syllables, letters, words or base pairs depending on application) from a given sample of text or speech. An n -gram of size one is called a unigram. In case of classification of text or documents, generally, words are treated as unigrams. In our application of tweet classification, we consider a word of the tweet as a unigram. We use unigrams as features as they have predicted power for the classification task. For instance, if a tweet contains words like “gas,” “gasoline,” “shortage,” “no,” “long,” “line,” it is likely the tweet is about “gasoline shortage.”

However, all the unigrams present in a tweet do not have predictive power. Hence, it is important to remove the less important terms in a tweet. There are two known ways in text mining to do this, using the measure of term frequency (tf) or the measure of term frequency-inverse document frequency (tf-idf). The number of times a term occurs in a document/tweet is called its tf. The elements of the document-term matrix in Table 4 are tfs. On the other hand, inverse document frequency (idf) is a measure of how much information the word provides, i.e., if it is common or rare across all documents. It is the logarithmically scaled inverse fraction of the documents that contain the word (obtained by dividing the total number of documents by the number of documents containing the term). So a high term frequency-inverse document frequency (tf-idf) is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents. In our case study, we found that using tf provided us better classification accuracy than tf-idf. This is because terms like “gas,” which were very common across all documents, have small tf-idf values.

We now present the process of filtering using tf. Suppose we define important unigrams as words that have been used at least ten times in the tweet corpus. This value “10” is our threshold. Suppose Table 4 shows the term-document matrix we exported from the corpus. Each element in the matrix is the tf of the term (represented by the column) in the tweet (represented by the row). Last row in Table 4 shows the total usage of each term in the matrix (sum of term frequency). Recall, we chose threshold to be 10. Now, “gas” is the only important unigram as it is used 12 times in the matrix. However, if we reduce the threshold to 3, the important unigrams are “got”, “gas”, “irma”, “just”, “station” and “water.” As the threshold increases, the number of important unigrams decreases.

4.2.2 Finding important topics

The other set of features is abstract topics that exist in the “gasoline shortage” tweet corpus. For example, in multiple instances, people were tweeting “to inquire which gasoline stations had gas.” In another topic, people were “complaining about long lines at the gas stations.” These topics, if identified, could have high predictive power and could be used as features for predicting tweets about gasoline shortage. Therefore,

in our methodology, we use topic models to identify hidden and abstract topics in our tweets.

An empirical comparison by Lee et al. shows the advantages and limitations of four different topic models, namely latent semantic analysis (LSA), probabilistic latent semantic analysis (PLSA), latent Dirichlet allocation (LDA) and correlated topic models (CTM) (Lee et al. 2010). Their comparison showed that LDA and CTM outperformed the other two techniques. LSA works well for unique and distinctive topics, and PLSA works well in identifying a single topic in document. Therefore, we modeled our tweets using LDA and CTM. LDA is a Bayesian mixture model which assumes that topics are not correlated (Blei et al. 2003). CTM eliminates the correlation assumption in LDA (Blei et al. 2007). We used the R package “topicmodels” for the implementation of LDA and CTM (Hornik and Grün 2011).

Since LDA and CTM are Bayesian models, we need to use Bayesian inference for parameter estimation. For estimation of CTM parameters, the package “topicmodels” uses the variational expectation minimization (VEM) algorithm, while for LDA both VEM and Gibbs sampling are available (Hornik and Grün 2011). This package currently provides an interface to the code for fitting an LDA model and a CTM with the VEM algorithm as implemented by Hoffman et al. (2010) and to the code for fitting an LDA topic model with Gibbs sampling written by Phan et al. (2008). Both VEM and Gibbs sampling provide approximate estimates. VEM is a deterministic method that converges faster but has higher bias in its estimate (Wainwright et al. 2008). Gibbs sampling is a Markov chain Monte Carlo sampling method (stochastic) which is computationally expensive but its bias and variance approach zero as you draw more samples (Geman and Geman 1987). The parameter α in LDA model is estimated by default in both VEM and Gibbs sampling methods of topicmodels package. Its starting value is kept at $50/k$, where k is the number of topics, as suggested by Griffiths et al. (2004). However, there is an option of fixing the value of α as $50/k$.

To find the best model, first, the document-term matrix was divided into training and testing sets in the ratio of 70:30. Next, the following four kinds of models were estimated using the training set:

1. LDA 1 (estimation using VEM),
2. LDA 2 (estimation using VEM with a fixed α parameter),
3. LDA 3 (estimation using Gibbs sampling),
4. CTM.

In addition to finding the best topic modeling paradigm, the hyperparameter “number of topics,” k , in the tweet corpus is determined. This is done by training each of the above models (LDA 1 through CTM) for 11 values of k (2, 4, 5, 8, 10, 12, 15, 20, 40, 50, 100). In all, 44 models were trained (11 for each modeling paradigm). The performance of each of these models is evaluated on the independent test set using the measure of perplexity. The perplexity is often used to evaluate the language models on held-out data and was also used in a seminal paper on LDA (Blei et al. 2003). Information theory uses geometric mean per-word likelihood to measure how well a probability distribution or probability model predicts a sample and in topic models (Hornik and Grün 2011), where a smaller perplexity indicates a better model fit. In

Sect. 5, we show which model was best suited for modeling our tweet corpus. The topics determined by the best model are used as features in the SVM classifier.

4.2.3 Model selection

In the previous two subsections, we explained how to filter out important unigrams and topics. In this subsection, we explain how model selection was performed (i.e., selecting the best set of unigrams and topics for classification). We used the standard technique in which we divided our tweet data into training and testing datasets. We trained multiple models with different sets of unigrams and topics and measured their performance using the F1 score measure. F1 score is the harmonic mean of precision and recall. In binary classification, “precision” is the number of true positives divided by the total number of true and false positives. In our application, it measures the fraction of tweets that were correctly classified as gasoline shortage tweets. “Recall”, also known as “specificity,” is defined as the number of true positives divided by the sum of true positives and false negatives. In our application, it measures the fraction of tweets that are correctly classified out of the tweets which were originally about gasoline shortage. In most classification problems, there is often a trade-off between “precision” and “recall.” When one tries to increase precision, “recall” decreases and vice versa. Since F1 score is the harmonic mean of both, it achieves a high value only when both “precision” and “recall” are reasonably high. Therefore, in our case study in Sect. 5 we compare the F1 score of different classifiers with different sets of unigrams and topics to find the best classifier for our application.

4.3 Stage 3: forecasting gasoline shortage tweets using a hybrid loss function (HLF)

In this stage, we aggregate the number of tweets about gasoline shortage for each city. Upon a Poisson regression analysis, we found that using gasoline shortage tweets identified in the previous stage shows us that the number of tweets about shortage (in a day, in a city) is a good predictor of the amount of shortage, i.e., the number of stations out of gasoline. Table 8 in our case study shows that the number of tweets along with other variables is statistically significant predictors of the number of stations out of gas. We can use this methodology to predict future shortage if we could forecast the number of tweets about gasoline shortage. This motivates the need to forecast the number of tweets about gasoline shortage.

We explored three methods to forecast the tweets, namely (a) a Poisson regression model, (b) time-series models like ARIMA and SARIMA and (c) regression model with hybrid loss function that we developed to combine the properties and results of the Poisson regression and time-series models. In the following subsections, we explain the motivation and details of the three methods. The results of the model selection (for each type of method) on the Irma data are discussed in Sect. 5. Furthermore, in Sect. 4.3.4, the model we describe the model selection methodology meant for selecting the best procedure among the three.

4.3.1 Poisson regression model

We analyzed the hourly and daily arrival of tweets in all the cities and found the distributions to be Poisson. The details of this analysis are given in Sect. 5. Thus, Poisson regression was a candidate method. The results in Table 8 confirm that Poisson regression allows us to conclude that the “number of tweets” and other variables (about hurricane path) can be used to predict the number of stations out of gas. This motivated us to explore that whether the number of tweets on the next day could be predicted using Poisson regression and variables like number of stations out of gas. We detail these results in Sect. 5.

We had access to multiple variables that could be used for features. These variables are listed in Sect. 3. Therefore, for model selection, we used the AIC and pseudo- R^2 (pseudo- $R^2 = 1 - \text{null deviance}/\text{residual deviance}$) which are measures of model fit. Equation (1) describes the selected Poisson regression model. Y is the dependent variable, namely the Number of Tweets about gasoline shortage in a city on the next day, followed by all the independent variables, namely x_1 : the average gasoline shortage one day prior, x_2 : the number of gas stations in the city, x_3 : a binary variable equal to 1 if the city is in the path of the Hurricane or else equal to 0, x_4 : a binary variable is equal to 1 if the city is inside the 3-day cone (i.e., within 3 days from potential hurricane strike) or else equal to 0, x_5 : the number of days to the arrival of the hurricane in the city, x_6 : a categorical variable for watches or warning issued in the city and x_7 : the maximum sustained wind speed of the hurricane in mph. In the Poisson regression model, the logarithm of the expected values of number of tweets about gasoline shortage is modeled to vary linearly with the independent variable described above. It is solved by minimizing Eq. (2) which is the negative of the log likelihood function. (Negative of the log likelihood is a convex function, so gradient methods can be applied for minimization.) In Eq. (2), θ is the vector of all β 's and X is a vector of all independent variables.

$$\log(E(Y)|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta_7 x_7 \quad (1)$$

$$-L(\theta, Y, X) = e^{\theta^T X} - Y \theta^T X \quad (2)$$

4.3.2 Time-series models

Natural candidates for forecasting gasoline shortage tweets are time-series models like ARIMA and SARIMA (Brockwell et al. 2002). We now provide some background for these models. The autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. The difference is that ARMA is used when the time series is known to be stationary, whereas ARIMA is used when the data show evidence of non-stationarity. ARMA models provides a description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the other for the moving average (MA) (Box et al. 2015). ARIMA involves an initial differencing step to eliminate the non-stationarity (Brockwell et al. 2002). We used the augmented Dickey–Fuller test (Said and Dickey 1984) to test stationarity (see Sect. 5). Our results showed that the

times series of the number of tweets (in 1 h) for a number of cities in Florida were stationary. For other cities, a differencing operation made the series stationary. Therefore, ARIMA models were suitable for modeling the time series of the tweets. Miami data also needed seasonality adjustment (24-h seasonality), and hence, SARIMA, a version of ARIMA that models seasonality, was employed to analyze and fit the Miami data.

Equation (3) is an autoregressive model, AR(p), in which Y_t is the number of tweets at hour t , ϕ 's are parameters, c is a constant and the random variable ϵ_t is white noise. Equation (4) is a moving averages model, MA(q), in which the new term μ is an expectation of Y_T (assumed as zero in most cases) and ρ 's are the parameters. Equation (5) describes an ARMA model, ARMA(p,q), which models time series with AR(p), MA(q) or a combination. Equation (6) simplifies Eq. (5), by using the lag operator defined in Eq. (7). Equation (8) describes the ARIMA model, ARIMA(p,d,q), which we use to model the times series of tweets in all cities. It introduces the differencing operation described in Eq. (9) which helps convert non-stationary series into stationary series.

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t \quad (3)$$

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^q \rho_i \epsilon_{t-i} \quad (4)$$

$$Y_t = \epsilon_t + \sum_{i=1}^q \rho_i \epsilon_{t-i} + \sum_{i=1}^p \phi_i Y_{t-i} \quad (5)$$

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) Y_t = \left(1 - \sum_{i=1}^q \rho_i L^i\right) \epsilon_t \quad (6)$$

$$LY_t = Y_{t-1} \quad (7)$$

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d Y_t = \left(1 - \sum_{i=1}^q \rho_i L^i\right) \epsilon_t \quad (8)$$

$$Y_t^{d=1} = Y_t - Y_{t-1} \quad (9)$$

For model selection in each city, we used the Box–Jenkins methodology (Box et al. 2015), which had the following four steps:

1. *Model identification* We ensured that the variables/differenced variables were stationary using the augmented Dickey–Fuller test. Seasonality was identified if present (seasonally differencing it, if necessary). Plots of the autocorrelation and partial autocorrelation functions of the time series were used to decide which components to use, among autoregressive, moving average, differencing and seasonality.
2. *Model selection* Multiple ARIMA models were fit to the time series and the best model was selected using the Akaike information criterion (AIC).
3. *Parameter estimation* This was done using maximum likelihood estimation.

4. *Model checking* We checked whether the selected model conforms to the properties of a stationary univariate series. In particular, we check that the residuals are uncorrelated and normally distributed using their autocorrelation functions for several lags. We further checked they were uncorrelated using the Ljung–Box test.

4.3.3 The hybrid loss function model

Our results from both Poisson regression and time-series models reflected that there was room to improve. Poisson regression captures variations due to the number of stations out of gas and hurricane path variables. Time-series models capture variation in the form of temporal covariances. Even though there is overlap in the explanation of variance through the two methods, it is possible that a combination of the two methods could potentially explain greater amount of variance in tweet data. This motivated us to combine the two methods. In the literature, we found multiple instances where ARIMA models were combined with other predictive algorithms. They have been combined with linear regression (Xu et al. 2016), a variety of neural networks (Zhang 2003; Tseng et al. 2002; Cadenas and Rivera 2010) and support vector machines (Pai and Lin 2005; Nie et al. 2012; Zhu and Wei 2013; Ni et al. 2017). However, there is no literature that builds a hybrid of ARIMA and Poisson regression. The application that comes closest to our work is the combination of SARIMA and SVM regression by Ni et al. which forecasts subway passenger flow using tweets (Ni et al. 2017).

In the HLF method, we combine the properties of Poisson regression method and ARIMA models using Eq. (10) which is the hybrid loss function. The function is convex and can be optimized using any gradient methods. We used the gradient descent method for the optimization using the gradients in Eqs. (11) and (12). In these equations, θ is the vector of parameters for the model, X_{train} is the matrix of the values of the independent variables in the training data [independent variables described in Eq. (1)], X_{test} is the matrix of the values of the independent variables in the test data, Y_{train} is vector of values of dependent variables in the training data [Y is the number of tweets about gasoline shortage on the next day as described in Eq. (1)], Y_{ts} is the vector of the values of the independent variables predicted by time-series methods on the test data, Y' is the vector of values of the dependent variables to be predicted by the combined method entered as a parameter of the loss function. $e^{\theta^T X} - Y\theta^T X$ is the negative of the log likelihood function for Poisson regression. Since this is a convex function, the weighted sum of the terms $Y_{\text{train}}\theta^T X_{\text{train}}, e^{\theta^T X_{\text{test}}} - Y'\theta^T X_{\text{test}}$ and $(Y' - Y_{\text{ts}})^2$ is also a convex function. The role of term $e^{\theta^T X_{\text{train}}} - Y_{\text{train}}\theta^T X_{\text{train}}$ in the loss function is to find the best θ for that minimizes the negative of the log likelihood function of Poisson regression. The term $e^{\theta^T X_{\text{test}}} - Y'\theta^T X_{\text{test}}$ minimizes the negative of the maximum likelihood function containing the prediction from Poisson regression method, $\theta^T X_{\text{test}}$. The term $(Y' - Y_{\text{ts}})^2$ in the loss function minimizes the sum of squared error between the prediction from HLF and (Y') the prediction from ARIMA methods (Y_{ts}). Hence, when we minimize the HLF, the terms $\Lambda_1(e^{\theta^T X_{\text{test}}} - Y'\theta^T X_{\text{test}}) + \Lambda_2(Y' - Y_{\text{ts}})^2$ find the Y' that includes the affect of ARIMA models into the and Poisson regression predictions. Here, the hyperparameters Λ_1 and Λ_2 are also used as regularization terms (to control bias and variance) and to

control the weight of time-series model and Poisson regression model in the HLF method.

$$\begin{aligned} -L(\theta, Y') &= e^{\theta^T X_{\text{train}}} - Y_{\text{train}} \theta^T X_{\text{train}} + \Lambda_1 (e^{\theta^T X_{\text{test}}} \\ &\quad - Y' \theta^T X_{\text{test}}) + \Lambda_2 (Y' - Y_{\text{ts}})^2 \end{aligned} \quad (10)$$

$$-\frac{\partial L(\theta, Y')}{\partial \theta} = (e^{\theta^T X_{\text{train}}} - Y_{\text{train}}) X_{\text{train}} + \Lambda_1 (e^{\theta^T X_{\text{test}}} - Y') X_{\text{train}} \quad (11)$$

$$-\frac{\partial L(\theta, Y')}{\partial Y'} = \Lambda_1 \theta^T X_{\text{train}} + 2\Lambda_2 (Y' - Y_{\text{ts}}) \quad (12)$$

The features used in the HLF method are the same as Poisson regression method as described in Eq. (1). To determine the hyperparameters Λ_1 and Λ_2 , cross-validation method is used with a training and testing set. The values of Λ_1 and Λ_2 are varied, and the performance of the resultant models is measured on the test set and compared.

4.3.4 Model selection

To select the best model for a city, the performance of the selected Poisson regression model, ARIMA model and HLF model is measured on a testing dataset. The measures of mean absolute percentage error (MAPE) and root mean squared error (RMSE) are compared. It must be noted that the training and testing data used in model selection of individual methods do not overlap with the testing data for the model selection between the three methods.

4.4 Stage 4: prediction of the gasoline shortage using the forecasted tweets

In stage 4, the number of tweets predicted in the previous stage is a predictor of the number of gasoline stations out of gas along with other features. Model selection in this stage was done using cross-validation. The data are divided into training and testing sets. Multiple models are estimated using the training data, and their performance is evaluated on the test data using MAPE and RMSE. Equation (13) describes the selected Poisson regression model. N is the dependent variable for the number of gas stations out of gas the next day, followed by all the independent variables namely, z_1 : the population of the city, z_2 : the number of gas stations, z_3 : the number of gasoline shortage tweets on the next day (predicted in stage 3), z_4 : days to arrival of the hurricane to the city and z_5 : a categorical variable for watches or warning issued in the city.

$$\log(E(N)|z) = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 \dots \alpha_5 z_5 \quad (13)$$

5 Case study

In this section, we describe the application of our methodology to predict the gasoline shortage in Florida during Hurricane Irma. While the landfall of Irma in Florida happened on the September 10, 2017, the shortage of gasoline in Florida was observed

Fig. 2 Word cloud of most frequent words in the tweet corpus of gas-related tweets



in multiple cities in the period September 6–15, 2017 (i.e., during onset and beyond landfall). The Web site Gasbuddy has the data about the percentage of gas stations out of gasoline on all these dates in all major cities of Florida (Gasbuddy 2017a). We use this data as ground truth about shortage to validate our findings.

We accessed more than 1 million tweets from Florida during this period, and the details of the tweet data are presented in Sect. 3. The National Hurricane Center (National Hurricane Centre 2017) Web site provided the data about the Hurricane path which we used as features and predictors in our model.

In stage 1, we filtered out gasoline-related tweets from the corpus of 1 million tweets. For this, we combined the tweets curated from hashtag and tweet search to filter down to 4070 relevant gasoline-related tweet. The hashtags and words that we found using the regular expressions included *gasoline*, *gas*, *gasinmiami*, *gaspricefixing*, *gasstation*, *gasservice*, *gastateparks*, *gasshortage*, *gasoil*, *gastation*, *gaswaste*, *nogas*, *outofgas*, *findgas*. We also applied the data cleaning procedure described in Sect. 4.1.

In stage 2, for tweet classification, we labeled the 4070 gasoline-related tweets. Out of the 4070 gasoline-related tweets, 2594 were gasoline shortage tweets. For classification, we extracted important unigrams on the basis of “term frequency” as described in Sect. 4.2.1. We filtered out 937 unigrams with “sum of term” frequency set to 5 in the dtm (threshold). These were the candidate features for our SVM classifier. Figure 2 shows a wordcloud of the top 50 unigrams in the tweet corpus by term frequency.

Next, we had to find “topics” to be used as features. For this step, we determined the number of topics and the best topic model among the four available models (LDA 1, LDA 2, LDA 3, CTM), using the model selection technique described in Sect. 4.2.2. We divided the document-term matrix into a training set and a testing set in the ratio of 70:30. We estimated the model with different values of k (number of topics) using the training set and measured their performance through the by “perplexity” value of the test data. Table 5 shows the comparison of their performance. The lowest perplexity value is achieved by LDA 3 models (LDA models with parameter estimation using Gibbs sampling). Further, a choice of 12 topics achieves the lowest perplexity value. These 12 topics became the candidate features for the SVM classifier.

Table 6 shows five abstract topics found using CTM described in Sect. 4.2.2 along with the top five words in each of the topics. Topic 1 is about people tweeting that they

Table 5 Perplexity measure of different topic models for different number of topics on test data

Number of topics	Perplexity (CTM)	Perplexity (LDA 1)	Perplexity (LDA 2)	Perplexity (LDA 3)
2	2.70E+36	618,944.7	620,294.5	656.9755
4	2.60E+36	621,846.1	631,295.4	619.1762
5	2.57E+36	622,815.2	637,068	609.0368
8	2.51E+36	625,236.7	655,194	591.0772
10	2.48E+36	626,552.5	667,867.9	583.622
12	2.45E+36	627,872.7	680,960.5	578.05
15	2.41E+36	629,781.3	701,214.6	580.2036
20	2.39E+36	632,543.1	638,136.2	584.5208
40	2.34E+36	639,615.3	663,342.7	627.7097
50	2.31E+36	639,811.1	685,070.5	651.7323
100	2.23E+36	659,971.7	800,349.3	787.987

Table 6 Topics identified by topic modeling techniques in the gas shortage tweet corpus

CTM				
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
gas	station	gas	gas	gas
cannot	gas	no	station	price
find	need	station	wait	high
know	hurricaneirma	line	line	got
irma	close	miami	irma	irma

cannot find gasoline due to Irma. On the other hand, topic 2 is about people tweeting that gas stations are closed and they need gas. Topic 3 is about no gasoline being there in Miami. Topic 4 is about waiting in line for gasoline because of Irma. Lastly, topic 5 is about high gasoline prices.

We had 937 unigrams and 12 topics as candidate features for the classifier. We performed model selection for the SVM classifier as described in Sect. 4.2.3 to identify the best set of features. To achieve this, we train multiple models with different sets of features (on the training data) and measure the F1 score on the test data. We divided the data into training and testing in the ratio 70:30. Next, we trained models with number of topics equal to 5 (top 5 out of the 12) and varied the number of unigrams on the basis of the sum of term frequency (threshold). Table 7 shows the performance of these models on test data. The best F1 score was achieved for five words. Next, we fixed the number of unigrams as 5 and varied the number of topics (out of the 12 candidate topics). Table 8 shows the performance of these models on test data. The best F1 score is achieved for five topics and five unigrams. The five best unigrams were gas, get, line, out and station.

Having classified the tweets, we studied the spatiotemporal dynamics of gasoline shortage tweet arrival in Florida. Figure 3 is a heat map showing the spatial distribution

Table 7 Performance of SVM using topics and unigrams (varied word frequency threshold, number of topics = 5, training/testing = 70/30)

Word frequency	Number of words	Precision	Recall	<i>F</i> score
5	937	0.941	0.714	0.811
6	797	0.961	0.788	0.866
7	710	0.950	0.762	0.846
10	519	0.969	0.771	0.859
20	282	0.963	0.767	0.854
50	109	0.972	0.775	0.862
100	38	0.972	0.779	0.865
350	5	0.985	0.789	0.876

Table 8 Performance of SVM using topics and unigrams (varied number of topics, word frequency threshold = 350, training/testing = 70/30)

Number of topics	Precision	Recall	<i>F</i> score
2	0.963	0.767	0.854
4	0.983	0.761	0.858
5	0.985	0.789	0.877
6	0.950	0.790	0.863
10	0.972	0.779	0.865
12	0.975	0.713	0.824

of gas shortage tweets in Florida in the period September 6–9, 2017. The change in tweeting behavior is clearly visible in the heat map. Until September 8, 2017, Miami was on the path of Hurricane and people were tweeting extensively about gasoline shortage as they were instructed to evacuate. However, on September 9, the hurricane changed its path and many people had evacuated Miami for the September 10 landfall. One can clearly see a reduction in gasoline shortage tweets in the Miami area on September 9. Figure 4 shows the geo-location of these tweets in four major cities of Florida. For certain tweets, a bounding box was available, shown in red, while for others an exact location was available shown by black dots.

For studying the temporal dynamics of the tweet arrival at the city level, we chose eight major cities in Florida, namely Tampa, Orlando, Jacksonville, Miami, Gainesville, Tallahassee, Naples and West Palm Beach. For each city, we grouped all the gasoline shortage tweets that came within the same hour for period of September 6–15, 2017. Figure 5 shows the frequency distribution histogram of the number of hourly tweets about gasoline shortage in the six cities. We tested whether the arrival of the tweets followed a Poisson distribution. For this, we calculated the mean value for the number of tweets arriving in an hour for each city. Using these values as arrival rates λ 's, a Poisson probability distribution was generated for each city. Next, for each city, we performed three goodness-of-fit tests to test the distribution of number tweets per hour against a sample generated Poisson distribution. We did the same tests for distribution of tweets per day for each city. We did the Chi-square test (Ch-sq), the Kolmogorov–Smirnov test (KS) and the Cramer–von Mises criterion (VM). In the

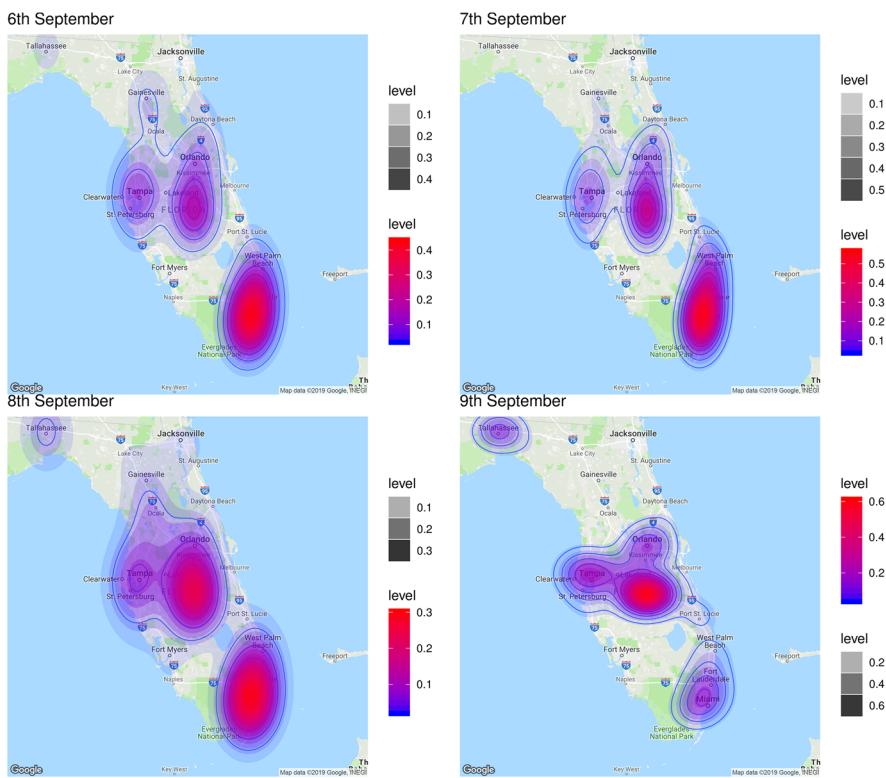


Fig. 3 Heat map of gas shortage tweets in Florida

Chi-square test, the p values were simulated by the Monte Carlo simulation method of Hope (1968). (This is advised for small reference sets.) In the Kolmogorov–Smirnov test, the p values are approximated as exact p values are not available for the two-sample case if one-sided or in the presence of ties (Conover 1971). In addition to the six cities, we also modeled the arrival of gasoline shortage tweets for the state of Florida as a whole.

Table 9 shows the results for the two Chi-squared tests for each city and the state of Florida. Assuming a significance level of 0.05 the Chi-square test fails to reject the null hypothesis for all the cities, i.e., our observations and samples from the Poisson distribution follow the same distributions. Similarly, using Von Mises criterion the arrival of our tweets follows a Poisson distribution. KS test shows that Orlando, Tallahassee, Jacksonville, Gainesville and West Palm Beach follow a Poisson distribution (as the test fails to reject the null hypothesis). Therefore, we conclude that the arrival of tweets follows a Poisson distribution. The same conclusion is drawn for the distribution of daily arrival of tweets for each city from Table 10.

We also observe that the λ values correlate with the amount of panic in a city. Miami, which was on the path of the hurricane initially and closest to landfall, has the highest value of λ , at almost 3.4 tweets per hour, indicating that people wanted to evacuate and were in desperate search for gasoline. In contrast, Jacksonville which

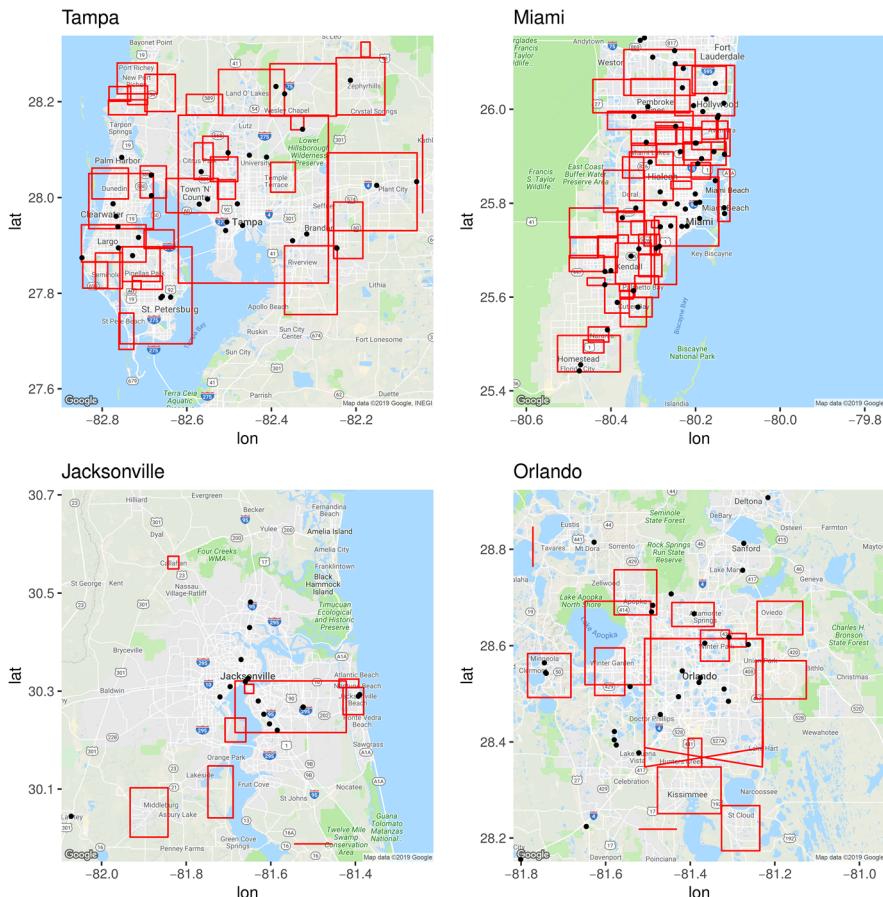


Fig. 4 Tweet locations city-wise (red box = bounding box tweets, black points = exact location tweets)

was neither on the path nor close to landfall had tweeted about gasoline shortage just 0.2 times per hour. Having observed this Poisson distribution, Poisson regression became a candidate method for predicting number of tweets in the future along with time-series methods.

As described in Sect. 4.3, in stage 3, we explore three methods for forecasting tweets about gasoline shortage : Poisson regression, time-series models, a HLF method that combines properties of Poisson regression and time-series models. First, we fit models of each kind using methodologies described in Sects. 4.3.1–4.3.3 on data for the period September 6–9, 2017. To find the best model among the three, we forecasted tweets about gasoline shortage for the period September 10–15, 2017, and compared it to ground truth.

For the fitting the ARIMA models, we use the Box–Jenkins methodology described in Sect. 4.3.2. Figure 6 shows time-series plots for four cities. For each city, in Step 1, we work on identifying the appropriate model. The time series for Miami and Tampa have a small negative trend for 6–9 September (72 h), while Orlando and Naples did

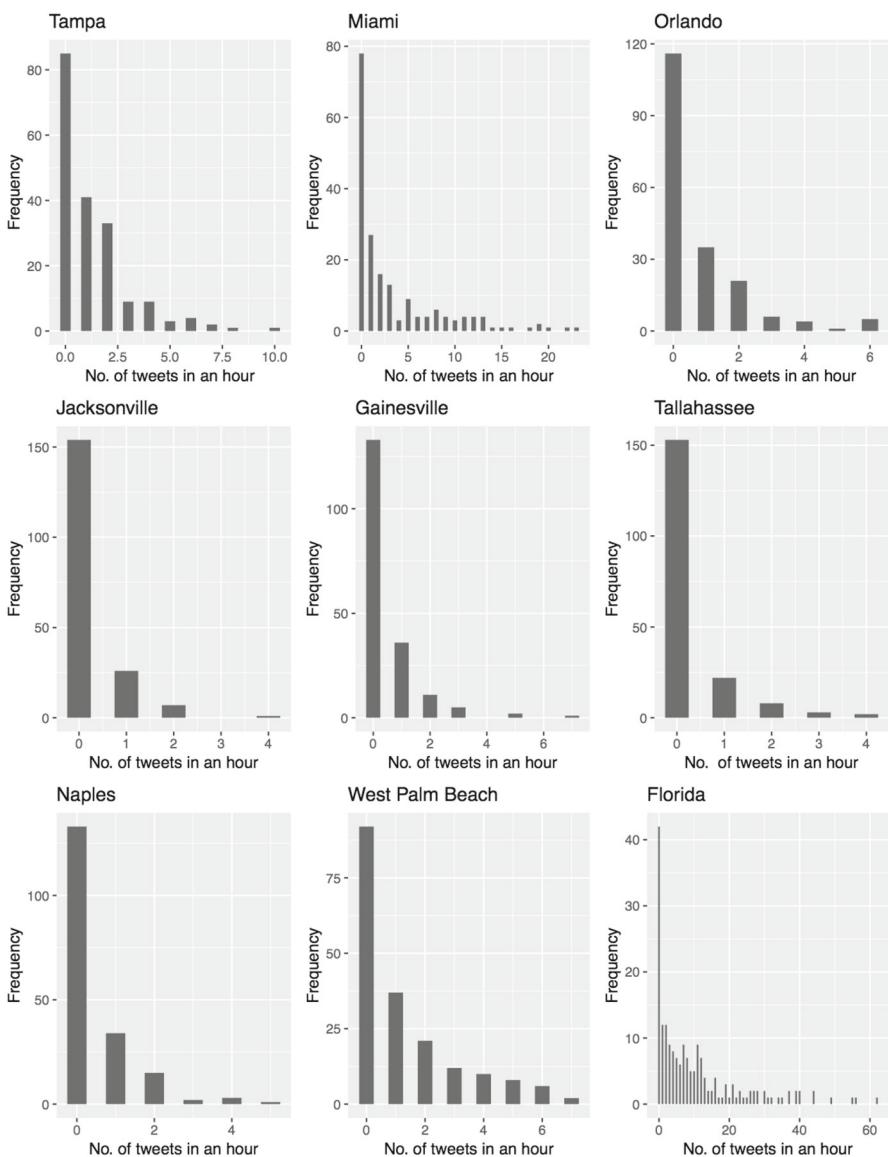


Fig. 5 Frequency of gas shortage tweets in the main cities of Florida

not. This is also evident in the autocorrelation (ACF) of the Miami time series in Fig. 7. The figure also shows the ACF and PACF of its differenced version. The differenced Miami time series does not have any trend. This is also verified by the augmented Dickey–Fuller test according to which the time series of Miami (6–9 September) is non-stationary (p value = 0.1). However, on differencing becomes stationary (p value = 0.01). Seasonality can also be inferred from the ACF for Miami time series. We tried multiple models to fit the observations seen in the ACF and PACF of the

Table 9 Goodness-of-fit tests for arrival of tweets (hourly)

City	Lambda	Chi-sq <i>p</i> value	VM <i>p</i> value	KS test <i>p</i> value
Tampa	1.281915	0.1894	0.000554025	0.0006129
Miami	3.356383	0.3073	0.000729358	2.22E-16
Orlando	0.7765957	0.1064	8.96E-05	0.09341
Tallahassee	0.2925532	0.3513	6.06E-09	0.5038
Jacksonville	0.2340426	0.4963	4.86E-09	1
Gainesville	0.4787234	0.1764	1.42E-06	0.6744
West Palm Beach	1.303191	0.2044	9.10E-04	0.01205
Naples	0.462766	0.2094	1.40E-06	0.5038
Florida	10.23936	0.1989	0.000347264	3.33E-15

Table 10 Goodness-of-fit tests for arrival of tweets (daily)

City	Chi-sq <i>p</i> value	VM <i>p</i> value	KS test <i>p</i> value
Tampa	0.2705	0.1220301	0.6994
Miami	0.2425	1.00E-01	3.36E-02
Orlando	0.1426	1.42E-01	0.3364
Tallahassee	0.3532	1.34E-01	0.6272
Jacksonville	0.3406	1.31E-01	0.6994
Gainesville	0.297	1.36E-01	0.6272
West Palm Beach	0.3675	1.22E-01	0.27
Naples	0.2578	1.49E-01	0.6272
Florida	0.2425	0.1001568	3.36E-02

Miami time series. Table 11 enlists the models and the AIC values they achieved. ARIMA((4,1,2),(1,0,0)) model with a period of 25 h (1 day) is the best fit for the Miami time series with the lowest AIC value. Residuals from the model are also normally distributed and uncorrelated (from the ACF) as shown in Fig. 8. *p* values from the Ljung–Box test are high, indicating that the autocorrelation between residuals is zero. Therefore, ARIMA((4,1,2),(1,0,0)) with period of 25 h was selected for forecasting tweets about gas shortage for Miami. For the other cities, the selected models are enlisted in Table 12.

Next, we estimated the Poisson regression models. Table 13 shows the estimation of the Poisson regression model that had the best fit for the data from 6–9 September. The column “Estimate” contains the maximum likelihood estimates of the regression coefficients for each variable under. “Asymptotic z-test” is used to determine whether the null hypothesis “that a regression coefficients is zero” can be rejected. The results of the test for each variables which included the standard error, z-score and *p* value have been tabulated. All predictors show high statistical significance for predicting the number of future gasoline shortage tweets. Null deviance = 2121.05, residual deviance = 473.84 and pseudo-*R*² = 0.78 (calculated using null and residual deviance) show

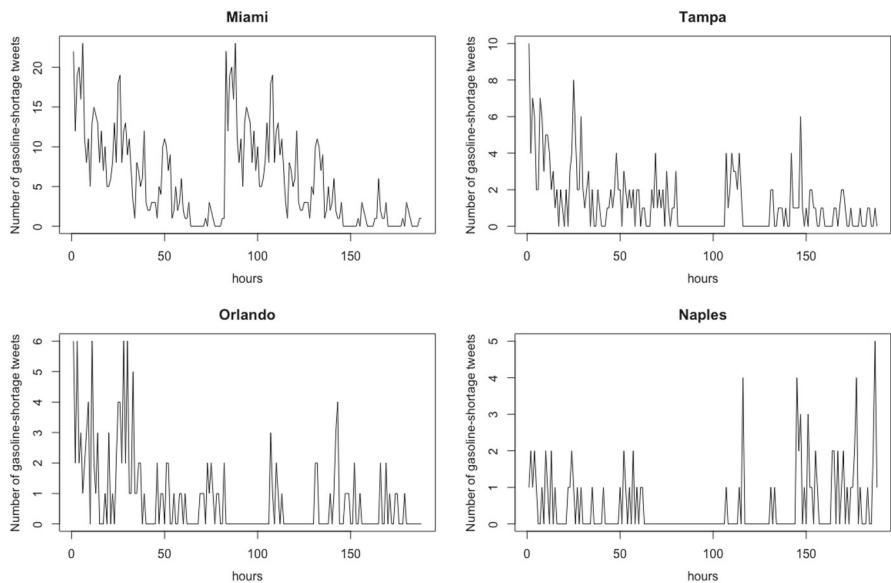


Fig. 6 Time-series plots of number of tweets about gasoline shortage in every hour in four cities

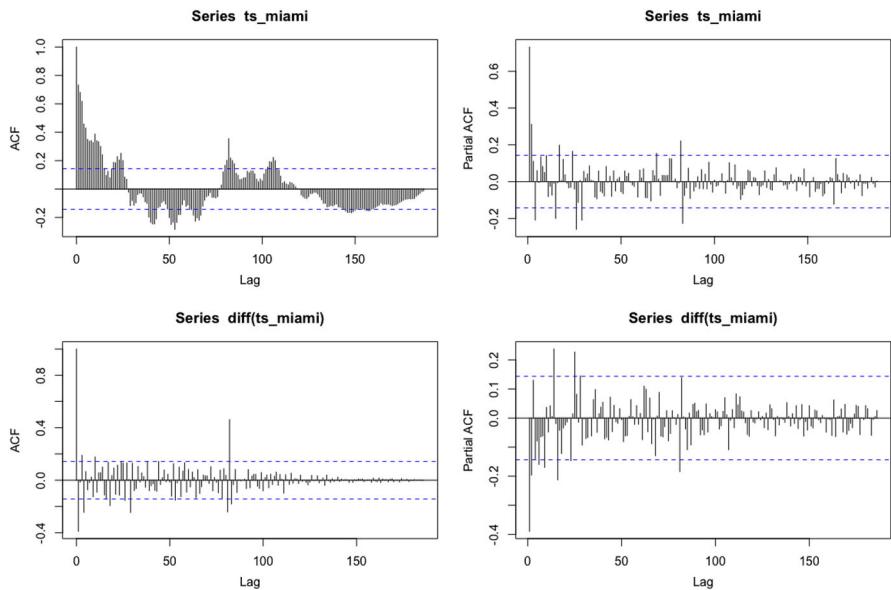
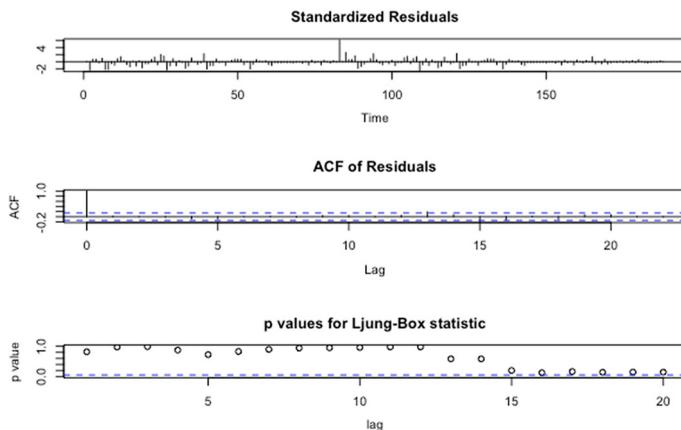


Fig. 7 Autocorrelation and partial autocorrelations for regular and differenced Miami time series

Table 11 Different time-series models for Miami tweets data and their AIC values

Model	AIC	Model	AIC
ARIMA(1,0,0) = AR(1)	1048.89	ARIMA(3,1,0)	1024.11
ARIMA(2,0,0) = AR(2)	1031.46	ARIMA(3,1,1)	1023.77
ARIMA(0,0,1) = MA(1)	1117.8	ARIMA((3,1,1),(1,0,0)) period =24	1012.12
ARIMA(0,0,2) = MA(2)	1095.47	ARIMA(4,1,2)	1004.52
ARIMA(3,0,0) = AR(3)	1029.65	ARIMA((4,1,3),(1,0,0)) period =25	1005.82
ARIMA(2,1,0)	1025.76	ARIMA((4,1,3),(1,0,0)) period =25	1002.79
ARIMA(2,1,1)	1022.19	ARIMA((4,1,2),(1,0,0)) period =25	1001.93

**Fig. 8** Residual distribution, its ACF and *p* value of Ljung–Box statistic for ARIMA((4,1,3),(1,0,0)) on Miami time series**Table 12** Time-series models selected for various cities of Florida

City	Model selected
Miami	ARIMA((4,1,2),(1,0,0))
Naples	ARIMA(0,0,2)
Jacksonville	ARIMA(0,0,4)
Tampa	ARIMA(3,1,0)
Gainesville	ARIMA(0,1,2)
West Palm Beach	ARIMA(3,0,0)
Tallahassee	ARIMA(1,0,0)

the model was a good fit. AIC = 707.31 is achieved which is the lowest among the competing models. The independent variable “gas shortage” in the table signifies the proportion of gas stations that were without gasoline on the day of the prediction. We also explored if tweets could be forecasted 2 days in advance (lag = 2 days) from the shortage and Hurricane information. Table 14 shows our results. The results showed that tweets can be predicted with lag = 2. However, variables like gas shortage and

Table 13 Results of Poisson regression model with the best fit to forecast tweets about gasoline shortage (lag = 1 day)

	Estimate	SE	<i>z</i> -value	<i>p</i> value	
(Intercept)	−5.632856917	1.342732525	−4.195069988	2.73E−05	***
Gas shortage	10.05553126	1.316275539	7.639381701	2.18E−14	***
Number of gas stations	0.006732827	0.000395052	17.04290458	3.95E−65	***
On hurricane path	0.679578601	0.142605366	4.765449025	1.88E−06	***
Inside 3-day cone	1.308414331	0.143248048	9.133906853	6.61E−20	***
Days to arrival	−1.71048558	0.148740094	−11.49982855	1.32E−30	***
Watches/warning	−5.763048725	0.418491761	−13.77099686	3.81E−43	***
Watches/warning	−3.698064077	0.265035222	−13.95310424	3.01E−44	***
Wind speeds	0.058446979	0.007923946	7.375993819	1.63E−13	***

Signif. codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ ” 1

Null deviance: 2121.05 on 63 degrees of freedom

Residual deviance: 473.84 on 54 degrees of freedom

AIC: 707.31

Table 14 Results of Poisson regression model with the best fit to forecast tweets about gasoline shortage (lag = 2 days)

	Estimate	SE	<i>z</i> -value	<i>p</i> value	Sig
(Intercept)	1.88E+00	1.342733	1.48E−01	< 2E−16	***
Gas shortage	7.88E−08	1.70E−07	4.63E−01	6.43E−01	
Number of gas stations	8.65E−06	2.06E−04	0.042	9.67E−01	
On hurricane path	7.39E−01	9.88E−02	7.4809	7.46E−14	***
Inside 3-day cone	1.49E+00	4.16E−01	3.571	3.55E−04	***
Days to arrival	−3.23E−01	4.51E−02	−7.162	7.95E−13	***
Watches/warning	−1.54E+00	2.01E−01	−7.681	1.58E−14	***
Watches/warning	−1.87E+01	1.72E−01	−10.895	< 2E−16	***
Wind speeds	5.44E−03	2.69E−03	2.021	4.33E−02	*

Signif. codes: 0 “***” 0.001 “**” 0.01 “*” 0.05 “.” 0.1 “ ” 1

Null deviance: 1021.00 on 55 degrees of freedom

Residual deviance: 561.87 on 46 degrees of freedom

AIC: 755.4

number of gas stations stopped being significant predictors. The AIC value increased and the pseudo- R^2 value dropped, indicating a worse model fit than the model with lag = 1.

Next, we estimated the best HLF model using the gradient descent algorithm. The features selected are the features of the model selected from the Poisson regression method which is described in Table 13. The data from 6–9 September were used for training data (X_{train} and Y_{train}). Data from Naples and Miami for the September 10

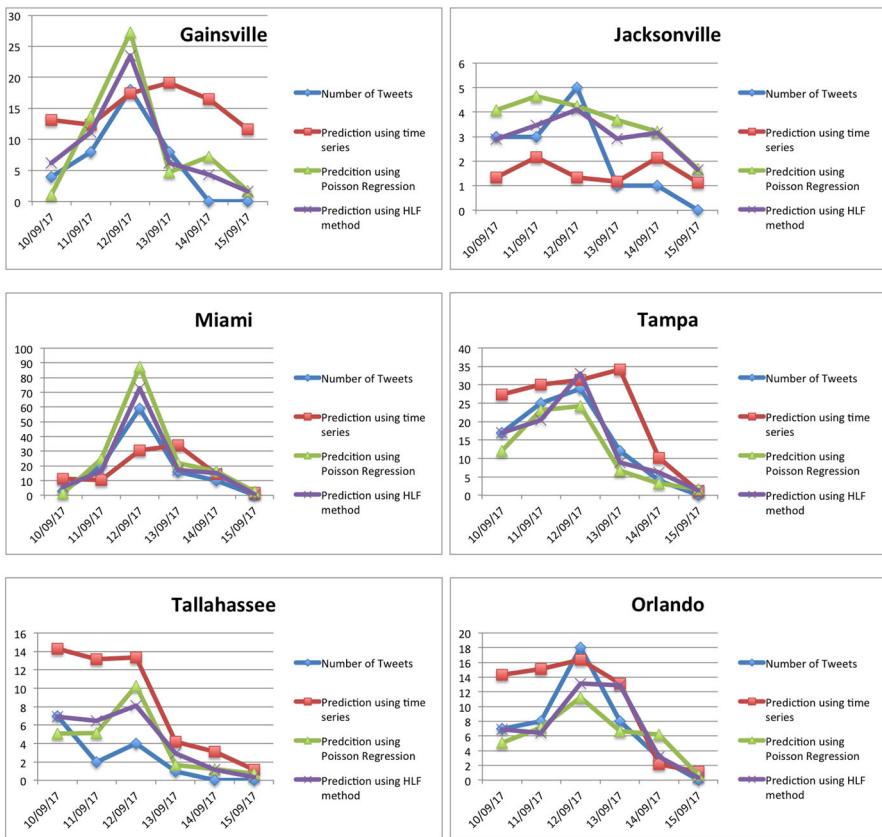


Fig. 9 Predictions and ground truth about number of tweets about gasoline shortage for six cities for September 10–15, 2017

are as testing data (X_{test}). Y_{ts} is determined by from ARIMA model predictions for Miami and Naples for 10 September. $\Lambda_1 = \Lambda_1 = 1$ achieves the smallest RMSE value on the test data. The gradient descent algorithm converges to an optimum fastest at a learning rate, $\omega_1 = \omega_2 = 10^{-5}$.

Next, the procedure described in Sect. 4.3.4 for selecting the best model from the three methods is employed. Prediction are made using the three methods for all the city on the testing data (10–15 September). Figure 9 shows the comparison of the performance of the all the three methods against the ground truth (about number of tweets) in six cities for the dates September 10–15, 2017. In all the cities, the purple line representing prediction by HLF method is closest to the blue line representing ground truth. Figure 10 shows the comparison of the overall MAPE and RMSE for the three methods. Clearly, HLF method is superior with the smallest MAPE and RMSE values.

In stage 4, we predict the number of stations out of gasoline the next day by using a Poisson regression (for the period September 13–15, 2017, for eight cities). To find the model with the best fit, we use the cross-validation technique described in

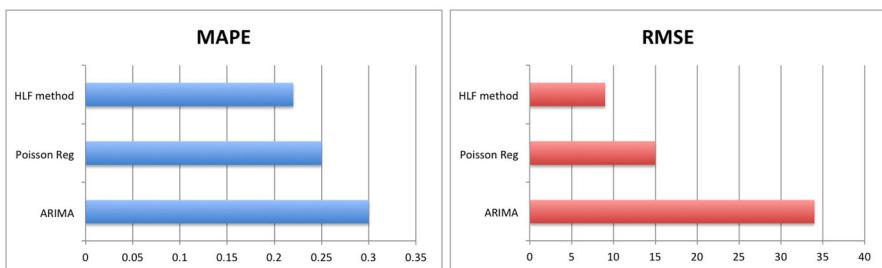


Fig. 10 MAPE and RMSE of the HLF, Poisson regression and ARIMA models)

Sect. 4.4. Data from 6–12 September are used as training data, and the data from 13–15 September are used as a testing set. Table 15 shows the estimation of the Poisson regression model which had the best fit and achieved the lowest MAPE and RMSE on the test set. It tabulates the maximum likelihood estimates of regression coefficients and the standard error, z-score and p value from the z-test. All predictors are statistically significant. Null deviance = 51,961.01, residual deviance = 689.17 and pseudo- R^2 = 0.987 show the model is a very good fit. On the test data, MAPE = 0.31 and RMSE = 9.13 are achieved. Figure 11 shows the comparison of the predictions on the test set with the ground truth.

We also wanted to explore that how much information tweets provide for predicting the shortage. For this, we trained a model without the data about number of tweets. The results are given in Table 16. Null deviance = 51,961.01 and residual deviance = 8758.68. Its residual deviance is larger than the original model, indicating that the original model is a better fit. Pseudo- R^2 value for this model is 0.831 as compared to 0.987 of the original model. This means that the original model with tweet information explains gasoline shortage 18.71 percent more than the model without tweet information. We also verified how can shortage be predicted using tweets and hurricane information from 1 day prior (lag = 1 day). Table 17 shows the results. Results show that the estimates of the coefficients are non-zero with statistical significance which means that the variables do have some predictive power in forecasting shortage of gasoline on the next day. However, the pseudo- R^2 of the model is 0.904, 8 percent less than 0.987 (the pseudo- R^2 of the original model with lag = 0)

6 Using social media data to drive decision-making models in the gas shortage domain

In our paper, we develop a coarse grain prediction of gasoline shortage, in that it only predicts the proportion of stations without gas in the city. If we had access to the ground-truth data at the individual gas station level, a finer grain prediction model that predicts gasoline shortage at individual gas station level could be validated. In the rest of this section, we assume that gasoline shortage predictions are available at the individual gas station level and outline the development of two key decision-making models, one related to supply of gasoline (by authorities) and the other related to search for gasoline (by individuals).

Table 15 Results of Poisson regression model with the best fit to predict gasoline shortage (lag = 0 day)

	Estimate	SE	<i>z</i> -value	<i>p</i> value	Sig
(Intercept)	4.255101001	0.056026652	75.94780113	0	***
Population	-1.18E-06	1.35E-07	-8.700698039	3.30E-18	***
Number of gas stations	0.00310295	0.000121362	25.56764236	3.50E-144	***
Number of tweets	0.002997528	0.000281172	10.66083059	1.55E-26	***
Days to arrival	-0.137963866	0.020294006	-6.798256761	1.06E-11	***
Warning	-0.20750846	0.049436483	-4.19747623	2.70E-05	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null deviance: 5961.01 on 71 degrees of freedom

Residual deviance: 689.17 on 60 degrees of freedom

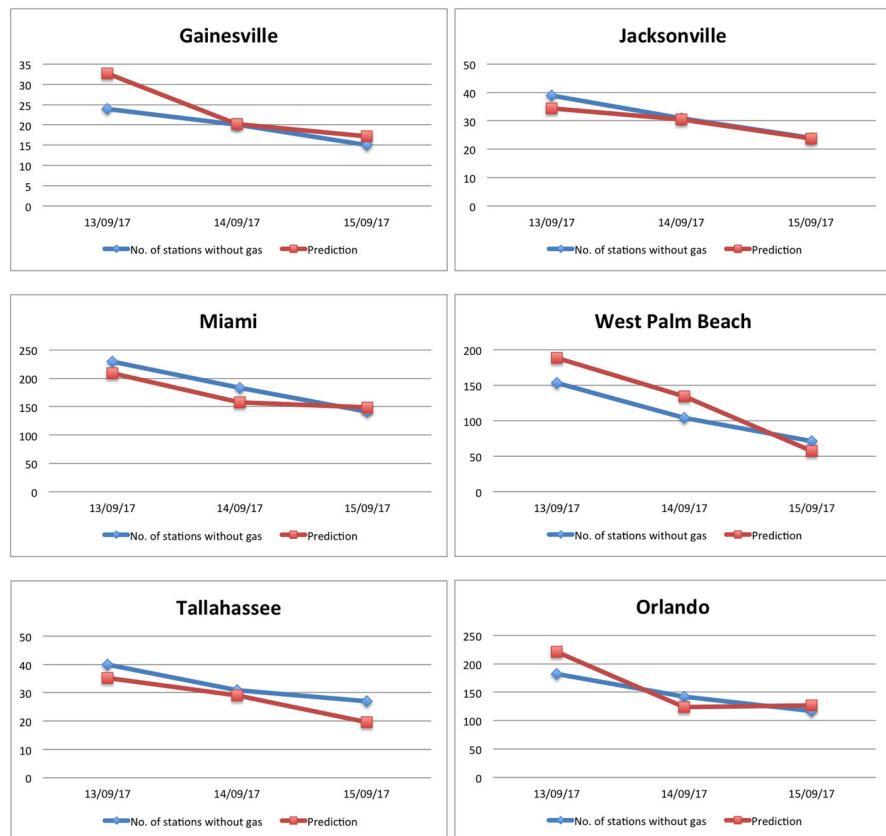
**Fig. 11** Predictions and ground truth about gasoline shortage for six cities for September 12–15, 2017

Table 16 Results of Poisson regression model with the best fit to predict gasoline shortage without tweets information (lag = 0 day)

	Estimate	SE	<i>z</i> -value	<i>p</i> value	Sig
(Intercept)	3.75E+00	2.73E-02	137.37	< 2E-16	***
Population	2.64E-07	5.28E-08	5.002	5.67E-07	***
Number of gas stations	0.00310295	0.000121362	25.56764236	3.50E-144	***
Days to arrival	0.2374	2.22E-02	10.714	< 2E-16	***
Warning	-1.01E-02	6.78E-04	-14.854	< 2E-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 5961.01 on 71 degrees of freedom

Residual deviance: 8758.68 on 61 degrees of freedom

Table 17 Results of Poisson regression model with the best fit to predict gasoline shortage (lag = 1 day)

	Estimate	SE	<i>z</i> -value	<i>p</i> value	Sig
(Intercept)	3.64E+00	3.03E-02	120.205	< 2E-16	***
Population	-3.56E-07	6.00E-08	-5.932	2.99E-09	***
Number of gas stations	2.83E-03	5.85E-05	48.46	< 2E-16	***
Number of tweets	2.20E-03	2.95E-04	7.442	9.90E-14	***
Days to arrival	-2.26E-01	2.16E-02	-10.43	< 2E-16	***
Warning	8.76E-03	7.57E-04	11.565	< 2E-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Poisson family taken to be 1)

Null deviance: 5445.45 on 63 degrees of freedom

Residual deviance: 521.64 on 52 degrees of freedom

6.1 Supply of gasoline

Analysis of Twitter data can yield either probabilistic inference of individual gas station shortages or deterministic inference, depending on which set of analytical methods is used. If gasoline shortage at individual stations is known on a probabilistic basis, the resultant vehicle routing problem can be modeled using a prize collection methodology, where the prize for a visiting a gas station is larger if it has a higher likelihood of a shortage. It could also be modeled using stochastic programming methods as applied to VRP approaches. If the gasoline shortage is known on a deterministic basis, the vehicle routing can be modeled using traditional VRP approaches.

6.2 Search for gasoline

An individual searching for gasoline presents an interesting modeling situation, because they start with some level of gasoline and have limited travel capability in the search process. Also, there can be significant waiting lines at a gas station that has gasoline and the consumer has to decide whether to wait (if they run out of gas while

waiting they would simply push their car in line) or travel to another gas station to seek gas. This can be modeled using a dynamic programming framework, in which each gas station can be in a variety of states with known probability. One of these states is the “no gas” state and the other states all have “gas” but different amounts of waiting time to obtain the gas.

In our upcoming work, we are building a model that can estimate the probability of gasoline shortage at an individual gas station using the spatiotemporal distribution of gasoline shortage tweets. This will provide needed data for both the models for supply of gasoline and search for gasoline.

7 Conclusions and future research

Our methodology helps answer two major questions. First, can social media be used to predict gasoline shortage during disasters and, second, what is a good methodology to make such a prediction. People tweet and use social media during emergencies. Hence, we believe this methodology can be generalized for other applications like predicting shortage of other commodities during and forecastable emergencies. Our methodology produces very accurate results for the case of gasoline shortage during Hurricane Irma in Florida in 2017. In particular, the method with HLF to predicts the number of future gasoline shortage tweets with high accuracy. ARIMA models successfully capture time-related covariance between number of tweets, while the Poisson regression captures the variation in number of tweets due to gasoline shortage and other variables that cause panic. The HLF model successfully combines these two properties and hence achieves more accurate results. For the gasoline shortage prediction, our model achieves MAPE = 0.31 and RMSE = 9.13.

In the future, there are several fruitful directions for future research. Our first suggested future research direction stems from the recognition that although the F1 score in the classification model is reasonably good, the recall values could be improved by decreasing the relatively high number of false positives. We believe that this can be improved by further analysis of stage 2 (Classification) model. Our second future research direction stems from the fact that our method does a coarse grain prediction of gasoline shortage, in that it only predicts the proportion of stations without gas in the city. If we had access to the ground truth data at individual gas station level, a finer grain prediction model that predicts gasoline shortage at individual gas station level could be validated. Therefore, in our upcoming work, we are building a model that can estimate the probability of gasoline shortage at an individual gas station using the spatiotemporal distribution of gasoline shortage tweets. Our third future research direction relies on successful completion of the second future research task. Once future shortage data are available at the individual gas station level, they can be fed into a decision-making model for gasoline delivery to gas stations to ensure adequate supply where it is needed. This would likely be a vehicle routing type of formulation.

Acknowledgements The authors would like to thank two anonymous referees who provided detailed comments that significantly enhanced our paper.

Funding Funding was provided by National Science Foundation (Grant No. 1663101).

References

- Ashktorab Z, Brown C, Nandi M, Culotta A (2014) Tweedr: mining twitter to inform disaster response. In: ISCRAM 2014 conference proceedings—11th international conference on information systems for crisis response and management (May), pp 354–358. <https://doi.org/10.1145/1835449.1835643>, <http://www.scopus.com/inward/record.url?eid=2-s2.0-84905845531&partnerID=40&md5=ee57e6c3d9498b083428cdae67d83396>
- Atefeh F, Khreich W (2015) A survey of techniques for event detection in twitter. *Comput Intell* 31(1):132–164
- Beigi G, Hu X, Maciejewski R, Liu H (2016) An overview of sentiment analysis in social media and its applications in disaster relief. In: Pedrycz W, Chen SM (eds) *Sentiment analysis and ontology engineering. Studies in Computational Intelligence*, vol 639. Springer, Cham, pp 313–340. https://doi.org/10.1007/978-3-319-30319-2_13
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3(Jan):993–1022
- Blei DM, Lafferty JD et al (2007) A correlated topic model of science. *Ann Appl Stat* 1(1):17–35
- Boulos MNK, Resch B, Crowley DN, Breslin JG, Sohn G, Burner R, Pike WA, Jezierski E, Chuang KYS (2011) Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *Int J Health Geogr* 10(1):67
- Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*. Wiley, Hoboken
- Brockwell PJ, Davis RA, Calder MV (2002) *Introduction to time series and forecasting*, vol 2. Springer, Berlin
- Cadenas E, Rivera W (2010) Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model. *Renew Energy* 35(12):2732–2738
- Caragea C, Squicciarini A, Stehle S, Neppalli K, Tapia A (2014) Mapping moods: geo-mapped sentiment analysis during hurricane sandy. In: ISCRAM 2014 conference proceedings—11th international conference on information systems for crisis response and management (May), pp 642–651. <http://www.iscram.org/legacy/ISCRAM2014/papers/p29.pdf>
- Cheng Z, Caverlee J, Lee K (2010) You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM international conference on information and knowledge management. ACM, pp 759–768
- Chowdhury R, Chowdhury SR, Castillo C (2013) Tweet4act : using incident-specific profiles for classifying crisis-related messages. In: Proceedings of the 10th international ISCRAM conference (May), pp 834–839
- Conover WJ (1971) *Practical nonparametric statistics*. Wiley, New York, pp 295–301
- Cordeiro M, Gama J (2016) Online social networks event detection: a survey. In: Solving large scale learning tasks. Challenges and algorithms. Springer, Cham, pp 1–41. https://doi.org/10.1007/978-3-319-41706-6_1
- Faulkner M, Olson M, Chandy R, Krause J, Chandy KM, Krause A (2011) The next big one: detecting earthquakes and other rare events from community-based sensors. In: 2011 10th international conference on information processing in sensor networks (IPSN). IEEE, pp 13–24
- Fdot (2017) Hurricane IRMA report by Florida department of transportation. http://www.fdot.gov/info/CO/news/newsreleases/020118_FDOT-Fuel-Report.pdf
- Feinerer I (2008) An introduction to text mining in R. *Newslett R Proj* 8/2:19
- Fessenden H (2017) Price gouging. https://www.richmondfed.org/-/media/richmondfedorg/publications/research/econ_focus/2017/q4/jargon_alert.pdf
- Flood R (2017) Express UK website. <https://www.express.co.uk/news/weather/850222/Hurricane-Irma-path-destruction-USA-Florida-panic-buying-storm>
- Gasbuddy (2017a) <https://business.gasbuddy.com/hurricane-irma-live-updates-fuel-availability-station-outages/>
- Gasbuddy (2017b) <https://tracker.gasbuddy.com/?q=Buffalo,%20NY>

- Gaynor M, Seltzer M, Moulton S, Freedman J (2005) A dynamic, data-driven, decision support system for emergency medical services. In: International conference on computational science. Springer, pp 703–711
- Geman S, Geman D (1987) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In: Readings in computer vision. Elsevier, pp 564–584
- Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proc Natl Acad Sci* 101(suppl 1):5228–5235
- Gu S, Pan C, Liu H, Li S, Hu S, Su L, Wang S, Wang D, Amin T, Govindan R, et al (2014) Data extrapolation in social sensing for disaster response. In: 2014 IEEE international conference on distributed computing in sensor systems (DCOSS). IEEE, pp 119–126
- Gupta A, Lamba H, Kumaraguru P, Joshi A (2013) Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd international conference on World Wide Web. ACM, pp 729–736
- Han B, Cook P, Baldwin T (2013) A stacking-based approach to twitter user geolocation prediction. In: Proceedings of the 51st annual meeting of the association for computational linguistics: system demonstrations, pp 7–12
- Hoffman M, Bach FR, Blei DM (2010) Online learning for latent dirichlet allocation. In: Advances in neural information processing systems, pp 856–864
- Hope AC (1968) A simplified Monte Carlo significance test procedure. *J R Stat Soc: Ser B (Methodological)* 30(3):582–598
- Hornik K, Grün B (2011) topicmodels: an R package for fitting topic models. *J Stat Softw* 40(13):1–30
- Hughes AL, St Denis LA, Palen L, Anderson KM (2014) Online public communications by police & fire services during the 2012 hurricane sandy. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1505–1514
- Imran M, Elbassouoni S, Castillo C, Diaz F, Meier P (2013) Practical extraction of disaster-relevant information from social media. In: Proceedings of the 22nd international conference on World Wide Web. ACM, pp 1021–1024
- Imran M, Castillo C, Lucas J, Meier P, Vieweg S (2014) AIDR: Artificial intelligence for disaster response. In: Proceedings of the companion publication of the 23rd international conference on World Wide Web companion (October), pp 159–162. <https://doi.org/10.1145/2567948.2577034>. https://mimran.me/papers/imran_castillo_lucas_meier_vieweg-www2014.pdf
- Kaigo M (2012) Social media usage during disasters and social capital: Twitter and the great East Japan earthquake. *Keio Commun Rev* 34(1):19–35
- Ki EJ, Nekmat E (2014) Situational crisis communication and interactivity: usage and effectiveness of Facebook for crisis management by fortune 500 companies. *Comput Hum Behav* 35:140–147
- Kumar S, Barbier G, Abbasi MA, Liu H (2011) Tweettracker: an analysis tool for humanitarian and disaster relief. In: Fifth international AAAI conference on weblogs and social media
- Lachlan KA, Spence PR, Lin X (2014) Expressions of risk awareness and concern through Twitter: on the utility of using the medium as an indication of audience needs. *Comput Hum Behav* 35:554–559. <https://doi.org/10.1016/j.chb.2014.02.029>
- Lee S, Song J, Kim Y (2010) An empirical comparison of four text mining methods. *J Comput Inf Syst* 51(1):1–10
- Liu BF, Fraustino JD, Jin Y (2016) Social media use during disasters: how information form and source influence intended behavioral responses. *Commun Res* 43(5):626–646. <https://doi.org/10.1177/0093650214565917>
- Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we RT?. In: Proceedings of the first workshop on social media analytics. ACM, pp 71–79
- Meyer D, Hornik K, Feinerer I (2008) Text mining infrastructure in R. *J Stat Softw* 25(5):1–54
- Morstatter F, Lubold N, Pon-Barry H, Pfeffer J, Liu H (2014) Finding eyewitness tweets during crises. [arXiv:1403.1773](https://arxiv.org/abs/1403.1773)
- National Hurricane Centre (2017) National hurricane centre website. <https://www.nhc.noaa.gov>
- Nazer TH, Xue G, Ji Y, Liu H (2017) Intelligent disaster response via social media analysis a survey. *ACM SIGKDD Explor News* 19(1):46–59
- Ni M, He Q, Gao J (2017) Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans Intell Transp Syst* 18(6):1623–1632
- Nie H, Liu G, Liu X, Wang Y (2012) Hybrid of ARIMA and SVMS for short-term load forecasting. *Energy Procedia* 16:1455–1460

- Olteanu A, Castillo C, Diaz F, Vieweg S (2014) CrisisLex: a lexicon for collecting and filtering microblogged communications in crises. In: Proceedings of the 8th international conference on weblogs and social media, p 376. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8091/8138>
- Pai PF, Lin CS (2005) A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* 33(6):497–505
- Panagiotopoulos P, Barnett J, Bigdeli AZ, Sams S (2016) Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technol Forecast Soc Change* 111:86–96. <https://doi.org/10.1016/j.techfore.2016.06.010>
- Phan XH, Nguyen LM, Horiguchi S (2008) Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In: Proceedings of the 17th international conference on World Wide Web. ACM, pp 91–100
- Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71(3):599–607
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World Wide Web. ACM, pp 851–860
- Sampson J, Morstatter F, Zafarani R, Liu H (2015) Real-time crisis mapping using language distribution. In: 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, pp 1648–1651
- Schulz A, Hadjidakos A, Paulheim H, Nachtwey J, Mühlhäuser M (2013) A multi-indicator approach for geolocalization of tweets. In: Seventh international AAAI conference on weblogs and social media, pp 573–582
- Starbird K, Stamerger J (2010) Tweak the tweet: leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In: Proceedings of the 7th international ISCRAM conference, information systems for crisis response and management Seattle, WA, vol 1, pp 1–5
- Stowe K, Paul MJ, Palmer M, Palen L, Anderson K (2016) Identifying and categorizing disaster-related tweets. In: Proceedings of The fourth international workshop on natural language processing for social media, pp 1–6
- Štířeský V, Stránská A, Drábik P (2015) Crisis communication on facebook. *Studia Commercialia Bratislavensis* 8(29):103–111
- Tien Nguyen D, Mannai KAA, Joty S, Sajjad H, Imran M, Mitra P (2016) Rapid classification of crisis-related data on social networks using convolutional neural networks. [arXiv:1608.03902](https://arxiv.org/abs/1608.03902)
- Tseng FM, Yu HC, Tzeng GH (2002) Combining neural network model with seasonal time series ARIMA model. *Technol Forecast Soc Change* 69(1):71–87
- Ushahidi (2017) Ushahidi. <https://www.ushahidi.com>
- Utz S, Schultz F, Glocka S (2013) Crisis communication online: how medium, crisis type and emotions affected public reactions in the Fukushima Daiichi nuclear disaster. *Public Relat Rev* 39(1):40–46
- van Gorp A, Pogrebnyakov N, Maldonado E (2015) Just keep tweeting: emergency responder's social media use before and during emergencies. In: Proceedings of the 23rd European conference on information systems (ECIS 2015), pp 1–15. <https://doi.org/10.18151/7217512>
- Wainwright MJ, Jordan MI et al (2008) Graphical models, exponential families, and variational inference. *Found Trends Mach Learn* 1(1–2):1–305
- Waze (2017) Waze. <https://www.waze.com>
- Xu Q, Tsui KL, Jiang W, Guo H (2016) A hybrid approach for forecasting patient visits in emergency department. *Qual Reliab Eng Int* 32(8):2751–2759
- Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175
- Zhu B, Wei Y (2013) Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega* 41(3):517–524
- Zook M, Graham M, Shelton T, Gorman S (2010) Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Med Health Policy* 2(2):7–33

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.