IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.Doi Number

A Hybrid Machine Learning Pipeline for Automated Mapping of Events and Locations from Social Media in Disasters

Chao Fan[1], Fangsheng Wu[2], and Ali Mostafavi[1], Member, IEEE
[1]Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840 USA
[2]Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77840 USA

Corresponding author: Chao Fan (e-mail: chfan@tamu.edu).

This material is based in part upon work supported by the National Science Foundation under Grant Number CMMI-1846069 (CAREER) and the Amazon Web Services (AWS) Machine Learning Award. The authors also would like to acknowledge the funding support from the National Academies' Gulf Research Program Early-Career Research Fellowship. 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 and Amazon Web Services.

ABSTRACT The objective of this study is to propose and test a hybrid machine learning pipeline to uncover the unfolding of disaster events corresponding to different locations from social media posts during disasters. Effective disaster response and recovery require a comprehensive understanding of disaster situations, i.e., unfolding of disaster events and geographic distribution of the disruptions. Existing studies have employed machine learning methods to conduct coarse-grained event detection and analyze the geographical location information from geotagged social media data. However, only a very small fraction of the entire set of social media data includes geotagged information, which may not directly correspond to events described in the content of posts. In addition, the coarse-grained information detected by existing approaches is token-based, which does not provide sufficient information for situation awareness. Hence, the detection of location and finer-grained event information could significantly improve the utility, credibility, and interpretability of social media data for situation awareness. To address these limitations, this study proposed a hybrid machine learning pipeline that makes use of all relevant tweets to uncover the evolution of disaster events across different locations. The pipeline integrates Named Entity Recognition for detecting locations mentioned in the posts, location fusion approach to extract coordinates of the locations and remove noise information, fine-tuned BERT model for classifying posts with humanitarian categories, and graph-based clustering to identify credible situational information. The application of the study is demonstrated using the data set collected from Twitter during the 2017 Hurricane Harvey in Houston. The results show the capability of the proposed hybrid pipeline for automated mapping of events across time and space from social media posts with considerable accuracy. The findings also suggest that the potential for forensic analysis of disasters using mapped events and their evolution, and based on the variation of social media attention to different locations in disasters. Hence, this method could provide a useful tool to support emergency managers, public officials, residents, first responders, and other stakeholders in rapid situation awareness across time and space.

INDEX TERMS Machine learning pipeline, social media, disasters, automated mapping.

I. INTRODUCTION

Natural disasters such as hurricanes, wildfire, and earthquakes cause large-scale disruptions over affected areas [1]. One key aspect of effective disaster response is situation awareness based on a timely and accurate assessment of the unfolding of events across time and space [2]. To this end, developing effective information processing tools to identify the timing and locations of events in disasters is critical. Over the past decade, social media has been increasingly used by people and responders to enhance information sharing and situational awareness [3], [4]. Social media platforms enable public users to report events and share personal experiences and reactions to disasters. Hence, the data generated on social media provide a unique opportunity to capture disaster situations with a relatively high temporal and spatial resolution to map different events across various locations [5], [6].

IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

The use of social media data for disaster management has led to the creation of a variety of analytics techniques for information processing [7]. There are two main streams of research in the existing literature that aim to leverage social media data for enhanced situation awareness during disaster response and recovery: (1) information extraction, and (2) geographical estimation. The first stream of studies focuses on developing analytic approaches to advance information extraction from social media in disasters. Specifically, text classification and sentiment analysis are the main tasks related to information extraction. Multiple machine-learning classifiers such as Naïve Bayesian classifier [8], Convolutional Neural Networks [9], and Generative Adversarial Network [10] have been deployed to classify the social media data with disaster-related topics. For example, Singh et al. proposed an algorithm to categorize the tweets into a high or low priority for evaluating the relevancy of the information in the tweets [11]. Mouzannar et al. proposed a multimodal deep learning framework to identify damage-related information [12]. The development of these techniques enables detecting disaster events using posts shared on social media. Another stream of studies has focused on estimating the geographical information for disaster-related tweets in order to assess disaster damages over different locations. A generally used approach for estimating and mapping locations of tweets is to learn the locations of users from their historical posts. For example, Ikawa et al. proposed a method to learn the association between a location and its relevant keywords from past messages, and estimate the location of new posts in disasters [13]. Morstatter et al. examined multiple machine learning techniques to infer the location of social media posts [14]. By integrating the geographical information from geotagged tweets and the estimated locations, researchers have come up with some disaster mapping and tracking systems to retrieve and visualize social media data for disaster situation awareness [15], [16].

Despite the advances in the techniques for information extraction and geographical estimation, reliable automated mapping of events across time and space during disasters remains a challenge. First, there are two types of information that are extracted from social media. One type of information is obtained using topic modeling, which provides a list of keywords to represent disaster events. While it is an efficient way to reduce the redundancy of the information from social media, a list of keywords is hard to be interpreted as a description of a disaster event. The other type of information is labeled tweets using topic classifiers. However, the size of the labeled tweets is often too large to be processed by humans and the credibility of the information needs improvement. Hence, existing techniques for information extraction have limitations in terms of interpretability and credibility of the situational information extracted from social media. Second, although existing techniques can estimate where a post came from based on geotags and users' historical information, the location information obtained based on users' location may not be consistent with the reported situation (sender-situation mismatch). Also, using the existing techniques, the resolution of an estimated location tends to be at the city or neighborhood level, and gaining more detailed information such as a street or a store is still challenging. Thus, better techniques are needed to improve the accuracy and resolution of location information obtained from social media data.

To address this gap, this study proposed a hybrid machine learning pipeline to extract credible and interpretable situational information for understanding the evolution of disaster events and locations from social media posts. The proposed pipeline enables (1) mapping the evolution of the disaster situations for different humanitarian event categories across time and locations, and (2) understanding the geographic distribution of social media attention to different locations in disasters. A case study of the 2017 Hurricane Harvey in Houston was conducted to examine the capabilities of the proposed pipeline in automated mapping of events and locations.

II. RELATED WORK

The existing literature related to the analysis of social media data for crisis response and disaster management is rapidly growing. This paper, however, focuses more narrowly on the state-of-the-art machine learning techniques that have been employed to detect and map disaster events and their geographical information. Hence, this section discusses related work in the extant literature focusing on this aspect.

The role of social media data for information propagation in disasters has been recognized in devastating natural disasters such as the 2010 Haiti Earthquake and the 2012 Hurricane Sandy [17]. The majority of studies focus on processing disaster-related social media data to gain situational awareness. One branch of existing studies focus on developing topic modeling and event detection approaches to extract frequently mentioned keywords corresponding to disaster events [18]. For example, Mishler et al. modeled the temporal changes in topics in the Ukrainian Crisis using Structural Topic Modeling [19]. Kireyev et al. conducted topic modeling on tweets during two earthquakes in American Samoa and Sumatra, emphasizing dynamic corpus refinement [20]. In addition, researchers also designed a number of platforms to perform automatic classification of crisis-related microblog communications [21]. AIDR (Artificial Intelligence for Disaster Response) is a platform that aims to classify social media posts into humanitarian categories in real-time [22]. The majority of the existing machine learning pipelines focus on grabbing and classifying real-time posts to improve the efficiency of situational awareness in disasters. However, these approaches have an important limitation related to providing detailed information about specific disaster events (such as which road is closed, when it will open, which building is damaged, and how long service was interrupted). In addition, existing classifiers and event detection pipelines do not provide the capability to map events to specific locations based on the content of posts.

Second, the majority of existing studies related to analyzing the geographical information on social media primarily employ the geotags from the posts to specify location insights. One branch of studies has focused on analyzing the geotagging behaviors of people in disasters [23], [24]. For example, Kumar et al. proposed an approach to identify whether a tweet is generated from crisis regions based on the historical geotags [25]. Another branch of the studies has examined the utility of geotagged social media posts for disaster response and recovery [26]. For example, Kryvasheyeu et al. examined the relationships among geotagged tweets, human sentiments, disaster damages, and the hurricane's path [1]. Wang et al. developed an approach for detecting disaster events solely based on geotagged tweets [27]. Hodas et al. extracted and classified highly relevant tokens used in geotagged tweets during emergencies [28]. Using the relevant tokens, people can craft more relevant geotagged tweets to capture the situation in disaster-affected areas. To further gain real-time and intuitive disaster situations, a number of studies developed tools to track and map the geotagged social media posts [29]. Tweet-Tracker is one of the popular tools for monitoring and analyzing location and keyword specific Tweets with near real-time trending [30]. TWRsms is another tool which supports data collection and early warning using geotagged tweets [31]. Despite the progress in exploring the capabilities of geotagged tweets for disaster management, there are important limitations related to drawing insights from geotagged tweets. In particular, the geotagged tweets only account for less than 1% of all types of tweets [32]. Situational information published by government officials, organizations, news media, and reporters tend not to include geotags. Individuals may not attach the geotags to their tweets, either, due to privacy concerns. Also, the sender of a post might report an event at a location different from her/his location, and thus, using geotags could create a mismatch between the location of the user and location of the event. In addition, Twitter recently removed the support for precise geotagging on mobile devices.

In summary, existing studies and techniques for analyzing disaster situations mainly focus on topic classification and Twitter geotags. These approaches have important shortcomings for exploring the evolution of the events for a specific location due to the limitations related to extracting fine-grained situational information and identifying specific locations. In order to address this gap, this paper presents a hybrid machine learning pipeline that enables mapping events across various locations based on the content of social media posts.

III. THE HYBRID MACHINE LEARNING PIPELINE

We propose a hybrid machine learning pipeline to detect the evolution and geographical distribution of disaster events using social media data (see Figure 1). The pipeline is composed of three modules: input, learning, and output.

First, the input module is designed to prepare the data for learning. There are multiple types of posts on social media. Take an example of Twitter. Users can post original tweets to share their experiences and observations in disasters. They can also retweet the posts from other users to transmit the situational information to their connections. In addition, users can quote and reply to the tweets that pique their interests. Except for the original tweets, all other types of tweets have repeated content which is retransmitted from the original tweets. To avoid processing repeated content, the first module filters out retweets, replies, and quotes, and only keeps original tweets to be analyzed in the following modules. In addition, since social media posts are made by users without any structure in the content, it is important to remove the noise from the data so that the results can be more reliable. For example, users might post emoji and multiple exclamation marks to indicate their emotions regarding a disaster or its impacts. In addition, users might add hyperlinks to share situational information from external sources such as news articles. Such heterogeneous posting behaviors lead to a challenge of removing noise for computational analysis. Hence, the first module of the pipeline employs regular expression to remove noise (such as emojis, punctuations, and hyperlinks) in the posts.

The second module, composed of two components, is designed to extract location entities and situational information from the posts. The first component is developed to identify and fuse location entities in the posts using Named Entity Recognition and Google Map Geocoding API. The second component adopts a classifier, fine-tuned BERT (i.e., Bidirectional Encoder Representations from Transformers) to classify the posts with different humanitarian categories, such as infrastructure and utility damages, rescue and donation, and injured victims. The classification of the posts can, to some extent, filter out some off-topic posts that may impair the precision of unfolding events and their geographic distribution. Finally, the posts with the same topic for a location are assembled for clustering, which enables identifying credible situational information. The following sub-sections explain the specific steps, equations, and algorithms in the learning module.

The third module is the output module, which enables interpreting and visualizing the results from the learning module to uncover the evolution and geographical distribution of disaster events. There are two components in this module, which are designed to represent both temporal and spatial information for disaster events. First, examining the unfolding of disaster events at a particular location requires temporal information as well as the event contexts. However, the temporal information for the occurrence of an event is sparse in social media posts. That is because users who post about the event may even not know the exact time when the event happened. To address this limitation, we consider the timing

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

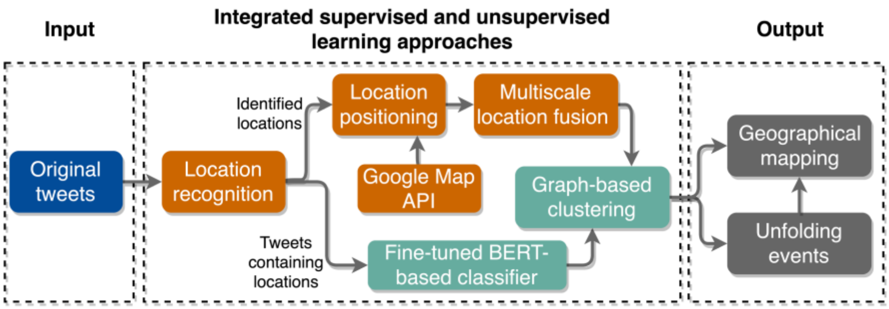

FIGURE 1. The proposed hybrid machine learning pipeline.

of the earliest related posts as an estimation of the timing of the event. By doing so, the unfolding of events can be mapped in a timeline for a particular location. Second, understanding the social media attention to a disaster-affected location requires detecting geographical information of the location in posts. By using the Google Maps Geocoding API, the coordinates of the recognized location entities can be saved. Accordingly, we can make use of the latitudes and longitudes for the recognized locations to locate a post on a geographic map to estimate the density of posts across various locations for different categories of events.

A. LOCATION RECOGNITION AND FUSION

Social media users tend not to share their locations due to privacy concerns. Therefore, a small fraction (less than 1%) of social media posts usually have geotags information. This leads to a challenge related to obtaining the location of events included in posts. However, a number of studies [33] have discovered that locations are mentioned in the content of posts, which provides an opportunity to complement the limited geotagged posts. Hence, we adopted the Stanford Named Entity Recognition (NER) tool [34] to extract the locations that are mentioned in the tweets. The Stanford NER tool is a Java implementation of Named Entity Recognizer. Based on the linear chain Conditional Random Field model, it labels sequences of words in the text into an entity type (i.e., “location” in this study).

By implementing the tool, we can obtain the locations mentioned in each tweet. However, due to the limitations of the tools, noise still persists. For example, some recognized locations are not in disaster-affected areas, and some locations are incorrectly identified as other entities and characters. These problems would negatively influence the accuracy in the analysis due to the loss of information or the insertion of noise. To reduce the noise, we deploy Google Map Geocoding API to match the identified locations with existing place names in disaster-affected areas. The locations that cannot be matched or matched with a place outside of the affected area are excluded from further analysis. Subsequently, the approach extracts the coordinates for the matched locations to enable the aggregation of frequent posts about a particular event in a specific location.

In addition to the above-mentioned noise, there are two additional issues in refining the list of locations. First, some location entities are exactly the same or close to each other, but are referred to different names in the posts, such as “hwy 10” and “Highway 10”. In general, the situation in nearby locations remains the same. For this type of location entities, the approach should identify them as the same location and combine the sets of posts mentioning the names of these entities. One feasible approach to solve this problem is to define a radius so that the location entities within the radius of the circle can be merged to represent a single location (see Figure 2). In the proposed pipeline, the radius can be customized based on the resolution needed for location mapping. However, the radius should also be defined within a reasonable range. If the radius is too small such as 0.2 miles, two similar location entities cannot be merged. This would lead to the redundancy of situational information. On the contrary, the location entities that are far away from each other would be merged if the radius is too large such as 10 miles. This would provide confusing information related to various events and location entities in a single representative location, and cause a low accuracy for mapping disaster events. To make a trade-off between the information redundancy and mapping accuracy, we set the radius as 1 mile based on several trials in the Hurricane Harvey case study.

Second, some locations represent a neighborhood area instead of a place with point coordinates. For this type of location entities, we need to develop and apply fusion rules to identify the geographical hierarchy of these location entities (e.g., neighborhoods, counties, and states). The proposed approach identifies the types of areas based on the length of the diagonal of the area (see Figure 2). By examining the sizes of different types of locations in Houston for the Hurricane Harvey case study, we find that the size of a neighborhood is significantly smaller than the size of a county, and in general, the diagonal length of a county is greater than 40 miles. Thus, we select 40 miles as the threshold to distinguish the neighborhood and county entities. In summary, we defined two rules for fusion and hierarchization of the location entities:

- *Rule 1:* two points (locations or the center of a place) are fused if their physical distance is less than 1 mile.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2965550, IEEE Access

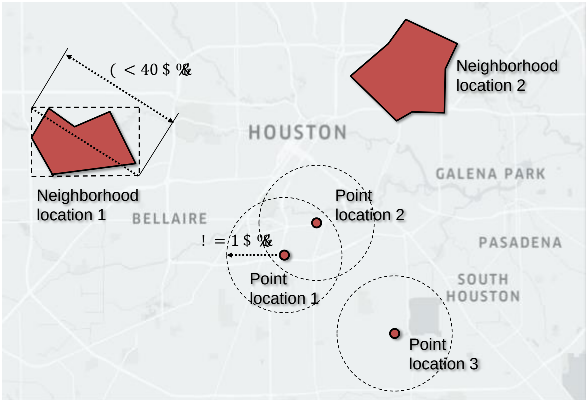

FIGURE 2. An illustration of location fusion rules (The distance on the graph is not proportional to the actual distance).

- *Rule 2:* a place (location) is categorized to a county if the distance between two diagonal points is larger than 40 miles. Otherwise, it is categorized into a neighborhood.

In these rules, a point represents a location of which Geocoding API returns a pair of precise longitude and latitude. A place represents an area of which Geocoding API returns the coordinates of its four corner points. By doing so, we can categorize the locations into three levels of scales: point, neighborhood, and county. Then, the posts mentioning the fused locations are merged into a single set.

B. FINE-TUNING BERT CLASSIFIER

Social media data contain a variety of situational information such as infrastructure damages and utility outages that could inform disaster relief and response [35]. To process the situational information posted on social media efficiently, it is essential to classify the posts in terms of events related to different humanitarian categories. After classification, the noise and non-informative posts can be removed, and the dataset would be more informative for assessing the situation in terms of specific humanitarian categories. A number of existing studies [36] have developed multiple classifiers for the classification tasks by employing high-performance deep learning techniques, such as Convolutional Neural Network (CNN) [9], Gated Recurrent Unit (GRU) [37], and Long Short-term Memory (LSTM) [38]. Despite the advances and applications of these deep learning approaches, a key element impairing the performance of these approaches is the input embedding. Existing studies tend to adopt commonly-used pre-trained embedding packages to encode the textual information from the posts. However, most of the pre-trained packages are unidirectional and their parameters are fixed so that they may not adapt to other domains without the tuning parameters.

A state-of-the-art model, BERT, is proposed recently [39] to enhance language representation. The BERT model is designed to train deep bidirectional representations by jointly conditioning on both left and right contexts in all layers [39]. In addition, the tokens, segments, and positions of the tokens are all embedded in the representation. Due to the benefits of the BERT model, we adopted a pre-trained BERT model with 12 transformer layers (12-layer, 768-hidden, 12-heads, 110M parameters) and an output layer with SoftMax to perform the classification task on social media posts (see Figure 3). The transformer layer has two sub-layers: a multi-head self-attention mechanism, and a position-wise fully connected feed-forward network, followed by a normalization layer [40]. The SoftMax function for the posterior probability of j_{th} class given the input vector $\boldsymbol{x}$ is:

$$P(y = j|\boldsymbol{x}) = \frac{e^{\boldsymbol{x}^T \boldsymbol{w}_j}}{\sum_{k=1}^{K} e^{\boldsymbol{x}^T \boldsymbol{w}_k}} \quad (1)$$

where K is the number of distinct functions and $\boldsymbol{w}$ is the weighting vector. The pre-trained BERT model is then fine-tuned on specific data sets.

To assess the performance of the fine-tuned BERT model, multiple baseline models are adopted to conduct the same classification tasks. The baseline models include Bidirectional GRU (BiGRU), Bidirectional LSTM (BiLSTM), Hybrid CNN and GRU, Hybrid CNN and LSTM, Deep Pyramid CNN (DPCNN), CNN with K-Max pooling (KMax-CNN), Region-based CNN (R-CNN). The accuracies for validation and testing results are calculated to evaluate the performance of different models:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ number\ of\ posts} \quad (2)$$

Using the classifier, we can label each post to a humanitarian category so that situational information can be reliably classified with no manual effort. The superior performance of the BERT-based classifier will be shown in the case study.

C. GRAPH-BASED CLUSTERING

A primary concern related to making use of social media data for disaster management is the credibility of the data. That is because social media posts can be generated by any users without fact-checking [41]. Therefore, it is important to have a measure to assess the credibility of the situational information obtained from social media. Previous work has discovered that credible situational information is usually discussed in different posts by different users [42]. Built upon this evidence, a graph-based approach which examines pairwise similarities among the posts is employed to identify credible posts (see Figure 4) [42].

Since previous steps have labeled the posts with different topics and location entities, it is efficient and precise to do clustering on the set of posts talking about the same topic and location. Thus, the algorithm starts with embedding the texts of the posts for the same topic and location into vectors which would be compatible for computers to process. Because of a limited number of words in this set of posts, low dimensionality of the embedding vectors can also reach high

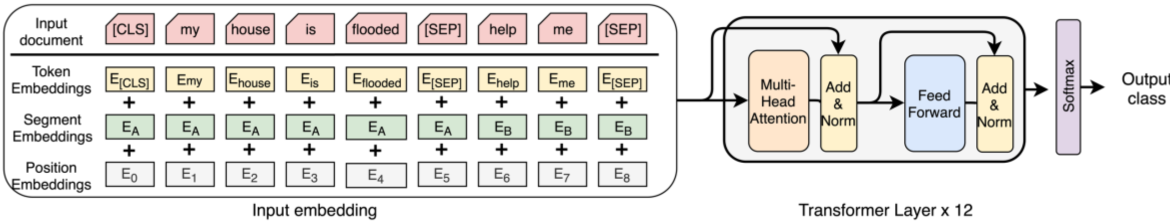

FIGURE 3. The architecture of the fine-tuned BERT classifier.

accuracy. To this end, this step adopts the well-defined and commonly used approach, TFIDF (short for term frequency–inverse document frequency), to construct vectors for each post. In this approach, the term frequency and inverse document frequency is calculated as:

$$\mathrm{tf}_{t,d} = \text{number of times that } t \text{ occurs in } d \tag{3}$$

$$\mathrm{idf}_{t,d} = \log_{10}\frac{N}{df_t} \tag{4}$$

where $\mathrm{tf}_{t,d}$ is the term frequency of the term t in the document d, $\mathrm{idf}_{t,d}$ is the inverse document frequency, df_t is the number of documents where the term t appears in the corpus, and N is the number of documents in the corpus. In this study, the posts are considered as documents for computing the embedding vectors. Then, the TFIDF weight for the term t in the document d can be obtained by:

$$\mathrm{tfidf}_{t,d} = \mathrm{tf}_{t,d} \cdot \mathrm{idf}_{t,d} \tag{5}$$

By calculating all terms in the corpus, we can obtain the vectors for each post. The next step is to examine the distance among the vectors and build the semantic graphs based on their similarities. There are multiple distance measures such as Euclidean distance and cosine similarity. Euclidean distance is a good proximity for the similarity between two vectors. However, since the posts do not have the same length, the Euclidean distance of two vectors with different lengths would be significantly large. Because of that, we measure the distance among the post vectors using cosine similarity:

$$\cos(\boldsymbol{p}_1, \boldsymbol{p}_2) = \frac{\boldsymbol{p}_1 \cdot \boldsymbol{p}_2}{|\boldsymbol{p}_1||\boldsymbol{p}_2|} = \frac{\sum_{i=1}^{|V|} p_{1i} p_{2i}}{\sqrt{\sum_{i=1}^{|V|} p_{1i}^2}\sqrt{\sum_{i=1}^{|V|} p_{2i}^2}} \tag{6}$$

where $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are the vectors of two posts, and p_{1i} and p_{2i} are the TFIDF weights of term i in two posts.

Then, we can build the semantic graph $\mathrm{G}(V, E)$ for the posts in the corpus based on the similarities among the posts. The edge between two posts is created if their similarity is greater than a threshold, which is specified case by case.

$$\mathrm{e}(\boldsymbol{p}_1, \boldsymbol{p}_2) \in E \text{ , } if \cos(\boldsymbol{p}_1, \boldsymbol{p}_2) > k \tag{7}$$

where k is a hyper-parameter which specifies the threshold of content similarity for creating an edge between two posts. The value of k ranges from 0 to 1, and can be customized for different cases.

Algorithm 1: Graph-based clustering

Input: $T = [t_1, t_2, \dots, t_n]$ # a list of tweets

Output: $\tilde{E} = [\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m]$ # clusters of tweets

$V \leftarrow$ []; E $\leftarrow$ [] # V for notes, and E for edges
$G \leftarrow \emptyset$
$vocabulary \leftarrow$ $all\ tokens\ in\ T$
for each $t \in T$
 $V = V \cup embedding(t, vocabulary)$
for each $v \in V$
 $s(v_i, v_j) \leftarrow cos(v_i, v_j)$
 if $s(v_i, v_j)$ > threshold k **then** E = E ∪ (v_i, v_j)
$G.add\_edges\_from$ E
$\tilde{E}$ ←Find_components(G)
return $\tilde{E}$

FIGURE 4. The algorithm of graph-based clustering.

Once the semantic graph is constructed, a set of credible posts can be identified by observing large components in the graph. The posts in the same component would have potential relationships: (1) the posts are talking about the same specific event such as road closure at different time stamps; (2) different posts may provide different observed evidence for the same specific event at the same time. The posts from different large components in the semantic graph may discuss different events. In addition, the posts might not be credible if they are isolated in the graph. Hence, to obtain credible posts, our approach does not include isolated posts in the analysis.

IV. CASE STUDY OF HURRICANE HARVEY

A. DISASTER CONTEXT AND DATASET

To demonstrate the application of the proposed pipeline, we conducted a case study of the 2017 Hurricane Harvey in Houston. Hurricane Harvey was a Category 4 hurricane that caused catastrophic flooding and many deaths [43]. More than 115 thousand buildings were damaged, more than 30 thousand people were displaced, and more than 17 thousand rescues were requested. We collected 21 million tweets from August 22 to September 30 in 2017 using Twitter PowerTrack API and filtering rules. The dataset includes all the tweets posted by the users whose profile locations are in Houston, and the tweets with geotags within the predefined bounding box over Houston. Furthermore, to examine the proposed pipeline and analyze the patterns of user attention to disaster events and locations on social media, we selected the period of time (i.e., August 27 – September 2) when Harvey landed in Houston.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2965550, IEEE Access

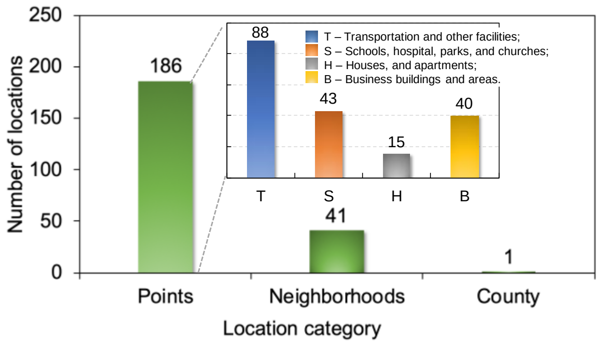

FIGURE 5. The categorical distribution of recognized locations.

Then, we filtered the original relevant tweets which were originally posted by uses without replying to, quoting to, or retweeting from any other tweets, using keywords "hurricane harvey", "hurricane", and "harvey". Subsequently, we obtain 95,606 original relevant tweets over the Houston for the selected time period to test the pipeline.

B. LOCATION RECOGNITION AND FUSION

By implementing the Named Entity Recognition approach and location fusion rules, we obtained locations that were mentioned in the relevant tweets, at three levels: point, neighborhood, and county. Because the tweets dataset was collected over Houston area which is in Harries County, obviously, there is only a single entity at the county level. In addition, we identified 186 point entities, and 41 neighborhood entities from the relevant tweets (see Figure 5).

To better understand which type of location entities gains more attention on social media, we labeled the location entities to four categories: (1) transportation and other facilities, such as roads, reservoirs, and water channels; (2) schools, hospitals, parks, and churches; (3) houses and apartments; (4) business buildings and areas, such as companies, malls, and restaurants. As shown in Figure 5, transportation facility entities gained most of the attention on social media during Hurricane Harvey. This could be due to the fact that transportation facilities play an important function in communities during crises. This is because the performance of transportation facilities does not only influence people's travel and evacuation, but also affect the dispatch of relief resources and personnel. The second and fourth categories which include schools, hospitals, churches and business buildings are also mentioned frequently on Twitter. These are notable public locations that play important functions for community response to disasters, such as serving as shelters, distributing meals, and providing healthcare. Conversely, only a few people posted about their houses or apartments because the situational information for houses and apartments are private and localized. On non-localized social media such as Twitter, users do not tend to share private information about their locations. Thus, the information related to houses and apartments is not as frequent as the information related to other location entities.

C. TWEETS CLASSIFICATION

Classifying the tweets into humanitarian categories is important for capturing the damages and needs of the affected areas. To make the analysis be comparable to other studies, we adopted the humanitarian categories and labeled data from a dataset paper which was published by Alam et al. [44] for training and validating the model. We define the categories and provide examples from our dataset as described below:

- Affected individuals: the text is related to the people who are in a disaster situation and need relief aid; for example, "*Does anyone have a canoe/boat/float of some sort to use tomorrow morning to get to our homes on Silvergate?*".
- Injured people: the text is related to injured people; for example, "*so throughout the whole hurricane harvey flooding and stuff, 5 people have been killed and 14+ have been injured.*"
- Missing people: the text is related to missing people; for example, "*Missing victims pulled from van in Greens Bayou*".
- Infrastructure and utility damage: the text is related to infrastructure disruptions such as road closure, power outage, and utility damages; for example, "*Has anyone found a way to get to I10 from south of the bayou? I am at Eldridge and Enclave*".
- Vehicle damage: the text is related to vehicle damages such as bus, car, train, and boats; for example, "*As many as 500k cars were damaged by Harvey*".
- Rescue, volunteering or donation: the text is related to the rescue efforts such as donations, search and rescue, and volunteering; for example, "*The Harris County Office of Homeland Security & Emergency Management invites residents affected by Hurricane Harvey to a recovery fair this weekend*".
- Other relevant: the text is related to the disaster, but it does not belong to any of the following humanitarian categories; for example, "*The weather is improving, and they are expecting 1-2" of rain on our end of town today and tonight so fingers crossed the water levels will remain or possibly start to go down slowly*".
- Non-relevant: the text is not related to the disasters; for example, "*I liked a @YouTube video*".

We tuned a BERT model using three data sets, Hurricane Harvey, Irma, and Maria, after merging them together and splitting them into training data and testing data. Then, the training data was further shuffled. 80% of the shuffled data was assigned for training and 20% was assigned for validation. As shown in Table 1, the number of labeled tweets for different humanitarian categories are uneven. Specifically, most of the tweets are in the category of "other relevant" and "rescue". This may lead to learning bias in our machine learning models. In this study, we employed oversampling to mitigate this issue.

The validation and testing accuracy for the fine-tuned BERT and baseline models are shown in Table 2. The validation accuracy of the fined-tuned BERT classifier is much higher than the accuracy of the existing models, which range from 70% to 75%. In addition, regarding the testing accuracy, the fine-tuned BERT model also outperforms other baseline deep learning models. Because this is an eight-class classification task, the accuracy is not very high compared to other binary classification tasks. The low accuracy primarily appears in the categories with low frequency, such as missing or found people, and vehicle damages. Due to the limitation of training data, the model would not be able to learn the features of these two categories very well. However, as shown in the following sections, the classification result is sufficiently good to capture the majority of the situational information in other categories. The results provide evidence for the capabilities of the model on precise classification. Hence, the model is capable of being applied to our learning tasks. The labeled tweets for our case study data using the fine-tuned BERT model is shown in the last column of Table 1.

TABLE 1. A BRIEF SUMMARY OF THE DATASETS FOR TRAINING AND TESTING THE MODELS.

Categories	# of tweets (Harvey)	# of tweets (Irma)	# of tweets (Maria)	# of tweets (Case data)
Affected individuals	167	100	189	3642
Injured or dead people	71	62	34	800
Missing or found people	1	6	9	0*
Infrastructure and utility damage	377	416	290	5250
Vehicle damage	25	24	8	0*
Rescue volunteering or donation	1153	830	1032	21170
Other relevant information	1540	2125	1382	50767
Not relevant	199	190	419	13977

* *Because the tweets related to missing people and vehicle damage are not identified from the case data, we removed these two categories for the following analysis.*

TABLE 2. THE VALIDATION AND TEST ACCURACY FOR BASELINE MODELS AND FINE-TUNED BERT MODEL

Models	Validation Accuracy	Test Accuracy
BiGRU	74.47%	67.88%
BiLSTM	73.40%	71.02%
CNN GRU	70.64%	67.63%
CNN LSTM	71.14%	58.09%
DPCNN	73.15%	63.74%
KMax CNN	73.40%	66.50%
R CNN	74.53%	71.77%
Fine-tuned BERT	**95.55%**	**75.37%**

D. GRAPH-BASED CLUSTERING

To extract credible situational information for the identified location entities, we implemented the graph-based clustering algorithm on the set of tweets mentioning the same location. By doing so, we could create the semantic graph in which tweets are the nodes and their content similarities are the weights of the edges. The edges are removed if the weights are lower than the selected threshold. The weighted degree of a node is the sum of weights for all edges that connect to this node.

Figure 6 depicts some examples of semantic graphs constructed for different humanitarian categories at different locations. Meyerland is a point, and Kingwood and Addicks are considered as neighborhoods. The graph of the tweets related to infrastructure damage at Meyerland is sparse, which is hard for verifying the credibility of all isolated tweets. But, the graph also has a few connected components that can indicate credible situational information. The results from these components are shown in the next section for an example related to the unfolding infrastructure damage events at Meyerland. In addition, we can observe a couple of structures in the semantic graphs for distinguishing the levels of credibility for different tweets. For example, there is a component with three nodes having the same weighted degree in both graphs for Meyerland and Kingwood. This indicates that these three tweets contain similar situational information. To reduce the information redundancy for unfolding events, we might only need one tweet from each component to represent the situation. The selection of the tweets would depend on the timestamps of the tweets and the completeness of the situational information included in the tweets. Another example of the component structure is that the three nodes in a component form a tree, instead of a fully connected component. In this case, the two leaf nodes have different situational information and both are related to the node in the middle. This indicates that the middle node provides more complete information regarding the disaster events. The structures of the three nodes are the basic building blocks which can be stacked and construct other structures involving more nodes. The example of tweets in the other relevant information category related to the Addicks entity shows a more complex giant component in the graph (see the last panel in Figure 6). As shown in the graph, there are a few nodes that have high weighted degrees. These high-degree nodes contain the situational information which is supported by other nodes. Based on the analysis of the component structures, we are able to identify the credible situational information from the tweets, which can improve the reliability of the results for unfolding events and geographical disparities.

E. EVOLUTION OF EVENTS AT DIFFERENT LOCATIONS

Understanding the evolution of disaster situations in a location is important for prioritizing disaster response and recovery actions. Based on the outcomes of the previous steps of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2965550, IEEE Access

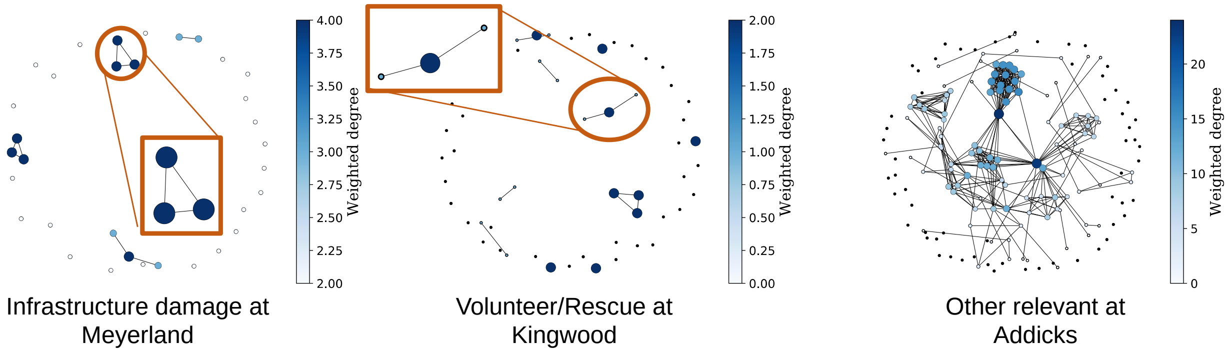

FIGURE 6. Visualization of clustering results for categorized tweets related to Meyerland, Kingwood, and Addicks.

proposed pipeline, we can identify credible situational information from the tweets and map them in a timeline for a specific location. Figure 7 provides two examples of the events at two locations: infrastructure and utility damages at Meyerland, and rescue and volunteer efforts at Beltway 8.

Meyerland is a district with a few stores and community facilities. As shown in the posts, the houses and streets were the primary entities mentioned in the tweets. At the beginning of Hurricane Harvey on August 27, Meyerland was flooded due to the heavy rainfall. Based on the posts on August 30 and September 1, the flooding was severe, and was sustained for a long duration in Meyerland. The situation timeline shows that Meyerland was a flood-prone area in which house damage was a significant problem for the residents during Harvey. The unfolding of the events related to infrastructure and utility damages at Meyerland can support the decision making for future disaster response and mitigation planning. Specifically, using the proposed pipeline, first responders can efficiently capture the disruptions and damages from the public posts. In addition, the extraction of the events and their evolution for a particular location can contribute to forensic analysis of disasters to inform future hazard mitigation and risk reduction planning.

Beltway 8 is a state highway in Houston, which also experienced flooding during Hurricane Harvey. Since many communities surrounding Beltway 8 were flooded, volunteers self-organized at Beltway 8 to help and rescue affected people. The posts show that volunteers from unflooded areas provided boats to assist affected people at the beginning of the Hurricane Harvey. After the flooding receded, volunteers also offered free meals to those affected by Harvey. In addition, the affected people also posted their requests for rescue on Twitter. These results indicate that Twitter could support the emergence of self-organization and further assist efficient relief actions in disasters. This result also implies the role that social media plays as a new way for communication of disaster relief information among individuals who did not know each other before. Our proposed pipeline can enhance the use of social media by volunteers to more efficiently identify the situation about disaster events across different locations.

To examine the credibility of the situational information identified on social media using the proposed pipeline, we check the related information in published news articles and government reports. For example, the Wall Street Journal reported that Meyerland was severely flooded due to the floodwater from Brays Bayou and needs to be built to withstand rising water [45]. In addition, the New York Times reported that volunteer rescue boats on flooded highways rescued many affected Texans [46]. The situational information on social media is validated by the recorded information from news articles, which also shows that the proposed method is capable of capturing the credible information and mitigate the effects of noise information.

In summary, the findings shown in these two examples not only demonstrate the performance of the machine learning pipeline for correctly labeling the humanitarian categories of the tweets and identifying credible situational information, but also provide insights for enhancing the efficiency and effectiveness of disaster management by creating a new communication channel for volunteers and people at risks.

F. MAPPING GEOGRAPHIC DISTRIBUTION OF EVENTS

Social media posts related to different location entities can represent the extent of disaster impacts across different locations. Greater attention to a location can be representative of the extent of damage in that location. In some cases, less attention could be due to other factors such as disruption in communication services, socio-demographic factors (the events affecting socially vulnerable populations get less attention), and the absence of points of attraction. Social media attention is not uniformly distributed across disaster-affected areas. Existing studies [47] have explored the use of geotags to examine social media posts related to a particular location. However, geotagged tweets are not sufficiently representative of social media attention in different locations. The geographical information extracted through the use of the hybrid pipeline proposed in this study provides the capability to more effectively examine the geographic distribution of social media attention to different locations.

In the case study of Hurricane Harvey, we used the proposed pipeline to automatically map the tweets to a geographical map using the coordinates for the locations mentioned in the tweets (see Figure 8). Meanwhile, we plotted the FEMA (stands for "Federal Emergency Management

IEEE Access
Multidisciplinary ⁞ Rapid Review ⁞ Open Access Journal

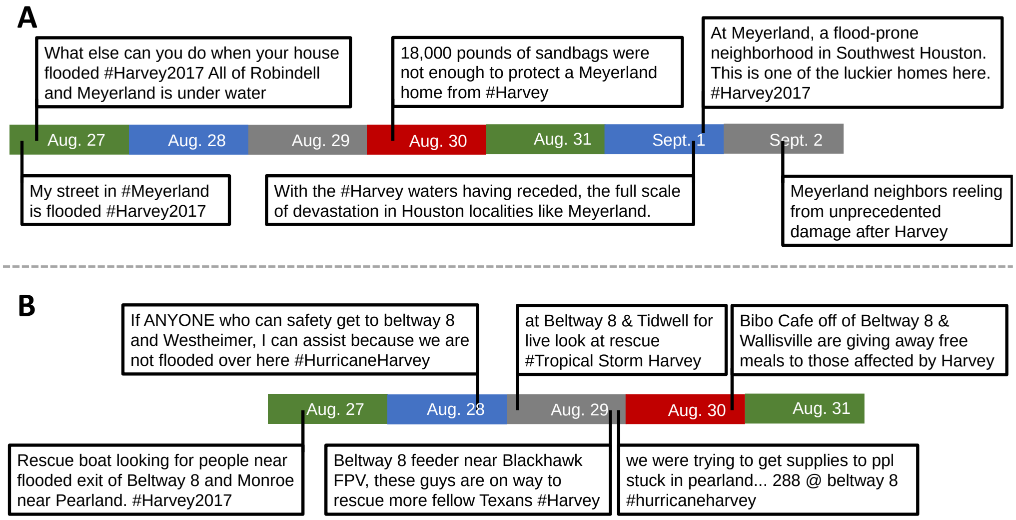

FIGURE 7. Unfolding disaster events. (A) Infrastructure and utility damage at Meyerland in Houston; (B) Rescue and volunteer efforts at Beltway 8 in Houston.

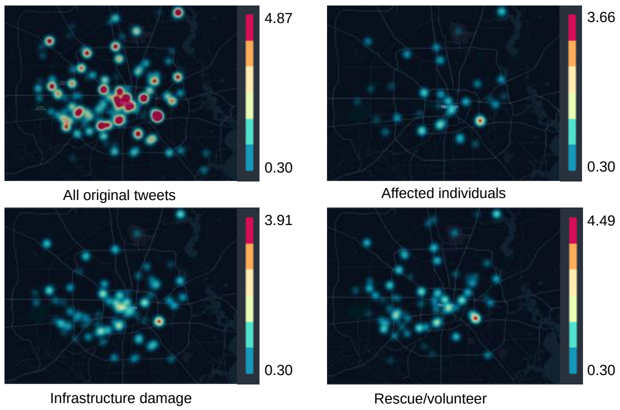

FIGURE 8. Geographical distribution of relevant tweets based on recognized entities and their topics (the values of the color bar represent the exponents of 10).

FIGURE 9. FEMA modeled building damage assessment during Hurricane Harvey (the values of the color bar represent the number of damaged buildings).

Agency") modeled building damages assessment in Houston during Hurricane Harvey (see Figure 9) [48]. Comparing the social media attention maps to the FEMA damage map, two important patterns can be observed. First, relevant tweets are unevenly distributed across Houston. The locations reported in relevant tweets are mainly located at the center and south of Houston. Although other areas were similarly affected, social media posts paid less attention to those locations. This finding suggests the influence of other factors on social media attention in disasters. Second, social media attention to location entities varies among different humanitarian categories. For example, the central part of Houston attracts more attention in terms of infrastructure damages than that in terms of affected individuals. In addition, the geographic distribution of social media attention regarding volunteering and rescue efforts is not consistent with the distributions of relevant tweets in damages or other humanitarian needs categories. This finding implies using the magnitude of social media attention to a location without considering the category of content could be misleading.

In addition to the cumulative mapping of locations and categories, the temporal information in the tweets enables us to dynamically examine the geographic distribution of social media attention to affected locations in terms of different humanitarian categories. Figure 10 shows an example of infrastructure and utility damages, and maps the relevant tweets across seven days during Hurricane Harvey. As shown in Figure 10, in general, multiple locations were mentioned by users during the first two days of Harvey, and the frequency of attention to different locations decreased with the dissipation of the disaster. This pattern can be observed from both the highlighted area and the scale of the color bar. In addition, social media attention to a few location entities (e.g., the top middle point in day 6) would also emerge at a late stage of Harvey, while these were not popular at the beginning. Although the geographic distribution of social media attention changes over time, some location entities keep a high profile (e.g., the red point at the bottom right corner). By combining the situational information identified from the graph-based clustering, we can further uncover the expansion of disaster

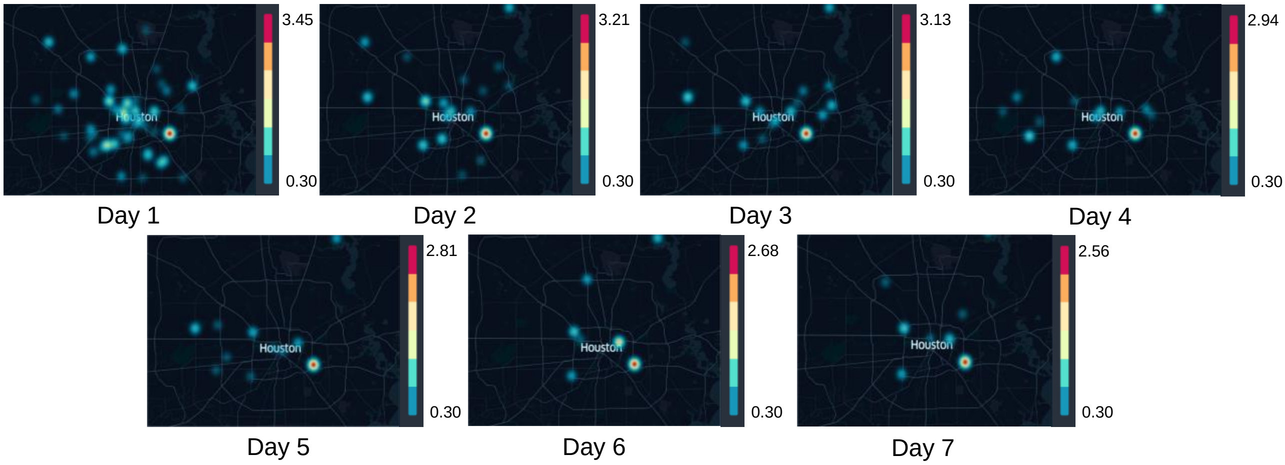

FIGURE 10. Temporal and spatial changes of social media attention across locations related to infrastructure and utility damages (the values of the color bar represent the exponents of 10).

impacts and the changes in the situation at different locations at different times.

The results and findings here show the capabilities of the proposed hybrid pipeline for using social media in disaster management. Due to uneven social media attention to location entities, the relief needs and the extent of damages based on the situational information obtained from social media might be different from the real situation of different locations in disasters. This geographic bias in social media content in disasters should be further examined along with other factors such as service availability and social demographics in future studies. Future studies can dive deeper into the problem of geographic bias on social media using the output of the hybrid pipeline presented in this study.

V. CONCLUDING REMARKS

This study presented a hybrid machine learning pipeline to automatically map the evolution of disaster events across different locations using social media posts. The proposed hybrid pipeline integrates named entity recognition, location positioning and fusion, fine-tuned BERT-based classifier, and graph-based clustering. This pipeline has two important capabilities: (1) ability to detect credible situational information for a location in evolving disaster conditions on social media; and (2) mapping the geographic distribution of social media attention in the disaster-affected area, which have not been achieved in previous studies. The application of the proposed pipeline was demonstrated in a case study of Hurricane Harvey in Houston. The results suggest: (1) integrating named entity recognition and Google Map Geocoding API, the pipeline can remove the noise corresponding to locations, and accurately obtain the coordinates for effective recognition of location entities; (2) the fine-tuned BERT-based classifier performs better than existing deep-learning classifiers in classifying the tweets into different humanitarian categories; (3) the pipeline enables the identification of credible situational information for different locations using graph-based clustering approach; and (4) the evolution and geographic distribution of disaster events can be automatically mapped by the proposed pipeline. In addition, the findings in the case study also suggest that the proposed pipeline is capable of capturing the spatial and temporal variation of social media attention to different locations and humanitarian categories. Specifically, we find that social media attention is unevenly distributed across different locations and the co-occurrence of different disaster events affects the attention to location entities.

This proposed hybrid pipeline and results in the case study have significant practical contributions and implications for existing disaster research and management processes. First, the proposed hybrid pipeline enables the automatic mapping of credible situational information for a location in disasters. This output of the pipeline can support multiple decision-making processes. For example, first responders can rapidly obtain situational information about events, and monitor the situation at different locations. In addition, by tracking the situation at different locations, volunteers and relief organizations can be aware of the allocation of relief efforts so that they can deploy their resources to the locations where people’s needs are not satisfied. Second, the evolution of the events for a location identified by the proposed pipeline from social media can be used for forensic disaster analysis. For example, the situational information on social media might indicate the interactions among multiple location entities. This information contributes to probing more deeply into the complex interdependencies of location entities, and the underlying causes of cascading impacts. Hence, the automatic mapping enabled by the proposed hybrid pipeline can enable a timely forensic investigation to inform disaster recovery and future risk mitigation. Finally, the pipeline can be further scaled to address automated mapping for several disasters occurring at the same time, when it is put into production. The relevant tweets for each disaster can be filtered using keywords and geographical bounding boxes that are specified by disaster responders with domain knowledge. Then, multiple pipelines can work parallelly with specified inputs for different disasters to automated map the situations.

The proposed hybrid machine learning pipeline offers multiple important directions for future studies. First, the BERT-based classifier was trained for hurricane cases in this study. To extend the capabilities of the proposed pipeline to other disasters such as earthquake and wildfire, existing domain adaptation approaches can be further integrated in the pipeline. Second, the proposed pipeline can be implemented on other sources of data, such as news articles, crowdsourced platforms, and remote sensing data, to extend the applications of the proposed pipeline and gain better situation awareness for disaster-affected areas. Finally, the findings in geographic mapping raise an important question related to bias in social media attention to locations in disasters. The problem of geographic biases on social media should be further addressed by examining the confounding effects of multiple related factors such as the extent of damages and socio-demographic characteristics in future studies.

REFERENCES

[1] Y. Kryvasheyeu *et al.*, "Rapid assessment of disaster damage using social media activity," *Sci. Adv.*, vol. 2, no. 3, 2016.

[2] H. Seppänen and K. Virrantaus, "Shared situational awareness and information quality in disaster management," *Saf. Sci.*, vol. 77, pp. 112–122, 2015.

[3] L. Palen and K. M. Anderson, "Crisis informatics-new data for extraordinary times: Focus on behaviors, not on fetishizing social media tools," *Science*, vol. 353, no. 6296. American Association for the Advancement of Science, pp. 224–225, 15-Jul-2016.

[4] S. A. Shah, D. Z. Seker, M. M. Rathore, S. Hameed, S. Ben Yahia, and D. Draheim, "Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?," *IEEE Access*, vol. 7, pp. 91885–91903, 2019.

[5] C. Fan, A. Mostafavi, A. Gupta, and C. Zhang, "A System Analytics Framework for Detecting Infrastructure-Related Topics in Disasters Using Social Sensing," in *Advanced Computing Strategies for Engineering*, 2018, pp. 74–91.

[6] B. A. Rachunok, J. B. Bennett, and R. Nateghi, "Twitter and Disasters: A Social Resilience Fingerprint," *IEEE Access*, vol. 7, pp. 58495–58506, 2019.

[7] C. Zhang, C. Fan, W. Yao, X. Hu, and A. Mostafavi, "Social media for intelligent public information and warning in disasters: An interdisciplinary review," *International Journal of Information Management*. 2019.

[8] M. Imran, S. Elbassuoni, C. Castillo, F. Diaz, and P. Meier, "Extracting information nuggets from disaster-related messages in social media," *ISCRAM, Baden- …*, no. May, pp. 791–800, 2013.

[9] X. Huang, C. Wang, Z. Li, and H. Ning, "A visual–textual fused approach to automated tagging of flood-related tweets during a flood event," *Int. J. Digit. Earth*, no. September, 2018.

[10] M. S. Dao, P. Quang Nhat Minh, A. Kasem, and M. S. Haja Nazmudeen, "A context-aware late-fusion approach for disaster image retrieval from social media," in *ICMR 2018 - Proceedings of the 2018 ACM International Conference on Multimedia Retrieval*, 2018, pp. 266–273.

[11] J. P. Singh, Y. K. Dwivedi, N. P. Rana, A. Kumar, and K. K. Kapoor, "Event classification and location prediction from tweets during disasters," *Ann. Oper. Res.*, May 2017.

[12] H. Mouzannar, Y. Rizk, and M. Awad, "Damage Identification in Social Media Posts using Multimodal Deep Learning," *15th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2018)*. 2018.

[13] Y. Ikawa, M. Enoki, and M. Tatsubori, "Location inference using microblog messages," in *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web Companion*, 2012, pp. 687–690.

[14] F. Morstatter, H. Gao, and H. Liu, "Discovering Location Information in Social Media," in *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2015, pp. 4–13.

[15] Q. Huang, G. Cervone, D. Jing, and C. Chang, "DisasterMapper: A CyberGIS framework for disaster management using social media data," in *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, BigSpatial 2015*, 2015, pp. 1–6.

[16] C. Fan, C. Zhang, A. Yahja, and A. Mostafavi, "Disaster City Digital Twin: A Vision for Integrating Artificial and Human Intelligence for Disaster Management," *Int. J. Inf. Manage.*, pp. 1–12, 2020.

[17] H. Gao, G. Barbier, and R. Goolsby, "Harnessing the crowdsourcing power of social media for disaster relief," *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 10–14, 2011.

[18] J. R. Ragini, P. M. R. Anand, and V. Bhaskar, "Mining crisis information: A strategic approach for detection of people at risk through social media analysis," *Int. J. Disaster Risk Reduct.*, vol. 27, no. August 2017, pp. 556–566, 2018.

[19] A. Mishler, E. S. Crabb, S. Paletz, B. Hefright, and E. Golonka, "Using structural topic modeling to detect events and cluster twitter users in the Ukrainian Crisis," in *Communications in Computer and Information Science*, 2015, vol. 528, pp. 639–644.

[20] K. Kireyev, L. Palen, and K. Anderson, "Applications of topics models to analysis of disaster-related twitter data," in *NIPS workshop on applications for topic models: text and beyond*, 2009, vol. 1.

[21] F. Ofli *et al.*, "Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response," *Big Data*, vol. 4, no. 1, pp. 47–59, Mar. 2016.

[22] M. Imran, C. Castillo, J. Lucas, P. Meier, and S. Vieweg, "AIDR: Artificial intelligence for disaster response," in *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 2014, pp. 159–162.

[23] R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, and D. Newth, "Understanding Human Mobility from Twitter," *PLoS One*, vol. 10, no. 7, p. e0131469, Jul. 2015.

[24] C. Fan, Y. Jiang, and A. Mostafavi, "Social sensing in disaster city digital twin: an integrated textual-visual-geo framework for situational awareness during built environment disruptions," *J. Manag. Eng.*, pp. 1–12, 2020.

[25] S. Kumar, X. Hu, and H. Liu, "A behavior analytics approach to identifying tweets from crisis regions," *Proc. 25th ACM Conf. Hypertext Soc. media - HT '14*, pp. 255–260, 2014.

[26] X. Guan and C. Chen, "Using social media data to understand and assess disasters," *Nat. Hazards*, vol. 74, no. 2, pp. 837–850, 2014.

[27] Y. Wang and J. E. Taylor, "Urban Crisis Detection Technique: A Spatial and Data Driven Approach Based on Latent Dirichlet Allocation (LDA) Topic Modeling," in *Proceedings of the 2018 Construction Research Congress 2018*, 2018, pp. 428–438.

[28] N. O. Hodas, G. Ver Steeg, J. Harrison, S. Chikkagoudar, E.

Bell, and C. D. Corley, "Disentangling the lexicons of disaster response in twitter," in *WWW 2015 Companion - Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1201–1204.
[29] J. Rogstadius, M. Vukovic, C. A. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo, "CrisisTracker: Crowdsourced social media curation for disaster awareness," *IBM J. Res. Dev.*, vol. 57, no. 5, 2013.
[30] S. Kumar, G. Barbier, M. Abbasi, and H. Liu, "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief," in *International AAAI Conference on Web and Social Media*, 2011.
[31] P. M. Landwehr, W. Wei, M. Kowalchuck, and K. M. Carley, "Using tweets to support disaster planning, warning and response," *Saf. Sci.*, vol. 90, pp. 33–47, Dec. 2016.
[32] M. M. Malik, H. Lamba, C. Nakos, and J. Pfeffer, "Population Bias in Geotagged Tweets," *ICWSM-15 Work. Stand. Pract. Large-Scale Soc. Media Res.*, pp. 18–27, 2015.
[33] J. Lingad, S. Karimi, and J. Yin, "Location extraction from disaster-related microblogs," in *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 1017–1020.
[34] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating non-local information into information extraction systems by Gibbs sampling," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, 2005, pp. 363–370.
[35] M. Imran, P. Mitra, and C. Castillo, "Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages," *Proc. Tenth Int. Conf. Lang. Resour. Eval. (LREC 2016)*, pp. 1638–1643, 2016.
[36] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transp. Res. Part C Emerg. Technol.*, vol. 86, no. November 2017, pp. 580–596, 2018.
[37] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," Dec. 2014.
[38] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[39] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Oct. 2018.
[40] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, vol. 2017-December, pp. 5999–6009.
[41] C. Fan, A. Mostafavi, W. Yao, and R. Huang, "A Graph-based Approach for Detecting Critical Infrastructure Disruptions on Social Media in Disasters," in *2019 52th Hawaii International Conference on System Sciences (HICSS)*, 2019, pp. 1975–1984.
[42] C. Fan and A. Mostafavi, "A graph-based method for social sensing of infrastructure disruptions in disasters," *Comput. Civ. Infrastruct. Eng.*, p. mice.12457, May 2019.
[43] K. Sullivan, A. R. Hernandez, and D. A. Fahrenthold, "Harvey leaving record rainfall, at least 22 deaths behind in Houston - Chicago Tribune," *Washington Post*, 2017. [Online]. Available: http://www.chicagotribune.com/news/nationworld/ct-hurricane-harvey-flooding-houston-20170829-story.html. [Accessed: 27-Jan-2018].
[44] F. Alam, F. Ofli, and M. Imran, "CrisisMMD: Multimodal Twitter Datasets from Natural Disasters," 2018.
[45] B. Olson, D. Frosch, and L. Cook, "Flooded Again, a Houston Neighborhood Faces a Wrenching Choice," 2017. [Online]. Available: https://www.wsj.com/articles/flooded-again-a-houston-neighborhood-faces-a-wrenching-choice-1504287854. [Accessed: 20-Nov-2019].
[46] M. Fernandez, "On Roads Turned Waterways, Volunteers Improvise to Save the Trapped and Desperate ," 2017. [Online]. Available: https://www.nytimes.com/2017/08/29/us/volunteer-rescue-crews-hurricane-harvey-houston.html. [Accessed: 20-Nov-2019].
[47] L. Zou *et al.*, "Social and geographical disparities in Twitter use during Hurricane Harvey," *Int. J. Digit. Earth*, vol. 0, no. 0, pp. 1–19, 2018.
[48] "FEMA Modeled Building Damage Assessments Harvey 20170829 | HIFLD for Harvey." [Online]. Available: https://respond-harvey-geoplatform.opendata.arcgis.com/datasets/1ac0b7c856a047e6bba2c66b32982f00_0. [Accessed: 29-Nov-2019].

CHAO FAN received the Master of Science degree in civil engineering from University of California, Davis, CA, USA, in 2017. He is a Ph.D. student in Zachry Department of Civil and Environmental Engineering at Texas A&M University, College Station, TX, USA, currently. He joined the Urban Resilience, Networks, and Informatics Lab in Fall 2017. He received a Doctoral Fellow and Best paper nomination at The HAWAII International Conference on System Sciences. His interests focus on cutting-edge interdisciplinary research at the interface of engineering, science, and policy, as well as integrated and resilience studies in disaster management using social media and data analytics methods.

FANGSHENG WU received the B.S. degree of Electronic Information Engineering from the Electronic Information School, Wuhan University, in 2018. He is currently pursuing the M.S. degree of Computer Engineering in the Department of Computer Science and Engineering, Texas A&M University. He also serves as a research assistant in the Urban Resilience, Networks, and Informatics Lab in the Department of Civil and Environmental Engineering. His research scope lies in social network data analysis and artificial intelligence techniques. He is also highly interested in exploiting various topics in social computing.

ALI MOSTAFAVI received the Master of Science in Industrial Administration degree from the Krannert School of Management and Ph.D. degree in civil engineering from Purdue University, West Lafayette, IN, USA, in 2013. He is an Assistant Professor with the Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, USA. He is the Director of the Urban Resilience, Networks, and Informatics Research Group. His research focuses on creating transformative solutions for addressing the grand challenges pertaining to the nexus of humans, disasters, and infrastructure systems.