DUET: Data-Driven Approach Based on Latent Dirichlet Allocation Topic Modeling

Yan Wang, A.M.ASCE[1]; and John E. Taylor, M.ASCE[2]

Abstract: Social networking platforms have been widely employed to detect and track physical events in population-dense urban areas. They can be effective tools to understand what happens and when and where it happens, either retrospectively or in real time. Correspondingly, a variety of approaches have been proposed for detecting either targeted or general events. However, neither type of event detection technique has been developed to detect urban emergencies that happen in specific geographic locations and with unpredictable characteristics. Therefore, we propose a spatial and data-driven detecting urban emergencies technique (DUET) for natural hazards, manmade disasters, and other emergencies. The method addresses both geographic and semantic dimensions of events using a geotopic detection module and evaluates their crisis levels on the basis of the intensity of negative sentiment through a ranking module. DUET was designed specifically for georeferenced tweets from a Twitter streaming application programming interface (API). To validate the technique, we conducted multiple experiments with geotagged tweets in different urban environments over a period of four to six consecutive hours. DUET successfully identified emergencies of different types among all the candidate geotopics. Our future work focuses on enabling online-mode detection with high scalability with large volumes of streaming data and providing interactive visualization through a GIS system. DUET can identify emergencies of general types and provide timely emergency reports both to first responders and to the public. The technique contributes to building an efficient and open disaster information system through a crowdsourcing effort and adding agility to urban resilience regarding crisis detection, situation awareness, and information diffusion. **DOI:** 10.1061/(ASCE)CP.1943-5487.0000819. *© 2019 American Society of Civil Engineers.*

Author keywords: Disaster informatics; Emergency detection; Latent Dirichlet allocation (LDA) topic modeling; Natural language processing; Sentiment analysis; Social media.

Introduction

Recent natural hazards and manmade disasters have had a devastating impact on human life and the environment. In 2017 alone, there were a series of hurricanes in the Atlantic coastal area (e.g., hurricanes Harvey, Irma, Maria, and Nate); destructive California wildfires; and severe infrastructure disasters such as the Grenfell Tower fire in London, and the Interstate 85 collapse in Atlanta. These extreme events have reminded disaster managers and the public of the importance of timely emergency communication and revealed the shortcomings in our ability to effectively detect crises, alert the population, and provide assistance (National Academies of Sciences, Engineering, and Medicine 2017).

In the United States, current emergency reporting and warning rely mainly on the integrated public alert and warning system and the wireless emergency alerts system, with calling 911 as the preferred approach for reporting emergencies. However, under some circumstances, people cannot make calls, the crisis is too time-critical for a 911 call to be effective, or calls have been made but the emergency responders are unable to assess the relative gravity of one crisis over another to deploy first responders efficiently. In a recent hurricane, Harvey, when emergency telephone hot lines were jammed, victims turned to social media for help, even posting their full addresses in desperation (Seetharaman and Wells 2017). Thus, it is necessary to incorporate these new communication platforms into existing disaster and emergency information management approaches and systems.

Additionally, a successful disaster management system should incorporate six capabilities: identification, prediction, mitigation, preparation, response, and recovery (Pradhan et al. 2007) however, many current studies have focused narrowly on the warning phase (Ghosh et al. 2013) and postdisaster phase (Peña-Mora et al. 2010, 2012). Few studies have addressed the identification phase of emergencies. We focus on new communication platforms and in this research develop a technique for detecting geographically constrained urban emergencies in order to supplement current disaster information management approaches and systems in terms of openness, effectiveness, and efficiency.

Newly available massive sets of data from geosocial platforms (e.g., Twitter, Facebook, Foursquare, and Instagram) have played an increasingly important role at different stages of disaster and emergency management (Ford et al. 2016). This crowdsourcing data has brought new opportunities in understanding human mobility patterns (Gonzalez et al. 2008), examining dynamic spatial networks (Wang and Taylor 2017), and tracking a population's sentiment (Kryvasheyeu et al. 2016; Wang and Taylor 2018) in the context of a disaster. These geosocial platforms have also played an increasingly critical role in early warning, monitoring, and evaluation of emergent events. It is especially crucial to enable early detection through social media because timely detection of emergencies can facilitate more immediate responses, may provide

[1]Assistant Professor, Dept. of Urban and Regional Planning, Florida Institute for Built Environment Resilience, Univ. of Florida, Gainesville, FL 32611. Email: yanw@ufl.edu

[2]Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, Mason 4140c, Atlanta, GA 30332 (corresponding author). Email: jet@gatech.edu

Note. This manuscript was submitted on March 20, 2018; approved on September 11, 2018; published online on March 13, 2019. Discussion period open until August 13, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

information to reduce potential casualties and damage, and may lead to more effective resource allocation (Li et al. 2017). Early detection also contributes to characterizing an event in terms of spatiotemporal scale, collective emotions, semantic topics, and the dynamic evolution of emergencies over time.

Among a variety of geosocial networking platforms, Twitter is especially suitable for emergency environments in terms of its open design, wide usage, geoenabled function, and limited message length environment (Kryvasheyeu et al. 2016). Georeferenced tweets can document geographical locations and collective reactions to crises unfolding in both spatial and temporal scales. A number of studies concerning event detection techniques in the context of Twitter were published in recent years. Some studies focused on targeted events with supervised methods (Sakaki et al. 2010; Sun et al. 2016), whereas others intended to identify general events that rapidly escalate in contents, time, and space (Maurya et al. 2016; Xie et al. 2016; Yu et al. 2017; Zhang et al. 2016). These proposed detection techniques can be built on the basis of clustering, supervised classification, latent Dirichlet allocation (LDA), or hybrid means. However, few have explored the context of urban disasters and emergencies (e.g., infrastructure failures, building fires, and natural hazards).

Compared with other events, urban disasters and emergencies can be regarded as more targeted but also as unpredictable in type and form. It is difficult to employ current supervised techniques designed for targeted events to identify an uncharacterized emergency. Moreover, detection approaches for general events have not stressed the distinct characteristics of disasters and emergencies in terms of their geographical and thematic impact as well as their high intensity of negative sentiment. Therefore, to address this methodological gap, we propose a data-driven technique to detect urban emergencies, with a focus on geotagged tweets from a Twitter streaming API. We describe the approach as the detecting urban emergencies technique (DUET) to highlight its specification in detecting emergency events that happen in the confined physical locations of cities and to describe the symbiotic sociotechnical relationship that can exist between first responders and the public during emergencies. The proposed detection technique can inform resource-constrained disaster managers and first responders of when and where a potential emergency exists, the emergency details, and the collective sentiment level. Such a technique could enable more intelligent and integrated disaster informatics and more agile predisaster, during-disaster, and postdisaster management and could contribute to building more resilient communities in the context of extreme events.

Background

Event detection using Twitter streams has been addressed in a mounting number of publications in recent literature. We classified recent approaches on the basis of their detection objectives into two types: targeted event detection and general event detection. Existing techniques for general event detection were either retrospective or real-time. We exclusively discuss the latter because most real-time methods were built on the basis of retrospective methods and our final goal is to enable real-time event detection.

Targeted Event Detection

Targeted event detection requires predefined keywords and adopts mainly supervised detection techniques. For example, Sakaki et al. (2010) proposed a targeted event detection system that monitored tweets and delivered prompt notifications. Their system was specifically applied in reporting earthquakes with tweets in Japan. They first devised classifiers to classify event-related tweets and unrelated tweets. Then the related tweets were used to develop a probabilistic spatiotemporal model for event detection and location/trajectory estimation. In other work, Sun et al. (2016) designed a novel method to detect and locate power outages from Twitter. The system was based on a heterogeneous information network, which included time, locations, and texts. Supervised LDA was then used to compute the probability of the topics of tweets that were related to a power outage. Gu et al. (2016) proposed a real-time traffic incident (TI) detection approach based on tweet texts. Each imported tweet was mapped into a binary vector of a dictionary and classified as TI-related or not. The TI-related tweets were further geocoded and classified into different incident categories.

Detection techniques for distinct targeted events are effective in identifying specific events with pre-envisioned and predefined characteristics. However, due to the diversity of the types of urban disasters, event detection through predefined characteristics may require a large volume of keywords to describe different types of potential events, not to mention unexpected types of events. Therefore, with the lack of predefined characteristics, it is impractical to employ a supervised approach to detect general and unknown disasters.

General Event Detection

Clustering-Based Approaches

Clustering-based approaches include threshold-based online approaches, graph-based clustering algorithms, and other new approaches. For example, Yu et al. (2017) proposed a real-time emerging anomaly monitoring system over microblog text streams, named RING. The system was based on a graph stream model. It was able to detect events at an early stage, to conduct correlation analysis between emerging events, and to track evolution of events over time. Specifically, the graph regarded keywords as nodes, their co-occurrence in each tweet as edges, and an accumulated frequency as weights of edges. A k-clique percolation method was then employed to identify communities (events) in the built graph. SigniTrend is a scalable detection technique developed by Schubert et al. (2014) that measured significance of terms to detect trending words on the basis of their co-occurrences using a hashing technique to track all the keyword pairs. The final stage of this approach was to cluster the detected keywords into larger topics. This method was used to detect emerging topics before they become hot tags. Some techniques included geolocation as a main dimension to capture real-world occurrences. EvenTweet (Abdelhaq et al. 2013) identified localized events using geotagged tweets. It extracted keywords based on the burstiness degree of words and then computed the spatial density distribution (spatial signature) over a keyword in a spatial grid. The event keywords were further partitioned on the basis of the cosine similarity of their spatial signatures. Finally, the clusters were scored to uncover the real-world local events. GeoBurst (Zhang et al. 2016) was also designed to extract local events from streams of geotagged tweets in real time. It identified candidate events based on the geographical and semantic impacts between each pair of tweets and ranked the candidates according to their spatial and temporal burstiness. However, most of the clustering-based approaches used co-occurrences of keywords to measure the semantic relationship between documents and, as such, they could not reveal the latent structure of topics underlying the text corpora.

J. Comput. Civ. Eng., 2019, 33(3): 04019023

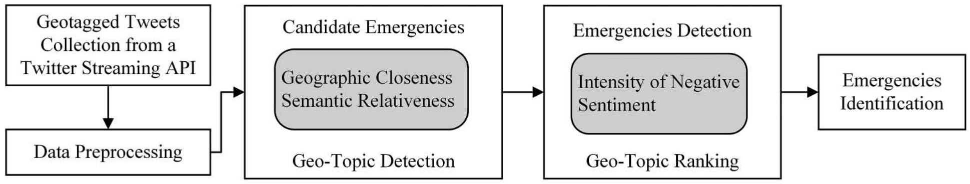

Fig. 1. Detecting urban emergencies technique.

LDA Topic Model–Based Methods

Latent Dirichlet allocation is a basic probabilistic topic model that analyzes the words from a sample of text to reveal the underlying themes and their connections (Blei 2012). Recently, researchers have explored the advantages of LDA in allowing the examination of multiple topics within a document and generating a probabilistic distribution of words under a topic. This has been carried out using LDA as a basis for extracting thematic content from social networks for event detection. For example, semantic scan (Maurya et al. 2016) was a contrastive topic-modeling approach based on LDA to identify new topics in a text stream; it then used a statistical scan to find the spatially localized events. The technique has been tested on Yelp and emergency department data sets; the moving window size was 3 days, a long time for detecting emergencies. Moreover, the method required a predefined number of topics for both background corpus and foreground topics. Topic sketch (Xie et al. 2016) was designed to detect bursty topics from Twitter, with the assumption that each tweet is related to only one latent topic. Topics were generated on the basis of sketch-based topic modeling using singular value decomposition of word pair frequency matrices or tensor decomposition of word triple frequency matrices. It also employed a hashing-based dimension reduction technique and conducted an effective sketch maintenance on the basis of the acceleration of word postings.

These methods included only temporal and semantic dimensions in the detection process without considering the geographic dimension. However, with regard to urban disasters, the physical locations and spatial patterns of an event may be as important as text contents and time. Current LDA-based event detection methods have not been tailored to detect disasters—which are different from other events such as celebrations, football games, and marathons—in terms of the intensity of negative emotions. Therefore, we propose a technique to specifically detect urban emergencies. The technique takes into consideration both the geolocations and the text contents of social media microblog postings, and the semantic correlation is measured on the basis of LDA. To separate urban crises from other events, we use the intensity of negative sentiment to filter the events.

Detecting Urban Emergencies Technique

In our study, we defined urban emergencies as events with the following properties: (1) they are geographically proximal and semantically related; and (2) they influence people's emotions and trigger a high intensity of negative sentiment over a certain time period. The properties also formed the basic assumptions of the detection approach. The main modules included in our proposed technique are shown in Fig. 1. The first module for collecting georeferenced tweets from a Twitter streaming API was a built system in our lab (Wang and Taylor 2015). On the basis of this collection module, we first built a data preprocessing module to normalize texts in tweets, and then we developed a geotopic detection module to identify distinct events (geotopics). This geotopic detection module combined geographic closeness and LDA-based semantic similarity to extract geotopics. To further evaluate the emergency level of candidate geotopics, we employed the intensity of negative sentiment to rank the events in a geotopic ranking module, in which events with higher average intensity of negative sentiment over certain consecutive time windows were identified as candidate urban emergencies.

Data Preprocessing Module

Twitter allows its users to share short 140-character messages (this length is in the process of being expanded). The texts can include words, URLs, @mentions, hashtags, emoticons, abbreviations, and so forth. To analyze this data, we cleaned the texts by removing URL links and user mentions (@), which are not relevant to the core meaning of the text in a large number of tweets. We also removed the special characters (#, $, %, ^, &, *, punctuation marks, and independent numeric numbers) that are unnecessary for further analysis. We tokenized the tweets to unigrams on the basis of regular expression patterns. Each tweet was segmented into its constituent words and converted to lowercase. We also removed the stop words that have no significance (e.g., *a*, *the*, and *me*) and words with fewer than two characters.

We further conducted parts-of-speech (POS) tagging and lemmatization. POS tagging classifies and labels POS tags for words in a sentence; the main tags include noun, verb, adjective, and adverb. In this study, the Penn Treebank notation was used for tagging; it is the most widely used POS tag set in various text analytics including tweets. Following the POS tagging, we converted inflectional words to their base forms, called lemma. This standardization process found the base form or lemma for a given word on the basis of the word and the POS by checking the WordNet corpus. It used a recursive technique for removing affixes from the words until a match was found in WordNet; words remained unchanged if no match was found. Thus, the lemmatization was largely influenced by the POS.

Before saving the cleaned texts into a corpus for analysis, we also deleted infrequently used words that appeared in fewer than 0.1% or 0.5% of tweets, depending on text frequencies in different cases (Denny and Spirling 2018), because the infrequently used terms do not contribute much information about text similarity. Discarding these terms can also greatly reduce the size of the vocabulary and speed up the text analysis. A Python package, natural language toolkit (NLTK), was employed to tokenize tweets, remove stop words, tag POS, and lemmatize words in tweets. Notably, these procedures were conducted to generate a dictionary and corpus for LDA topic modeling. The text normalization steps for sentiment analysis had some differences: only URL links and user mentions (@) were removed from the original tweets because we aimed to keep as much information as possible for sentiment analysis. Fig. 2 shows the processes involved in the data preprocessing module.

J. Comput. Civ. Eng., 2019, 33(3): 04019023

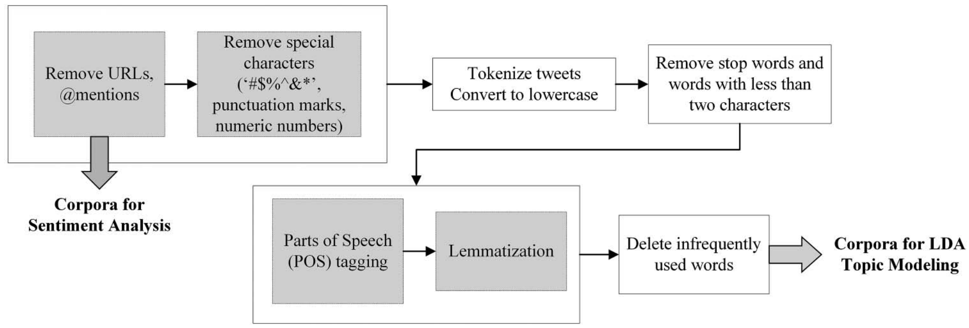

Fig. 2. Data preprocessing.

Candidate Geotopics Generation

LDA-Based Topic Modeling

We employed latent Dirichlet allocation (Blei et al. 2003) to identify latent topic information in the corpora of tweets. A tweet is treated as a bag of words, without regard for the word order and grammar. The basic idea of this generative probabilistic model in our research context was that each tweet is represented as a probability distribution over a predefined number of topics and that each topic is represented as a probability distribution over words. We employed a graphical model (Fig. 4 in Blei 2012, p. 81) to formally describe LDA. A vocabulary was formed on the basis of the collection of tweets, which was indexed by $\{1, \ldots, V\}$. The vth *word* was denoted by a V-vector w ($w^v = 1$ and $w^u = 0$, $u \neq v$). A *tweet* is a sequence of N words denoted by $\mathbf{w} = (w_1, w_2, \ldots, w_N)$. A tweets' corpus is a collection of M *tweet* denoted by $\mathbf{D} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M\}$. We developed the topic model according to Blei and colleagues (Blei 2012; Blei et al. 2003). The basic assumptions of LDA that for *tweet* in a corpus $\mathbf{D}$ include

1. Choose $N \sim$Poisson(ξ).
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each word $w_{d,n}$ in tweet d
   a. Choose a topic $z_d \sim \text{Multinomial}(\theta)$.
   b. Choose a word $w_{d,n}$ from $p(w_n|z_d, \beta)$, a multinomial probability conditioned on the topic z_d.

$\beta_{1:K}$ denotes topics, where β_K is a distribution over the tweets' vocabulary index. θ_d represents the topic proportions for the dth tweet, and $\theta_{d,k}$ is the topic proportion for topic k in tweet d. The topic assignments for the dth tweet are z_d, where z_d is the topic assignment for the nth word in tweet d. Last, the observed words for tweet d are w_d, where $w_{d,n}$ is from the fixed vocabulary and the nth word in tweet d. In the graphical model, only the shaded node ($w_{d,n}$) is observed, and the other nodes (the topic proportions, assignments, and topics) are hidden nodes.

LDA-Based Topic Similarity

Compared with other semantic measurements, the result of the LDA-based cosine similarity is much closer to the human perception of document similarity (Towne et al. 2016). We used the cosine similarity of tweets' topics for our similarity measure

$$C(t \to t') = \cos(\text{topic}_t, \text{topic}_{t'}) \tag{1}$$

where t and t' = two distinct tweets; and topic_t = topic vector of a tweet and the topic vector is a probability distribution over the predefined number of topics.

Geographical Closeness

We adopted the Epanechnikov kernel to measure the geographical closeness of two tweets due to its high efficiency and simplicity (Zhang et al. 2016)

$$K(t \to t') = 0.75 \times (1 - c(\|loc_t - loc_{t'}\|)^2) \tag{2}$$

where $\|loc_t - loc_{t'}\|$ = Haversine distance between locations of two distinct tweets; and c = scaling function that transforms the distance to the range (0, 1).

Geotopic Clustering

We built an undirected graph named Tweet geotopic graph of $G = (V, E)$, where V is a set of tweets and E is a set of edges. The weight of an edge is the product of semantic similarity and geographical closeness. If the weight is less than the predefined threshold, the edge is not built. The node with the highest sum weight is regarded as a hub. Each hub with each of its connected neighbors is regarded as a candidate geotopic.

Geotopics Ranking and Crisis Detection

Kryvasheyeu et al. (2016) revealed that negative average sentiment can indicate an emergency situation on the basis of their examination of sentiment change during hurricane Sandy. Research by the authors has also identified a significant negative correlation between sentiment level and the intensity of an earthquake (Wang and Taylor 2018). We therefore ranked the candidate geotopics on the basis of their intensity of negative sentiment, as follows: Geotopics with higher average intensity of negative sentiment over certain time windows have a higher rank.

We adopted an unsupervised lexicon-based method to measure sentiment. The method was based on the affective word list AFINN to assign sentiment scores to words in tweets (Nielsen 2011). The AFINN word list includes 2,477 words. The valence of a word ranges from −5 (very negative) to +5 (very positive) as an integer. The sum of valence without normalization of words represents the combined sentiment strength for a tweet. A Python package, AFINN, was used to compute the sentiment scores. AFINN is a Twitter-based sentiment lexicon including Internet slang and obscene words. It has been tested in different types of tweets corpora and performs consistently at a satisfactory level of accuracy both for two classes (positive and negative) and for three classes (positive, negative, and neutral), in comparison with other unsupervised methods for sentence-level sentiment analysis (Ribeiro et al. 2015). Moreover, the AFINN word list has shown its advantages in analyzing tweets for disaster and crisis sentiment detection (e.g., Nagy and Stamberger 2012; Walther and Kaisser 2013). We therefore selected the AFINN lexicon to evaluate the sentiment polarity of our collected tweets.

Because we were concerned mainly with negative tweets (in an emergency situation), we converted the numeric scores of sentiment to binary scores. Specifically, we used 0 to represent both positive and neutral sentiment, and 1 for negative sentiment. We employed intensity of negative sentiment under each candidate geotopic to quantify its crisis level

$$S_t = \frac{n_{\text{tweet}^{(-)}}}{n_{\text{tweet}}} \quad (3)$$

where $n_{\text{tweet}^{(-)}}$ = number of tweets with negative sentiment in each cluster; and n_{tweet} = total number of tweets in each cluster. To include the impact duration of each crisis, we calculated the average values of the intensity over consecutive time windows

$$\bar{S} = \left(\sum_{1}^{i} S_t\right) \Big/ i \quad (4)$$

where i = number of consecutive time windows, and $i \geq 1$.

Validation

To assess and validate the functionality of the proposed detection technique, we implemented DUET to detect different types of emergencies in multiple scenarios using georeferenced tweets. These emergencies had happened in the past, and we had already collected this data using a system running in our lab (Wang and Taylor 2015). The time windows of data used to assess and validate DUET included the Interstate 85 collapse in the city of Atlanta; the Grenfell Tower fire in London; the magnitude 5.8 earthquake in Oklahoma City, Oklahoma; and the Tubbs Wildfire in northern California (Table 1). Although we knew that these emergencies had occurred, no keywords or other means were used to instruct DUET to search for these specific emergencies. To assess and validate DUET, we hypothesized that these specific crises would emerge from the data. The georeferenced tweets were further filtered in the spatial bounding box of selected urban areas, which were characterized by different geographical scales. We specifically selected a short and hourly-based period as the detection period during which the targeted detection events happened. Each studied period was sliced into several equal-length time windows in order to analyze the average intensity of negative sentiment for geotopics. Details about the cases are provided in Table 1, which also includes the data volume of English tweets and the size of the cleaned vocabulary corpus. Notably, words with frequency lower than 0.1% or 0.5% of the total number of tweets were removed from the vocabulary corpus [percentage determined as per M. J. Denny and A. Spirling (unpublished data, 2017)], and in each different case, distinct common words were deleted from the corpus. Our experiments were conducted with Python 2.7 in the Anaconda environment. The new version of Python (version 3.3) has required some GIS-based scripts from earlier versions of Python to be changed. However, we did not use Python scripts in ArcGIS for filtering or analysis; therefore, the Python version change will not impact DUET's results.

Candidate Event Generation

To generate geotopics based on both semantic similarity and geographical closeness, we set the threshold of semantic similarity to 0.7 and set geographical closeness to 0 to generate neighbors of each tweet. The selection of thresholds was based on findings of a former study (Zhang et al. 2016) coupled with our validation goal—to identify as many emergency-related topics as possible in the selected detection geographical scale for each event. The number of LDA topics was set differently because the diversity level of Twitter topics varies by geographic area. Specifically, the number of LDA topics is place-specific, and as such, it may be used consistently in a certain area, but there is a difference between distinct regions. We determined the number of topics by comparing the results from topic modeling with different numbers of topics (from 5 to 15) on the basis of the highest topic coherence (Röder et al. 2015), which measures the semantic coherence of topics coupled with human judgment. Finally, we obtained a different number of unique geotopics for each data set (Table 2). We also calculated the geotopics' average number of tweets in each case (Table 2). It is noteworthy, however, that one tweet might belong to no geotopic or to more than one geotopic.

Geocrisis Ranking

We further ranked all candidate geotopics on the basis of the average intensity of negative sentiment over consecutive time windows for each case. Table 3 provides examples of the hub tweets related to the emergencies (tweet with the highest sum of weights) and the

Table 1. Geotagged tweets corpora specification for validation test cases

Targeted cases	Detected area (bounding box)	Period	Number of tweets (English)	Number of vocabulary words (cleaned)
Interstate 85 collapse	Atlanta	23:00 (UTC) on March 30 to 4:00 (UTC) on March 31, 2017 (6 h)	596	326
Grenfell Tower fire	Greater London	5:57 to 9:56 (UTC) on June 14, 2017 (4 h)	1,309	1,274
Oklahoma earthquake	Oklahoma City, Oklahoma	12:00 to 17:57 (UTC) on September 3, 2017 (6 h)	539	244
Tubbs fire (wildfires)	Northern California megaregion	12:00 to 16:00 (UTC) on October 9, 2017 (4 h)	1,035	241

Table 2. Parameters setting and generation of geotopics

Targeted emergency events	Number of LDA topics	Topic coherence	Threshold of semantic similarity	Threshold of geographical closeness	Number of unique geotopics	Average number of tweets under each geotopic
Interstate 85 collapse	10	−15.96	0.7	0	36	89
Grenfell Tower fire	15	−15.91	0.7	0	68	113
Oklahoma earthquake	7	−11.66	0.7	0	20	136
Northern California megaregion wildfire	15	−12.76	0.7	0	72	82

Table 3. Examples of hub tweets and intensity of negative sentiment for top-ranked emergency-related geotopics

Targeted emergency events	Hub tweet for emergency-related geotopics	Average intensity of sentiment
Interstate 85 collapse	(#1) I85 fire. Here's what was burning. But why? #i85 #i85collapse @mention <url>	0.368
Grenfell Tower fire	(#2) #grenfell towerfire as seen from white city #iglondon #ig_london @mention	0.248
Oklahoma earthquake	(#1) A M5.6 magnitude earthquake occurred in Oklahoma Details: <url> Map: <url>	0.172
Northern California megaregion wildfire	(#2) Huge wild fire in the north bay this morning!!! Saw it in the climb gnarly winds in the night	0.352

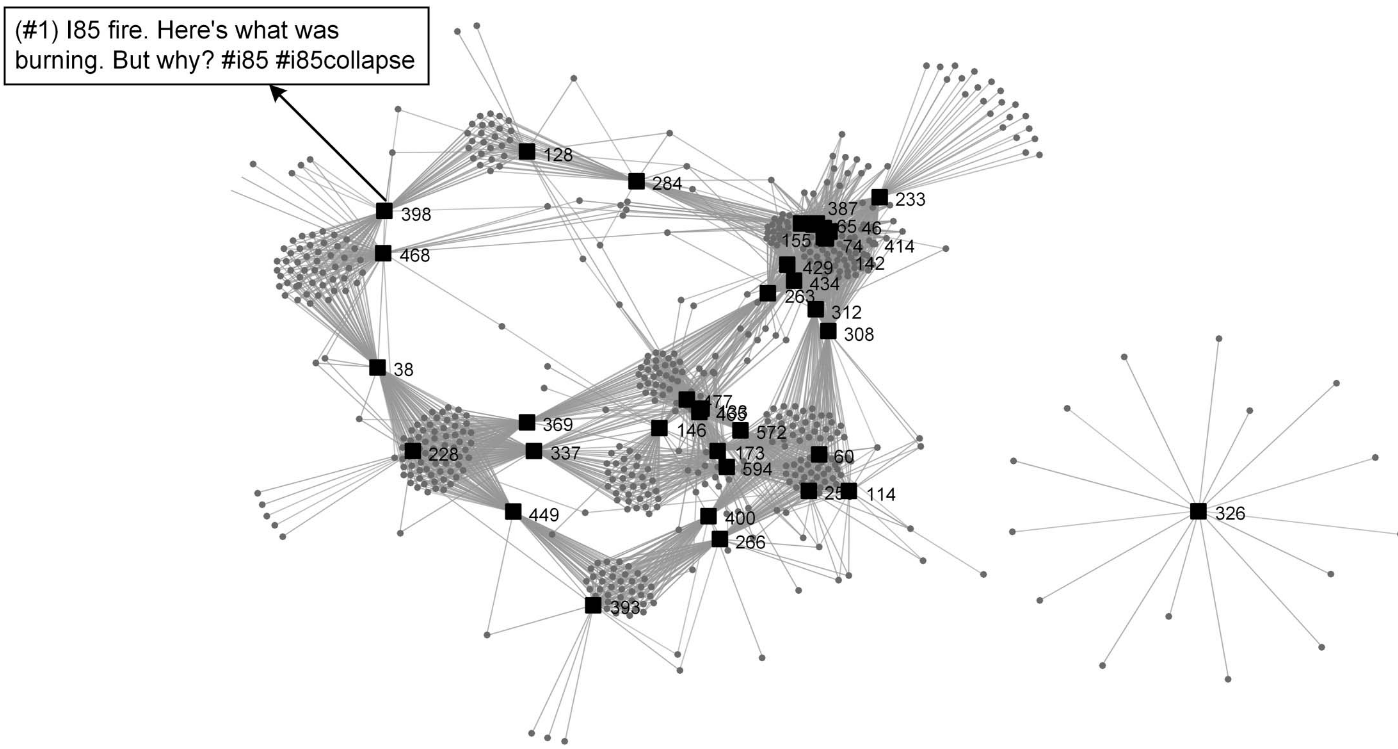

Fig. 3. Network visualization of geotopics in Atlanta.

average intensity of negative sentiment over time windows for top-ranked geotopics.

During the 6 h encompassing the Interstate 85 bridge collapse in Atlanta, four of the top five geotopics characterized by the highest intensity of negative sentiment were related to this event and traffic situation. Similarly, for geotopics in London, four of the top 10 geotopics mentioned the Grenfell Tower or the fire. In northern California, our detection technique identified several fire spots in the affected area of the Tubbs wildfire. Finally, in the Oklahoma City area magnitude 5.8 earthquake, the emergency was also detected by DUET; the top five geotopics were all related to the earthquake. The network structure of geotopics in Atlanta is shown in Fig. 3. The solid squares refer to the hub tweet, and the gray circles represent individual tweets. One tweet may belong to multiple geotopics. The number close to the hub is the number of its corresponding tweet. A hub tweet and all its corresponding tweets construct a geotopic. The geotopics are intercorrelated rather than independent. These geotopics were ranked on the basis of their intensity of negative sentiment. By retrieving the content of the hub tweet and the associated tweets of top-ranked geotopics, DUET could identify the emergencies. For example, Tweet 398 is the hub tweet of the top-ranked geotopic over the time period, which is also given in Table 3.

Discussion

DUET, our proposed detecting urban emergencies technique, built upon a Twitter streaming API, comprises three modules: a data-preprocessing module, a geotopic detection module, and a sentiment-based geotopic ranking module. We implemented the detection technique in distinct scenarios (e.g., different types of crises, geographical areas, spatial scales, and data volumes) and validated its capability in identifying general types of crises effectively. Compared with former targeted-event detection techniques (Gu et al. 2016; Sakaki et al. 2010; Sun et al. 2016), DUET does not require predefined keywords regarding the emergency's type and characteristics; it detects emergencies on the basis of rankings of the intensity of negative sentiment over specific time windows. This can improve the situation awareness of emergency managers by providing a range of candidate crises to consider. Additionally, DUET is more targeted on detecting focal emergencies/crises than are current general-event detection systems (Abdelhaq et al. 2013; Schubert et al. 2014; Yu et al. 2017; Zhang et al. 2016). Specifically, we extended general-event detection from large corpora of text into the field of disaster and emergency response. Our technique can identify geotopics that relate to general events, as well as detect candidate emergencies among all emergent geotopics on the basis of the degree of negative sentiment.

The identified emergent geotopics can be reported to both the public and first responder organizations to increase awareness of when and where an emergency is occurring, to share voluntarily provided details of an emergent crisis, and to provide an assessment of the sentiment level associated with the emergency. Extending functionality beyond current disaster information systems, DUET bridges the gap by detecting emergencies from social media postings to supplement current emergency reporting during disasters. In doing so, it also addresses the paucity of research focused on the identification phase in disaster management systems in terms of agility, openness, and effectiveness (Ghosh et al. 2013; Peña-Mora et al. 2010, 2012; Pradhan et al. 2007). By basing its emergency detection on crowdsourced data, DUET extends beyond and integrates across previous empirical studies on volunteered geolocation contents and sentiment from social media (Kryvasheyeu et al. 2016; Wang and Taylor 2018), extending them toward practical implementation.

To enable the application of DUET in more diverse scenarios with more agility, we still need to make some improvements in future studies. First, we assumed that emergencies are geographically proximal and semantically related and that emergencies influence people's emotion and trigger a high intensity of negative sentiment over a certain time period. For this reason, it is difficult to detect crises that do not trigger a high-intensity level of negative sentiment. To enable effective detection when a high intensity of sentiment may not exist, we plan to adjust the ranking module to rank the geotopics on the basis of the change in negative sentiment intensity and to adjust DUET to consider other dimensions, such as the geographic spreading speed.

Second, we focused exclusively on tweets in English, and this may lead to bias in identifying emergency events incorporated in tweets in other languages. However, the English tweets that we examined represent nearly 90% of the collected geotagged tweets from our streaming API. Therefore, they are suitable for our analyses. Nevertheless, with the development of topic modeling and sentiment analysis in other languages (e.g., Chinese, French, and Spanish), we hope to enhance DUET in order to detect emergencies in multilingual corpora.

Additionally, our proposed detection technique has detected geotopics related to the Interstate 85 bridge collapse, Grenfell Tower fire, Oklahoma earthquake, and Tubbs wildfires, and these geotopics are ranked among the highest intensity of negative-sentiment clusters. We evaluated the geotopics on the basis of the content of the hub tweets as well as all the connected tweets that mostly concern the emergent events. Although in the earthquake case, one hub tweet did not directly show its relevance to the earthquake, other tweets under the same geotopic were related to the earthquake. This can be caused by the centrality of its geographical location and the generality of the tweet's content. For this reason, to understand the emergent crisis, it may be necessary to check other tweets' content beyond the hub tweet in evaluating an identified emergency.

We generally set the thresholds for semantic similarity and geographical closeness to 0.7 and 0 in the four cases, but different selections of the thresholds can influence the number and scale of geotopics. We hope to analyze the sensitivity of geotopics to the thresholds in terms of their geographical scale and average intensity of negative sentiment in different urban areas with more available data sets in the future. Additionally, because the number of LDA topics is place-specific, we plan to examine the number of topics over different time periods (e.g., minutes, hours, and days) for distinct cities in order to select optimal settings in various detection scenarios. In the next stages of this research, we will meet with disaster response and emergency management personnel to work with real data from prior emergencies to determine the optimal settings for specific cities. We will further explore with these practitioners how to best integrate the evolving signals for help from social media into disaster response approaches and systems, with the ultimate goal of providing them with a technique to increase the speed of emergency detection, improve their ability to assess the severity of crises when there are multiple concurrent crises occurring across a city, and ultimately to save more lives and reduce injury in the critical first minutes and hours of an emergency.

Conclusion

In this paper, we developed and validated a novel technique for detecting urban emergencies using data collected from a geosocial networking platform (Twitter). The technique is capable of detecting different types of emergencies (e.g., infrastructure failure, building fire, city earthquake, and wildfire near a city) in distinct urban environments over short time periods. This new technique for detecting urban emergencies leverages both semantic and geographical similarity in generating candidate events and evaluates the crisis level on the basis of the intensity of negative sentiment. The next step of this study is to expand the technique for scalable geocrisis detection in high-volume real-time tweets. We will also incorporate an online module to identify emergencies in real time and add modules for disaster tracking, visualization, and assessment. We will build an open platform to inform emergency management personnel regarding the type, content, location, and sentiment level of an emergency.

The proposed urban emergencies detection technique has the potential to provide first responders and emergency management agencies in affected areas with an updated understanding of the role georeferenced social media can play in increasing the effectiveness of disaster response efforts. DUET contributes to improving disaster and emergency management, enhancing situation awareness, and taking steps toward achieving urban resilience. The completed detection platform can also be incorporated into current disaster information systems in the future in order to build an open, integrated, connected, and agile disaster management system.

Acknowledgments

This study is supported by the National Science Foundation under Grant No. 1760645 and a Virginia Tech BioBuild IGEP grant. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or Virginia Tech.

References

Abdelhaq, H., C. Sengstock, and M. Gertz. 2013. "Eventweet: Online localized event detection from Twitter." *Proc. VLDB Endowment* 6 (12): 1326–1329. https://doi.org/10.14778/2536274.2536307.

Blei, D. M. 2012. "Probabilistic topic models." *Commun. ACM* 55 (4): 77–84. https://doi.org/10.1145/2133806.2133826.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet allocation." *J. Mach. Learn. Res.* 3 (Jan): 993–1022.

Denny, M. J., and A. Spirling. 2018. "Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it." *Political Anal.* 26 (2): 168–189. https://doi.org/10.1017/pan.2017.44.

Ford, J. D., et al. 2016. "Opinion: Big data has big potential for applications to climate change adaptation." *PNAS* 113 (39): 10729–10732. https://doi.org/10.1073/pnas.1614023113.

Ghosh, J. K., D. Bhattacharya, P. Boccardo, and N. K. Samadhiya. 2013. "Automated geo-spatial hazard warning system GEOWARNS: Italian case study." *J. Comput. Civ. Eng.* 29 (5): 04014065. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000372.

Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabasi. 2008. "Understanding individual human mobility patterns." *Nature* 453 (7196): 779. https://doi.org/10.1038/nature06958.

Gu, Y., Z. S. Qian, and F. Chen. 2016. "From Twitter to detector: Real-time traffic incident detection using social media data." *Transp. Res. Part C: Emerging Technol.* 67: 321–342. https://doi.org/10.1016/j.trc.2016.02.011.

Kryvasheyeu, Y., H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian. 2016. "Rapid assessment of disaster damage using social media activity." *Sci. Adv.* 2 (3): e1500779. https://doi.org/10.1126/sciadv.1500779.

Li, T., et al. 2017. "Data-driven techniques in disaster information management." *ACM Comput. Surv.* 50 (1): 1–45. https://doi.org/10.1145/3017678.

Maurya, A., K. Murray, Y. Liu, C. Dyer, W. W. Cohen, and D. B. Neill. 2016. "Semantic scan: Detecting subtle, spatially localized events in text streams." Preprint, submitted February 13, 2016. http://arxiv.org/abs/1602.04393.

Nagy, A., and J. Stamberger. 2012. "Crowd sentiment detection during disasters and crises." In *Proc., 9th Int. ISCRAM Conf.*, 1–9. Brussels, Belgium: ISCRAM.

National Academies of Sciences, Engineering, and Medicine. 2017. *Emergency alert and warning systems: Current knowledge and future research directions*. Washington, DC: National Academies of Sciences, Engineering, and Medicine.

Nielsen, F. Å. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." Preprint, submitted March 15, 2011. http://arxiv.org/abs/1103.2903.

Peña-Mora, F., A. Y. Chen, Z. Aziz, L. Soibelman, L. Y. Liu, K. El-Rayes, C. A. Arboleda, T. S. Lantz Jr., A. P. Plans, and S. Lakhera. 2010. "Mobile ad hoc network-enabled collaboration framework supporting civil engineering emergency response operations." *J. Comput. Civ. Eng.* 24 (3): 302–312. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000033.

Peña-Mora, F., J. K. Thomas, M. Golparvar-Fard, and Z. Aziz. 2012. "Supporting civil engineers during disaster response and recovery using a Segway mobile workstation Chariot." *J. Comput. Civ. Eng.* 26 (3): 448–455. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000117.

Pradhan, A. R., D. F. Laefer, and W. J. Rasdorf. 2007. "Infrastructure management information system framework requirements for disasters." *J. Comput. Civ. Eng.* 21 (2): 90–101. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:2(90).

Ribeiro, F. N., M. Araújo, P. Gonçalves, F. Benevenuto, and M. A. Gonçalves. 2015. "SentiBench —A benchmark comparison of state-of-the-practice sentiment analysis methods." Preprint, submitted December 6, 2015. http://arxiv.org/abs/1512.01818.

Röder, M., A. Both, and A. Hinneburg. 2015. "Exploring the space of topic coherence measures." In *Proc., Eighth ACM Int. Conf. on Web Search and Data Mining*, 399–408. New York: ACM.

Sakaki, T., M. Okazaki, and Y. Matsuo. 2010. "Earthquake shakes Twitter users: Real-time event detection by social sensors." In *Proc., 19th Int. Conf. on World Wide Web*, 851–860. New York: ACM.

Schubert, E., M. Weiler, and H.-P. Kriegel. 2014. "Signitrend: Scalable detection of emerging topics in textual streams by hashed significance thresholds." In *Proc., 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 871–880. New York: ACM.

Seetharaman, D., and G. Wells. 2017. "Hurricane Harvey victims turn to social media for assistance." *Wall Street Journal*, August 29, 2017.

Sun, H., Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao. 2016. "Data-driven power outage detection by social sensors." *IEEE Trans. Smart Grid* 7 (5): 2516–2524. https://doi.org/10.1109/TSG.2016.2546181.

Towne, W. B., C. P. Rosé, and J. D. Herbsleb. 2016. "Measuring similarity similarly: LDA and human perception." *ACM TIST* 8 (1): 7:1–7:28. https://doi.org/10.1145/2890510.

Walther, M., and M. Kaisser. 2013. "Geo-spatial event detection in the Twitter stream." In *Proc., European Conf. on Information Retrieval*, 356–367. New York: Springer.

Wang, Q., and J. E. Taylor. 2015. "Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms." *J. Comput. Civ. Eng.* 30 (2): 04015004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000469.

Wang, Y., and J. E. Taylor. 2017. "Tracking urban resilience to disasters: A mobility network-based approach." In *Proc., 14th Int. Conf. on Information Systems for Crisis Response and Management*, edited by T. Comes, F. Bénaben, C. Hanachi, M. Lauras, and A. Montarnal, 97–109. Brussels, Belgium: ISCRAM.

Wang, Y., and J. E. Taylor. 2018. "Coupling sentiment and human mobility in natural disasters: A Twitter-based study of the 2014 South Napa earthquake." *Nat. Hazard.* 92 (2): 907–925. https://doi.org/10.1007/s11069-018-3231-1.

Xie, W., F. Zhu, J. Jiang, E.-P. Lim, and K. Wang. 2016. "Topicsketch: Real-time bursty topic detection from twitter." *IEEE Trans. Knowl. Data Eng.* 28 (8): 2216–2229. https://doi.org/10.1109/TKDE.2016.2556661.

Yu, W., J. Li, M. Z. A. Bhuiyan, R. Zhang, and J. Huai. 2017. *Ring: Real-time emerging anomaly monitoring system over text streams*. New York: IEEE.

Zhang, C., G. Zhou, Q. Yuan, H. Zhuang, Y. Zheng, L. Kaplan, S. Wang, and J. Han. 2016. "Geoburst: Real-time local event detection in geo-tagged tweet streams." In *Proc., 39th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, 513–522. New York: ACM.