2020

Understanding the Socio-infrastructure Systems During Disaster from Social Media Data

Kamol Chandra Roy
University of Central Florida

UNDERSTANDING THE SOCIO-INFRASTRUCTURE SYSTEMS DURING DISASTER FROM SOCIAL MEDIA DATA

by

KAMOL CHANDRA ROY

B.Sc. Bangladesh University of Engineering and Technology, 2014

M.Sc. University of Central Florida, 2018

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term

2020

Major Professor: Samiul Hasan

ABSTRACT

Our socio-infrastructure systems are becoming more and more vulnerable due to the increased severity and frequency of extreme events every year. Effective disaster management can minimize the damaging impacts of a disaster to a large extent. The ubiquitous use of social media platforms in GPS enabled smartphones offers a unique opportunity to observe, model, and predict human behavior during a disaster. This dissertation explores the opportunity of using social media data and different modeling techniques towards understanding and managing disaster more dynamically. In this dissertation, we focus on four objectives. *First*, we develop a method to infer individual evacuation behaviors (e.g., evacuation decision, timing, destination) from social media data. We develop an input output hidden Markov model to infer evacuation decisions from user tweets. Our findings show that using geo-tagged posts and text data, a hidden Markov model can be developed to capture the dynamics of hurricane evacuation decision. *Second*, we develop evacuation demand prediction model using social media and traffic data. We find that trained from social media and traffic data, a deep learning model can predict well evacuation traffic demand up to 24 hours ahead. *Third*, we present a multi-label classification approach to identify the co-occurrence of multiple types of infrastructure disruptions considering the sentiment towards a disruption—whether a post is reporting an actual disruption (negative), or a disruption in general (neutral), or not affected by a disruption (positive). We validate our approach for data collected during multiple hurricanes. *Fourth*, finally we develop an agent-based model to understand the influence of multiple information sources on risk perception dynamics and evacuation decisions. In this study, we explore the effects of socio-demographic factors and information sources such as social connectivity, neighborhood observation, and weather information and its credibility in forming risk perception dynamics and evacuation decisions.

ACKNOWLEDGMENTS

I would like to convey my heartiest gratitude to my honorable supervisor Dr. Samiul Hasan for his excellent supervision and constant support in this dissertation. In every phase of this dissertation, his guidance, resourceful insights, and wisdom directed me to the way to complete my work in time. I would also like to acknowledge the support and encouragement from my family, friends, and all the people I look for whenever I go through difficult times.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

Extreme weather events have become more common these days due to climate change and other related causes. Since 1980 the U.S. has sustained 279 billion-dollar (cost exceeds $1 billion) climate and weather disasters with a total cost exceeding $1.825 trillion [1]. Recently hurricanes Harvey, Irma, and Maria have affected millions of people in several states in the USA. These extreme events have caused significant physical and socio-economic losses [2–5]. Effective disaster management is crucial to minimize the damage and save human lives. Traditionally, disaster management is considered as a cyclic process consisting of four main phases: mitigation, preparedness, response, and recovery [6, 7]. Among these four phases the preparedness phase includes activities, training, planning for an event that cannot be mitigated and often these types of plans are taken within very short times because of the dynamic and sometimes unpredictable nature of a disaster. For example, considering the severity and forecasted path of an approaching hurricane, emergency officials often declare evacuation orders for to save human life. Such evacuation orders are expected to propagate through multiple sources (traditional media, social networks, social media, etc.) to inform people living in the risk zone. In response to the evacuation order and forecasted risk, households take evacuation decisions, which depend on a complex and dynamic process varying over time and household locations [8–10]. The delivery of accurate and timely information is crucial to create situational awareness in the affected communities. Information about the time and severity of an incident greatly helps in taking organized decisions and increases coordination among the responding organizations during disaster preparedness, response, and recovery [11, 12].

Online social media platform facilitates fast and easy exchange of information by allowing content creation, sharing, discussion, and with the ubiquitous use of social media platforms (e.g. Twitter, Facebook etc.), a massive volume of real-time data is available. Such data can provide valuable insights on human behavior during extreme events such as a hurricane [13–15].

In disaster management, social media data have been used in different contexts such as understanding and detecting natural disasters [16–18], modeling human mobility [14, 19], evacuation detection [20], monitoring epidemics [21], responding to crises [22–24], analyzing sentiment [25, 26], crisis mapping [27–29], damage assessment [18, 30–32], and so on. Social media users can also serve as social traffic sensors that traditional sensors cannot provide [33–35]. Moreover, traffic information from social media can supplement traditional physical sensors installed in road networks [36, 37]. Recently many researchers have used social media data to understand evacuation and to monitor damage/disruption. However, existing literature still lacks appropriate modeling approaches to explore the full potential of such data. Thus, novel approaches are needed to get the best of these real-time data such as social media, traffic sensors, etc.

On the other hand, researchers have adopted computational modeling approaches such as agent based model [38–40] to capture the dynamics of the collective evacuation behavior that evolves from the social interactions among households because of the challenges in collecting such empirical data [41, 42]. However, most of these existing models are based on hypothetical agents and hazards and did not consider the credibility of the information received from the mutual interaction between agents; these limitations suggest continual need of improvements of such model.

1.2 Motivation

Broadly, the motivation behind this dissertation is to make disaster management more dynamic and pro-active by using social media data. To do that, we work on four inter-dependent studies. The motivation of choosing these studies are discussed below.

1.2.1 Data-driven Evacuation Dynamics Model

Traditionally, evacuation study largely depends on post disaster survey, which is both time consuming and costly. Thus, large-scale social media data can be used for a better understanding of evacuation behaviors during hurricanes [20]. However, one of the major challenges of using social media data is to reliably model evacuation decisions from such data. To date, the studies investigating social media data are limited to inferring evacuation choices. These studies [20, 43] have mainly adopted clustering approaches that locate a user during pre-evacuation and evacuation periods. A recent case study[44] on hurricane Sandy Twitter data shows the relationship between social connectivity and evacuation decision without specifically modeling the real-time dynamics of evacuation decision-making. Using geotagged Facebook data from hurricane Irma, Harvey, and Maria, another recent study [45] has analyzed the influence of social ties on evacuation behavior. Although these studies have demonstrated the significant benefits of using location-based social media data in an evacuation context, they do not provide a modeling framework that can answer what, when, and where users participate in different activities during a hurricane. That motivate us to develop a modeling approach to understand the dynamics of evacuation decision using real-time social media data.

1.2.2 Crowdsourced and Traffic Sensor Data for Evacuation Traffic Forecast

During a hurricane, mandatory or voluntary evacuation orders are issued over a large region so that potentially impacted people can move to safer places. Under a hurricane evacuation, it is

critical for emergency agencies to ensure smooth operations of interdependent infrastructure systems and emergency services. Efficient traffic operations can maximize the utilization of existing transportation infrastructure, reducing evacuation time and stress due to massive congestion. Accurately predicting evacuation traffic is critical to plan for effective traffic operations strategies. Most previous studies in traffic prediction focus on short-term prediction (< =1 hour) which is not often useful in disaster scenario. But because of the complex dynamic nature of evacuation participation, predicting evacuation traffic demand long ahead of the actual evacuation is a very challenging task. That motivate us to develop a model for predicting long-term traffic demand by taking advance of real-time social media data and traffic sensor data.

1.2.3 Monitoring Disruptions of Critical Infrastructure During Hurricane

For effective disaster response and recovery operations, coordinated actions are required from the responsible organizations. Disruptions to infrastructure systems such as electricity/power, cell phone, internet, water, waste water, and other systems significantly affect the recovery time of a community [46]. Due to the interdependence among infrastructure systems, multiple types of disruptions (e.g., power outages, internet/cell phones, water service) are likely to co-occur during a disaster. To ensure an expedited recovery of the systems, rapid identification of the co-occurrence of disruptions is necessary so that coordinated actions can be taken by multiple agencies.

Although infrastructure performance data can be collected through physical sensing technologies such as drones, satellite, UAV etc. [47, 48], they might not be feasible due to the rapidly evolving nature of a disaster spreading over a large area [49]. Social media users have been used as sensors during disasters and several studies have found its potential for understanding situational awareness [16, 50]. Previous studies investigated social media sensing

for damage assessment [30], recovery [51], and inundation mapping [47]. Studies have also proposed query based approaches to identify topics related to critical infrastructure disruptions [49, 52]. However, these studies have not considered the co-occurrences of the types and extent of infrastructure disruptions.

During an unfolding disaster, people from the affected regions share their opinions, views, concerns, and eye witnessed events in social media platforms. Such user-generated content can provide valuable information to extract disruption-related information. However, during a disaster, emergency managers face challenges to monitor the massive volume of social media posts in real-time [53]. Thus, to get actionable information, it is important to identify whether a post indicates an actual disruption or simply expresses user views or opinions about a disruption. Recent studies have mainly focused on identifying whether a particular social media post is damage related or not [31, 54]. However, since infrastructure systems are more interconnected, co-occurrences of disruptions in multiple infrastructures are more likely. That motivate us to develop an approach for identification and visualization of multiple types of infrastructure disruptions considering whether a particular damage related post is actual or not using real-time social media data.

1.2.4 Agent-based Modeling

Although, data-driven approach has the potential in unveiling the evacuation dynamics, it is difficult to understand the complex decision making process induced by risk perception that depends on individual's socio-demographic attributes, and the complex interplay between the influence of information resources and social connection in forms of neighbors or other social networks. The agent-based modeling approach allows to understand the behavior of the agent/human under different scenario which are difficult to observe directly.

Few studies- including statistical model and agent based model - have considered multiple information sources [42, 55, 56]. The statistical models [55] of this type lack the dynamics of these information sources; the agent-based models are mostly based on synthetic data and do not consider the socio-economic and demographic factors. Most of the agent-based models rely on random seed values of evacuated households that might not be true regarding the socio-demography of a region. Moreover, the role of multiple information sources is not well-known regarding shadow evacuation. That motivates us to develop an agent-based model that combines hydrologic characteristics, socio-demographic characteristics, and multiple information sources to understand the risk perception and evacuation behavior of the households of an area.

1.3 Dissertation Objectives

This dissertation presents studies to improve our understanding, and to fill the gaps of existing studies towards making better decisions during a disaster. The dissertation focuses on the following specific objectives:

I. Develop a model using social media data that can capture the dynamics of hurricane evacuation by answering what, when and how users participate in different activities during evacuation.

II. Develop a model that can predict evacuation traffic for a longer time horizon (> 1 hour) utilizing real-time data from traffic sensors and social media.

III. Develop a method to identify and monitor the co-occurrence of multiple types of infrastructure disruptions during a disaster from social media data.

IV. Develop an agent-based model to understand the effect of multiple information sources (social network, neighborhood, weather forecasts, etc.) and its credibility in risk perception dynamics and evacuation decisions.

1.4 Contributions

The dissertation has contributions in the following ways:

- The dissertation will provide a better understanding and approach to collect evacuation decision, evacuation destination, and evacuation timing from textual, and location information of real-time social media data. Additionally, this research has proposed a model to understand the dynamics of hurricane evacuation by answering what, when and how users participate in different activities during evacuation. Such model in capturing evacuation dynamics does not exist in the literature. While the traditional evacuation study is mainly survey based, which is costly and time consuming, this model can be applied in real-time with low cost.
- We developed a data driven model to predict evacuation traffic for a longer time horizon utilizing real-time data from traffic sensors and social media. Current approaches for evacuation demand modeling are based on fixed set of expectations and cannot adjust to any changes real-time. Existing data-driven traffic prediction model is mainly focused on short-term prediction or the long-term prediction model is not optimized for disaster scenario. Thus, this research can greatly help in proactive decision making during an evacuation scenario.
- Developed an identification and visualization approach of multiple types of infrastructure disruptions using social media data which may reduce the disaster monitoring and response time.
- Developed an agent-based model to understand the influence of multiple information sources (social networks, neighbors, forecasts) and its credibility on risk perception

dynamics and evacuation decisions. This study may help in policy making by understanding complex decision-making process during an emergency event.

1.5 Structure of the Dissertation

The remainder of the dissertation is divided into five chapters which shows how each chapter contributes to the goal of making disaster management more dynamics and pro-active. From chapter two to chapter five, each chapter outlines our specific objective in the larger context of disaster management, points out the limitation of earlier research, and finally our inter-disciplinary framework and modeling approach and estimation of the result are discussed in details to illustrate how our objective and proposed approach/framework/methodology contributes to the existing literature as well as to the practical applications.

In chapter 2, we develop an input output hidden Markov model (IO-HMM) to infer evacuation decisions from user tweets. This chapter contributes to the objective one of this dissertation. To infer the underlying evacuation context from tweet texts, we first estimate a word2vec model from a corpus of more than 100 million tweets collected over four major hurricanes. Using input variables such as evacuation context, time to landfall, type of evacuation order, and the distance from home, the proposed method infers what activities are made by individuals, when they decide to evacuate, and where they evacuate to. To validate our results, we have created ground truth data, collected during hurricane Irma, of 324,012 tweets posted by 4,046 unique users. Our findings show that the proposed IO-HMM method can be useful in inferring evacuation behavior in real time from social media data. As traditional survey-based studies are infrequent, costly, and often performed at a post-hurricane period, the proposed method can be very useful to practitioners for predicting evacuation behavior as a hurricane unfolds in real time.

In chapter 3, we develop a machine learning approach Long-Short Term Memory Neural Networks (LSTM-NN) and trained it using different combinations of input features and forecast horizon. We compare our prediction results against a baseline prediction and existing machine learning models. Results show that the proposed LSTM-NN model can predict evacuation traffic demand well up to 24 hours ahead. The proposed LSTM-NN model can significantly benefit future evacuation traffic management. This chapter contributes to the objective two of this dissertation.

In chapter 4, we develop a method to detect co-occurrence of multiple types of disruptions and their locations. We propose a multilabel classification approach that can detect multiple types of disruptions along with the disruption status (actual disruption or not). We validate our approach using Twitter data collected during two real-world hurricanes: Hurricane Matthew and Irma. This chapter contributes to the third objective of this dissertation.

Chapter 5 contributes to the final objective of this dissertation by developing an agent-based model that captures the influence of multiple information sources in risk perception dynamics and evacuation decisions. In this study, we have integrated socio-demographic factors, forecasted flood depth, social network opinion dynamics, and neighbor's activity observation to dynamically model risk perception and evacuation decisions.

Finally, in chapter 6 we conclude the dissertation by discussing our overall finding the chapters, citing the limitations, and providing directions for future studies.

CHAPTER 2: MODELING THE DYNAMICS OF HURRICANE EVACUATION DECISIONS FROM TWITTER DATA: AN INPUT OUTPUT HIDDEN MARKOV MODELING APPROACH

2.1 Introduction

[1]Large-scale social media data can be used for a better understanding of evacuation behaviors during hurricanes [20]. However, one of the major challenges of using social media data is to reliably model evacuation decisions from such data. To date, the studies investigating social media data are limited to inferring evacuation choices. These studies [20, 43] have mainly adopted clustering approaches that locate a user during pre-evacuation and evacuation periods. A recent case study [44] on hurricane Sandy Twitter data shows the relationship between social connectivity and evacuation decision without specifically modeling the real-time dynamics of evacuation decision-making. Using geotagged Facebook data from hurricane Irma, Harvey, and Maria, another recent study [45] has analyzed the influence of social ties on evacuation behavior. Although these studies have demonstrated the significant potential of using location-based social media data in an evacuation context, they have not developed any modeling framework that can answer *what*, *when*, and *where* users participate in different activities during a hurricane.

In this chapter, we present a modeling approach for understanding the dynamics of hurricane evacuation from social media data. In particular, we have developed an input-output hidden Markov model (IO-HMM) to infer evacuation behavior from social media data. We have gathered large-scale Twitter data during hurricane Irma and used the spatio-temporal and contextual sequences from this data to run the proposed model. Hurricane Irma, the largest storm ever recorded in the Atlantic Ocean, made its landfall on the southern coastal areas of Florida. The

[1] Roy, KC, Hasan, S. Modeling the dynamics of hurricane evacuation decisions from twitter data: an input output hidden markov modeling approach. Unser 2nd review in Transportation Research Part C: Emerging Technologies

storm generated a massive amount of social media posts nationwide, especially in Florida. This chapter has the following contributions:

- We implement a process to gather hurricane evacuation information from geo-tagged Twitter data. We validate the results by manually checking locations and tweet texts of the users. As traditional survey data is costly and often confined with small geographic region, this type of data can be used for understanding evacuation behavior during hurricane alongside with traditional approach.
- We develop a Word2Vec model to extract contexts based on the tweets collected from multiple hurricanes (Sandy, Matthew, Harvey, and Irma). The model has been trained using more than 100 million tweets having about 882.54 million words (after filtering out the stop words, punctuations, emoticon, URLs). This model can contribute in future research to determine disaster contexts from Twitter data.
- We develop an input output hidden Markov model from the sequences generated from user tweets. To the best of our knowledge, this is one of the first studies that use social media data for modeling the dynamics of hurricane evacuation decisions. The model can capture the dynamics of hurricane evacuation by answering what, when, and how users participate in different activities during a hurricane.

2.2 Literature Review

During a hurricane, timely evacuation is critical to reduce hazard risks and save human lives [57, 58]. Despite the importance of evacuation, some people choose not to evacuate [59]. Therefore, a thorough understanding of the determinants of evacuation behavior is needed to protect the loss of lives, especially for the vulnerable communities [60]. Many studies have investigated population response during hurricanes from different perspectives, particularly focusing on evacuation

choices [61]. These topics include: evacuation decision making [60, 62–64], evacuation time [65–67], evacuation demand [68], destination choice [69], and mode and route selection [70, 71]. However, most of these studies are based on post-disaster household surveys collecting information on population behavior instead of real-time dynamics. Studies [72–74] have developed high fidelity agent-based models to predict population responses in future hurricanes. One of the major shortcomings of these models is that factors influencing evacuation decisions do not change over time. Although a few models [75, 76] considered the dynamics of evacuation decision-making process, these models depended on post-disaster surveys, mainly focusing on household characteristics with limited transferability (across regions, communities, and disaster contexts) [20, 77]. Survey data have limitations in capturing the dynamic nature of the evacuation decision-making process [8].

However, hurricane response is a dynamic event with significant changes and uncertainties involving parameters beyond household characteristics. During a hurricane, emergency agencies and weather services issue frequent advisories providing information on the hurricane's projected trajectory and category, wind speed, rainfall, storm surge, evacuation warning etc. Local and national news channels disseminate information on the present condition of the hazard and traffic situation. Context awareness, considering all these dynamic factors, plays a critical role for a large number of populations to decide whether to leave or not. Lee et al. explored the dynamics of visiting patterns to the weather-related websites during Hurricane Katrina [78]. Yabe et al. developed a web-search query-based evacuation prediction model [79]. These studies mainly focused on understanding risk perceptions without modeling the spatial-temporal dynamics of evacuation behavior. As an alternative to relying on static post-disaster surveys, dynamic predictive models can be built employing real-time information received from multiple sources

including individuals, transportation facilities, and emergency services. For instance, Meyer et al. studied the dynamics of risk perception by using survey data collected during an approaching hurricane [80]. Studies also developed a physics-based hazard modeling approach to simulate evacuation uncertainty considering the physical interaction among multiple hazard components [81, 82]. However, these studies were based on simulated environments and did not use real-time information available from different sources[83–85]. Evacuation models can utilize the vast amount of streaming data available from social media, giving us real-time insights on individual actions during evacuations [20, 43, 86, 87].

Recently, the role of social media in a disaster management context has gained a significant attention, mainly from the perspectives of crisis communication [12, 15, 87, 88], human mobility analysis [19, 89–91], nowcasting damage assessment [30], and event detection [86, 92]. However, its potential in understanding evacuation behavior is still underexplored. Existing studies on inferring evacuation decisions from social media data found home locations and displacements to determine if a user has evacuated or not. Chaniotakis et al. [43] used a density based clustering approach to identify home and geotagged tweet counts during an evacuation order to identify evacuation decision. Using hurricane Matthew data, Martin et al. [20] showed that Twitter data can be used to understand evacuation compliance behavior. This study considered user median locations during a normal period as their homes and median locations during a hurricane as their evacuation destinations. Using similar approach on hurricane Sandy twitter data, Kumar and Ukkusuri [44] studied the evacuation decision of New York City residents in relation to the social connection of the users, distance from coastline, and time to evacuation. They have found that higher number of social ties (number of friends, followers) decrease the likelihood to evacuate. A recent study using Facebook data of hurricane Irma, Harvey, and Matthew found a similar result

that social ties decrease the likelihood to evacuate [45]. However, these studies did not capture the dynamics of individual evacuation decisions requiring a modeling framework that can infer evacuation choices from geo-location data.

In this chapter, we present an input output hidden Markov model to infer evacuation behavior from Twitter data. Hidden Markov models (HMMs) relate a sequence of observations to a sequence of hidden states that explain the observations [93]. HMMs have been widely used in speech recognition [94], protein topology [95], social science [96], and activity modeling [97]. HMMs have been used to classify activity categories considering spatiotemporal features [98] and to determine activity-location sequence from geo-location data [99]. Duong et. al [100] introduced a switching hidden semi-Markov model for online activity recognition and abnormality detection. Input-output hidden Markov model is an extension to the standard hidden Markov model for using the HMM in a supervised fashion [101]. IO-HMM has shown the added advantages over HMM to map the output sequences with the inputs in studies such as audio-visual mapping [101], price forecasting [102], hand-gesture [103] etc. Yin et. al [97] proposed an input output based modeling framework to infer urban activity patterns.

2.3 Data Preprocessing and Description

In this study, for inferring evacuation choices from social media posts, we have used Twitter data from hurricane Irma. Using its streaming API, we collected around 1.81 million tweets made by 248,763 users between September 5, 2017 and September 14, 2017. We collected the data using a bounding box covering Florida, Georgia, and South Carolina. To obtain user activities during a pre-disaster period, we also collected user-specific historical data using Twitter's rest API which allows to collect the most recent 3,200 tweets for a given user. We collected user specific data for 19,000 users who were active for at least three days between the day the first evacuation

order was issued and the landfall day, so that we have enough data for capturing the activity dynamics during the evacuation.

For our analysis, we have considered only the tweets with geo-location information. The geolocation information is provided either as a point (latitude, longitude) or a bounding box (area defined by two latitude and longitudes pairs). The point location is the exact location whereas the bounding box has different level of precision of where a tweet has been posted. We use the center point of a bounding box as the latitude and longitude of that place. To convert all the locations to a region under a geocoding system and to protect the privacy of the users, we have used geohash geocoding system with a precision of ~5 kilometers. Geohash converts a latitude, longitude pair into a short string of letters and digits depending on the precision (length of the strings) [104]. In our study, we have used a geohash of length 5, which is equivalent to a region surrounded by ~ 5 km × 5 km area and has a reasonable resolution to capture the spatial dynamics.

2.3.1 Preparing Evacuation Data

From the historical tweets of a user, we extracted the most visited place during office hours (9:00 AM to 6:00 PM) on weekdays and the most visited place during nighttime (10:00 PM to 7:00 AM). For each user, we assigned the most frequent office hour place and night hour place as office location and home location, respectively. For some users, the office and home location can be same because users may not be a worker or may have their offices within 5 km from home.

Every year Florida attracts millions of visitors from home and abroad. We adopt several steps to remove the users who came from outside of Florida (international visitors and domestic users coming from states other than Florida). Through the filtering steps, we consider only the users whose home and office locations are within Florida, whose evacuation distance is less than

2,400 km (chosen based on the literature [105, 106]), and who have returned to their home after the landfall.

In this study, we have focused on capturing the evacuation demand that is most likely to affect traffic flows on highways. Short distance evacuations (e.g., going to a nearby shelter) are not likely to impact highway traffic. Also, previous studies found that short distance evacuations are only a small percentage of the total evacuation count. During hurricane Floyd, very few evacuations were found less than 50 miles (~80.5 km); about 3.5% of the respondents chose a shelter or a church as an evacuation destination [105]. Based on hurricane Matthew Twitter data, a recent study [106] has found that evacuees are likely to move more than 200 km for an evacuation. During hurricane Irma, only 4% of the respondents were found to evacuate to a shelter [107, 108]. Moreover, some of the geotagged tweets do not have the necessary granularity (tweets with locations as a bounding box) to detect short distance evacuation. Thus, we select a threshold of 200 km to identify evacuation. After returning home, a user may not have any tweets posted from her home but may have posted from nearby locations. Thus, we select a 20 km distance threshold from someone's home to identify the return of an evacuee.

Starting from the beginning (10 days prior to landfall) of the location sequence to the landfall day, if a user has not tweeted from home or office but tweeted from somewhere else with a displacement of at least 200 kilometers, we consider that the user evacuated and the corresponding time as evacuation time. After landfall, a return is considered as the time when an evacuated user is first seen within the 20 kilometers from her home or office. We collect the information on evacuation orders from the official Twitter account of each county. We have considered the timings of the evacuation orders issued by each county. So, if someone evacuates

before the first official order, it is considered as an evacuation without an official order. We have found that 252 users have evacuated among 2,571 identified Florida users.

We have manually checked the results of the above approach of identifying a user's home location and evacuation (if any), it's destination, and timing. Please see the supporting information section for details of the manual checking process. We compare the results from the manual checking process with the results obtained from this approach. We find that both the results match with respect to evacuation time and displacement traveled during evacuation. We use this resulting data as labeled dataset for the purpose of model estimation and validation. After all the processing, our final dataset contains 38,256 geotagged tweets, posted by 2,571 users from Florida.

2.3.2 Data Exploration

Figure 2.1 shows the origins and destinations of the evacuated users. Here the identified home location of an evacuee is considered as the origin and the evacuation destination place is considered as the destination. Figure 2.1 shows the result of 252 Florida-based users after filtering out the tourists/visitors. Residents of Florida evacuated to Georgia (Atlanta was one of the major destinations), Alabama, South Carolina, and North Carolina. Some users (at right bottom of Figure 2.1, near coast) moved to places that are closer to the coast than before. This is reasonable as the projected path of hurricane Irma changed overnight on September 8, 2017. Initially, Irma was expected to hit from the east coast, but later it changed its path and was predicted to hit from the west coast. These results seem plausible according to the news update from different source during hurricane Irma [109, 110]. The majority of the evacuees were from Miami, Tampa, West Palm Beach etc. (see Figure 2.1) where mandatory evacuations were ordered.

Figure 2.2 shows the distribution plot of evacuation time and return time of the users who evacuated during Irma. Figure 2.2(a) shows the marginal and joint frequency distributions of

evacuation time and return time. Figure 2.2(b) and 2.2(c) show the probability distribution of evacuation and return time considering the type of evacuation order received. Most evacuees left within 100 hours before the landfall (September 10, 2017); 18 to 42 hours before landfall was the most frequently chosen evacuation time window. On the other hand, 78 to 102 hours after the landfall was most frequently chosen return time window. People started evacuating before the official evacuation order (see Figure 2.2 (b)). Although the pattern of evacuation time is different for voluntary and mandatory orders, the patterns of return times are almost similar (see Figure 2.2 (c)). The resulting distributions are aligned to the actual evacuation time and return time according to the concurrent news reports during hurricane Irma [111, 112].

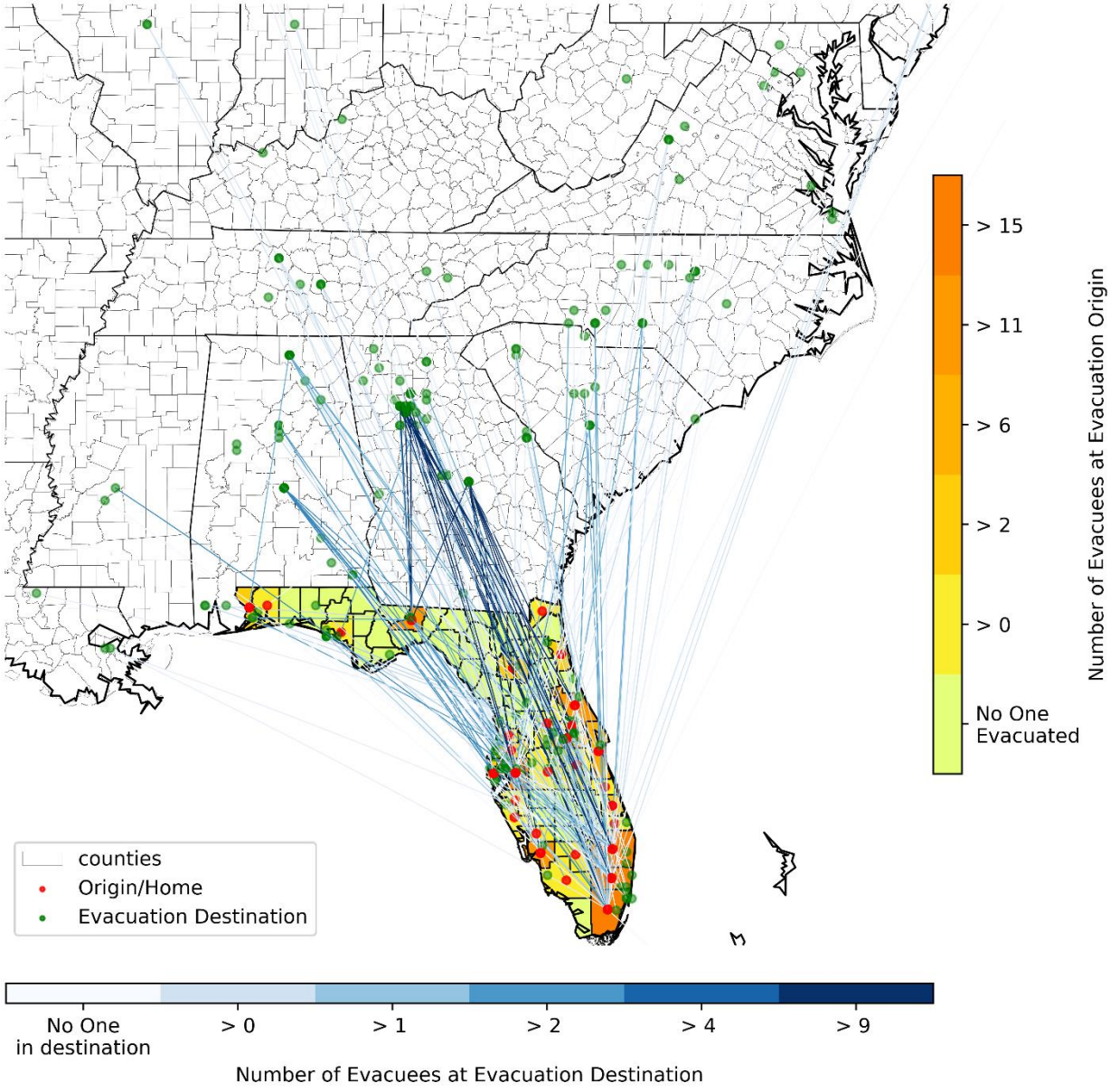
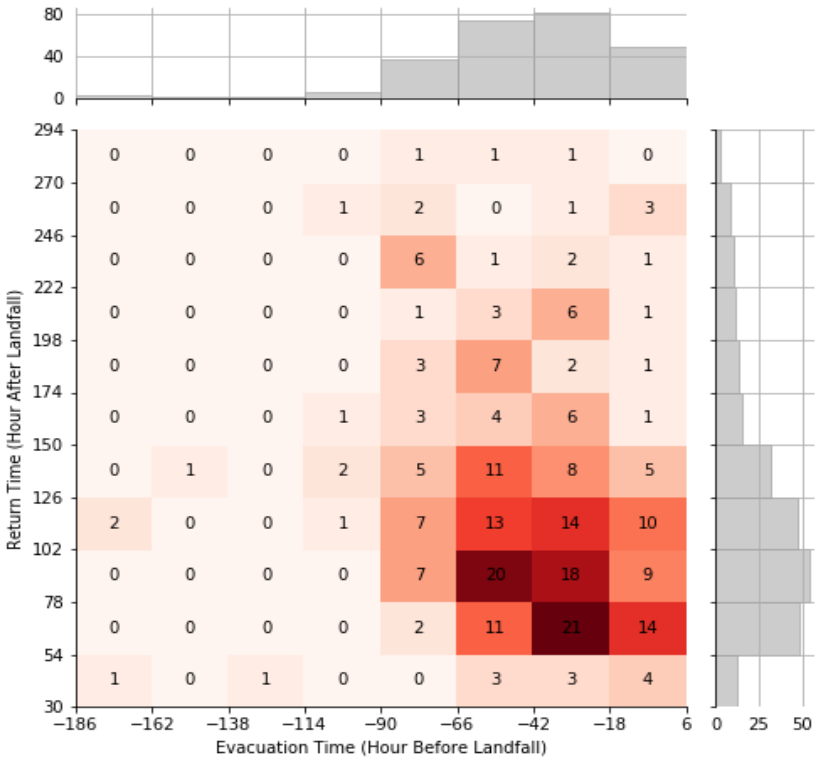
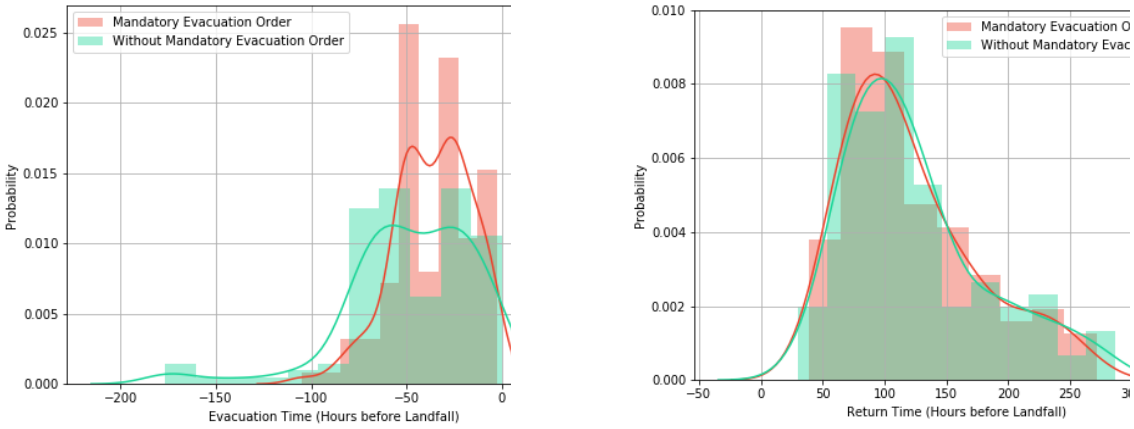


Figure 2.1 Evacuation Origin and Destination



(a)



(b) **(c)**

Figure 2.2 Distributions of Evacuation Time and Return Time during Hurricane Irma (a) Joint distribution of evacuation time and return time; the top histogram on *x* axis shows the distribution of evacuation count in 24 hour intervals; the right histogram on *y* axis shows the distribution of return time in 24 hours interval; each cell in the heatmap shows both evacuation

count and return count with respect to the corresponding 24 hour evacuation interval on *x* axis and return interval on *y* axis. (b) Probability distribution of evacuation time for mandatory and without mandatory evacuation order, and (c) Probability distribution of return time for mandatory and without mandatory evacuation order; **t**he evacuation time and return time are expressed as the time difference from landfall time (September 10, 2017), a negative value indicates a period before the landfall and a positive value indicates a period after the landfall.

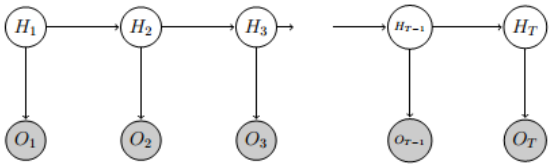
2.4 Methodology

We have used an Input Output Hidden Markov Model (IO-HMM) to identify activity sequence during a hurricane. We compare the results with a standard Hidden Markov Model (HMM). The model architecture is shown in Figure 2.3. The IO-HMM is similar to HMM, but it maps the input sequence to output sequences and applies the expectation maximization algorithm (EM) in a supervised fashion.

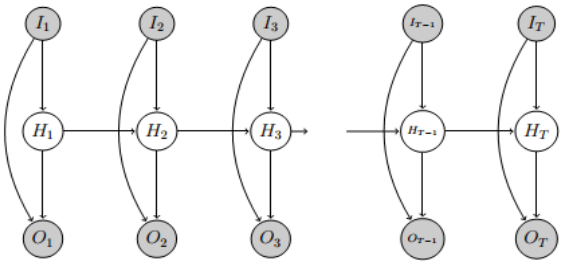
In an HMM modeling framework, the system being modeled follows a Markov process with unobserved (i.e., hidden) states. Figure 2.3(a) shows a graphical representation of an HMM. The solid circles represent the observed information and the transparent circles represent the hidden state latent variables, in our case the activity types of a user. Here, the hidden states, $(H_1, H_2 \ldots . H_T)$ are assumed to follow a Markov process that means a state, H_t depends only on the previous state, H_{t-1}; i.e., $H_t = f(H_{t-1})$. On the other hand, for the observations $(O_1, O_2, \ldots O_T)$, an observation, O_t depends only on its current state, H_t; i.e., $O_t = f(H_t)$.

Unlike the standard HMM, in IO-HMM, the hidden state H_t at time t, depends on the previous state H_{t-1} and the input I_t at time t; i.e., $H_t = f(H_{t-1}, I_t)$. Observation O_t at time t depends on both the hidden state H_t and I_t at time t; i.e., $O_t = f(H_t, I_t)$ (see Figure 2.3(b)).

Here, $I_t \in R^m$ is the input vector at time t. $O_t \in R^m$ is an output vector, and $H_t \in \{1,2, \ldots . T\}$ is a discrete state. Similar to HMM, IO-HMM has three set of parameters (θ): initial probability parameters (α), transition model parameters (β), and emission model parameters (γ).



(a) HMM



(b) IO-HMM

Figure 2.3 Graphical Model Specifying Conditional Independence Properties (a) For a Hidden Markov Model (b) For an Input Output Hidden Markov Model

The likelihood of a data sequence given the model parameters (θ) is given by:

$$L(\theta, O, I) = \sum_{H} \left(Pr(H_1|I_1;\alpha). \prod_{t=2}^{T} \Pr(H_t|H_{t-1}, I_t, \beta) . \prod_{t=1}^{T} \Pr\left(O_t|H_t, I_t;\gamma\right) \right) \qquad (2.1)$$

The model parameter is learned by expectation maximization algorithm [113]. For initial and transition models, we have used a multinomial logistic regression model. If we assume that there are k hidden states, the equation of initial probability model becomes the following:

$$\Pr(H_1 = i|I_1;\alpha) = \frac{e^{\alpha^i I_1}}{\sum_k e^{\alpha^k I_1}} \qquad (2.2)$$

where α is a coefficient matrix for initial probability model with α^i represents the coefficients for the initial state being at state i.

The transition from the state i to the state j can be modeled as:

$$\Pr(H_t = j \mid H_{t-1} = i, I_t; \beta) = \frac{e^{\beta_i^j I_t}}{\sum_k e^{\beta_i^k I_t}} \tag{2.3}$$

where β represents the transition probability matrices with the (β_i^j) being the coefficients for transitioning to next state j given the current state is i.

For the output model, we have used a linear model for a continuous outcome:

$$\Pr(O_t | H_t = i, I_t; \gamma_i) = \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(O_t - \gamma_i . I_t)^2}{2\sigma_i^2}} \tag{2.4}$$

where, γ_i represents the emission coefficient when the hidden state is i. For a hidden state i, γ_i and σ_i denote the arrays of model coefficient and standard deviation of the linear model.

And a logistic regression model is used for a categorical outcome:

$$\Pr(O_t | H_t = i, I_t; \gamma_i) = \frac{e^{\gamma_i I_t}}{\sum_k e^{\gamma_k I_t}} \; ; if \; O_t \; is \; categorical \tag{2.5}$$

where, γ_i denotes the model coefficient when the hidden state is i.

Detailed descriptions of HMMs and the associated solution algorithms can be found in ref [94]. The IO-HMM model architecture and its formulation can be found in this ref [101].

2.5 Model Development

An IO-HMM model considers data as sequences of inputs and outputs for each user. For that purpose, we need to process the data from raw tweets in that specific form. Figure 2.4 shows the sequence generation process. For a user, $\{t_1, t_2, \ldots\ldots t_T\}$ represent the times of the tweets posted at locations $\{l_1, l_2, l_3 \ldots\ldots l_T\}$, respectively. We have collected hurricane related information for each of the location (county level) such as whether the location had a mandatory

evacuation order or not, whether the location had a voluntary order or not, whether the time is before landfall or after landfall. This information is encoded as a binary variable in the sequence. Other information associated with each location is distance from home and time difference from landfall, similarity score of the posted tweet text with the evacuation context words. For simplicity, all information is not shown in Figure 2.4. We calculate the similarity score by training a word to vector model using tweets from 4 hurricanes (Irma, Matthew, Harvey, and Sandy).

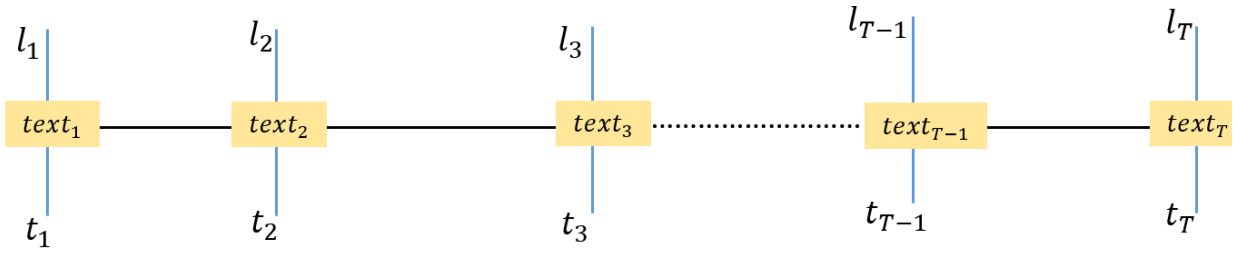


Figure 2.4 Schematic diagram of sequence generation. Here, l_i=location of the user when posting tweet i ; $text_i$= texts of the tweet i ; t_i= time of the tweet i ; and T = total number of tweets posted.

2.5.1 Inferring Evacuation Context from a Tweet

In general, the text of a tweet may reflect the underlying context such as hurricane awareness, evacuation intent, information sharing/seeking, power outage etc. We use a similarity score to quantify how similar a tweet is to an evacuation context (e.g., words such as 'evacuate', 'evacuating', 'sheltering'). We have used a vector space model called word2vec to learn the word vectors of an evacuation-related tweet.

Vector Space Model [114] is a natural language processing tool to represent texts as a continuous vector where words that appear in the same contexts share semantic meaning [115],[116]. A detailed description of how the model works is given in the supporting information.

Once a model is trained, every word in the vocabulary will have a vector representation of a length equal to the vocabulary size (see supporting information). We train a word2vec model using CBOW architecture (please see the supporting information for details) on a corpus of 100 million disaster-related tweets, collected during multiple hurricanes (Hurricanes Sandy, Harvey, Matthew, and Irma). We calculate the cosine similarity [117] between two word vectors $Word_i \; and \; Word_j$ by the following equation:

$$Similarity = \cos(\theta) = \frac{Word_i . Word_j}{||Word_i||Word_j||} \tag{2.6}$$

Here, $word_i, Word_j \in \{vocabulary \; in \; the \; model\}$,

$$k = dimension \; of \; the \; word \; vector.$$

In our study, to calculate the similarity of a word to an evacuation context, $word_i \in \{'evacuation', 'evacuating', 'sheltering'\}$ and $word_j \in \{words \; in \; a \; tweet\}$. If a window size of n is selected, then score of a tweet is calculated by summing up to n^{th} top score for the words present in the tweet. We have used the top one score to represent the similarity of a tweet with respect to an evacuation context. Similarity scores of the tweets posted at locations $l_1, l_2, l_3 \ldots \ldots l_T$ are denoted by $\{Score_{text_1}, Score_{text_2}, \ldots \ldots, Score_{text_{T-1}}, Score_{text_T}\}$.

2.6 Model Estimation

As we obtain the sequences of all the information needed, we need to specify the inputs and outputs. In IO-HMM, both inputs and outputs are available at the training stage; but after training, the model should infer the outputs given its inputs. In general, the inputs are known before the start of a transition to a new state/activity, but the outputs are not known. In our model, we choose input variables that are likely to influence the decision of a user's next activities. We select 5 input variables including: the time difference from landfall in hours represented as negative to

positive where negative means a pre-landfall period (I^1), a binary variable representing a pre-landfall or a post landfall period (I^2), an interaction variable representing the time difference from landfall only for a pre-landfall period (I^3), a binary variable representing if the user's home location is under a mandatory evacuation order (I^4), and a binary variable representing if the user's home location is under a voluntary evacuation order (I^5). As outputs, we choose two variables such as: current location's distance from home (O^1) and evacuation similarity score (word2vec score) of the tweets posted in the location (O^2).We make several assumptions for selecting the dependencies among the initial, transition, and output models of the IO-HMM structure. We assume that the transition (see Figure 2.3b) between the hidden states depend on the current state and the input variables I^1(*time difference from landfall*), I^2(*post landfall*), I^4(*user's home is under mandatory evacuation*), I^5(*user's home is under voluntary evacuation*). We did not explicitly study the research question of what factors impact people's evacuation behavior, rather we model this behavior as part of an activity dynamics process, where evacuation is considered as an activity type. The coefficient of these five input variables corresponding to an evacuation activity transition indirectly captures the factors impacting evacuation behavior. Existing literatures have also found that variables like mandatory evacuation order, voluntary evacuation orders, time of landfall significantly affect people's evacuation decisions[59, 107, 118]. We also assume that output variables include O^1(*home distance*) and O^2(*word2vec score*). These output variables depend on the current state and the input variables including I^1(*time difference from landfall*), I^3(*time difference from landfall during pre-landfall*), I^4(*user's home is under mandatory evacuation*) and I^5(*user's home is under voluntary evacuation*). Moreover, no input is chosen for the initial probability model and thus parameters will be learned by the EM algorithm only. Multinomial logistic regression is used as the transition and initial models. Since both the outputs distance from

home (O^1) and word2vec evacuation similarity score(O^2) are continuous, linear regression models are used as output models.

In the given setting of IO-HMM, to unfold the dynamics of hurricane evacuation, we choose four types of latent activities (*home activity*, *office activity*, *evacuation*, and *others*) as hidden states. Home activity and office activity represent the activities when the user stays at home and office, respectively; any activities participated at other locations are defined as "other" activities. Starting from the first evacuation order to the landfall day, if a user has only other activity but no home/office activity and is found at a location of 200 kilometers or more away from home, we labeled it as an evacuation activity. We train the IO-HMM model using the labeled sequences of 80% (*n*=202) of the evacuated users and 80% (*n*=1855) of the non-evacuated users. To validate the model, we use the data from the 20% (*n*=50) of the evacuated users and 20% (*n*=463) of the non-evacuated users. We implement the models in Python programming language using IO-HMM package [97] available here https://github.com/Mogeng/IO-HMM.

2.7 Results

In this section, we describe the results of our evacuation dynamics model. First, we apply the standard HMM model to find the learned distribution of the selected outputs. Then we interpret the results of IO-HMM.

2.7.1 HMM Results

In the HMM structure, we have four latent states/activities considering the output variables as mixtures of gaussian distributions. Figure 2.5 shows the posterior distribution of home distance and word2vec score for each latent activity. Mean distances from home is 0, 22.64, 129.18, 699.97 kilometers for home activity, office activity, other activity, and evacuation, respectively. The model has estimated higher average distance of 699.97 kilometer for evacuation activities.

Average distance for office activity is 22.64 kilometers with dispersion of 32.64 kilometer. We choose other activities as a broad category for simplicity of the model; it may include grocery shopping before hurricane, eating at restaurants, short or long trips etc., thus 129.18 kilometer of average distance with the highest dispersion of 340.44 kilometer seems reasonable.

We are interested in learning how users respond to evacuation warning in their tweets. We find the word2Vec evacuation similarity scores as 0.54, 0.48, 0.48, and 0.50 for home activity, office activity, other activity, and evacuation, respectively. Although the difference is not that much, we see a higher score during home activity and evacuation activity. This means that users have tweeted about evacuation more during evacuation or home activity (0.54 and 0.50 word2vec similarity score). It is expected since evacuated users are more likely to share posts about evacuation. Also, during a hurricane, people are more likely to share evacuation related updates from their homes.

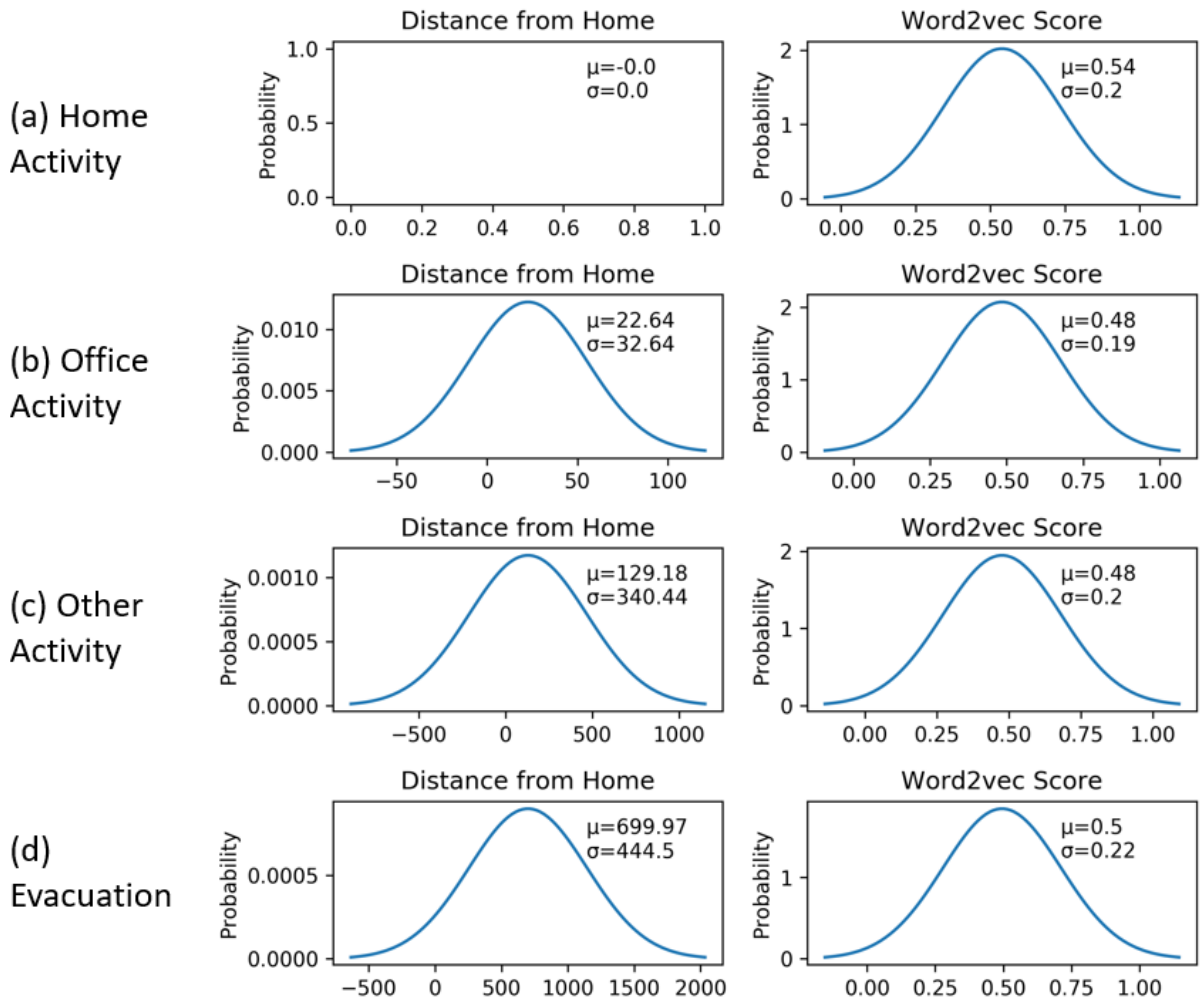


Figure 2.5 Posterior Distributions of Output Variables (Distance from Home and Word2Vec Score) for Different Activity Types (a) Home Activity (b) Office Activity (c) Other Activity (d) Evacuation

2.7.2 IO-HMM Results

Although an HMM can learn the latent activities from the observed tweets, it does not allow to incorporate contextual input variables to infer the latent activities and their relationship with the outputs. Table 2.1 shows the coefficients of the output model when applied an IO-HMM structure. We have considered other variables such as friends count and follower count; but the estimated coefficients for these variables are not significant. We have excluded these variables

from our final model. The output variable, *distance from home*, given the current state is a home activity has no coefficient. This is plausible since, for any user at home, distance from home is always zero. As expected, among all activities, the evacuation activity has the highest intercept for the *distance from home* output variable. The coefficients of I^1 *and* I^3, for an evacuation activity, are statistically insignificant, which indicates that evacuation distance does not depend on the time difference from landfall. For an evacuation activity, a positive coefficient of (I^4) indicates that if all other variables remain constant users from a mandatory evacuation zone are likely to evacuate longer distances compared to the users from no evacuation order zone. Same is true for the users from voluntary evacuation order zones (I^5). Thus, positive coefficients for both mandatory and voluntary order indicate that evacuated users from mandatory and voluntary evacuation zones will travel longer than a user from a zone with no evacuation order.

Table 2.1 Coefficients of the Output Models for IO-HMM

Output Variables	**Latent Variables**		**Input Variables**			
		Intercept	**Time difference from landfall (hour), I^1**	**Time difference from landfall*Pre-landfall period (hour), I^3**	**Home location under mandatory evacuation order, I^4**	**Home location under voluntary evacuation order, I^5**
Distance from Home	Home Activity	0	0	0	0	0
	Office Activity	22.565***	0.015***	-0.021*	1.893*	-5.724***
	Other Activity	111.639***	0.189***	-0.457***	-24.910***	10.991***
	Evacuation	585.724***	-0.125	-0.572	160.968***	176.748***
word2vec Score	Home Activity	0.667***	-0.0006***	0.002***	0.035***	-0.0272***
	Office Activity	0.577***	0.0005***	0.001***	-0.030***	0.019**
	Other Activity	0.571***	-0.001***	0.011***	0.015***	-0.013***

	Evacuation	0.563***	-0.0005***	0.002***	-0.040***	-0.036*
Note: ~p*<0.1; **~p<0.05; *~p<0.01;*						

For word2vec evacuation similarity score, home activity has the highest intercept value. It means that if all the independent variables are equal to zero-- equivalent to the landfall day and no evacuation order has been issued-- users are more likely to post about evacuation from their homes. This is reasonable as users who are not required to evacuate are more likely to stay at home and may post evacuation related tweets. For evacuation activity, a negative coefficient of I^1(-0.0005) and a positive coefficient of I^3(0.002) indicate that evacuated users post more about evacuation during a pre-landfall period as time approaches to landfall compared to a post-landfall period. In other words, an evacuated user posts more about evacuation before landfall, probably because they have already evacuated and expressing concerns who are yet to evacuate. These variables (I^1, I^3) have similar effect for home activity and other activities, indicating that in general users are expected to post more about evacuation before the landfall than in a post-landfall period. For home activity and other activity, mandatory evacuation order (I^4) has a positive coefficient and voluntary evacuation order (I^5) has a negative coefficient for word2vec evacuation similarity score. It means that if all other variables remain constant, while staying at home or participating in other activity, compared to a user from no evacuation order zone, a user from a mandatory evacuation order zones is likely to post more about evacuation whereas a user from voluntary evacuation order zones is likely to post less about evacuation. On the other hand, for evacuation activity, variables representing mandatory order zone and voluntary order zone have negative coefficients for word2vec evacuation similarity score. This indicates that evacuated users from a mandatory or voluntary order zone post less about evacuation compared to evacuees from a zone

with no evacuation order. This is plausible since evacuated users may have less time to tweet while traveling.

Table 2.2 shows the coefficients of multinomial logistic regression (MNL) models for the transition models of IO-HMM. Given the current state, there are 4 MNL models to capture the transition among the hidden states (activity types). Here, any positive coefficient means that an increase in the associated variable will increase the probability to make a transition between the corresponding states. We have 80 different coefficients to capture the dynamics of transition between any two states. We mainly focus on interpreting the coefficients associated with evacuation. For instance, if we observe the coefficients of home: evacuation (see Table 2.2), a negative sign of the variable I^2(i.e., a post landfall period) represents that if a user's current state is a home activity, in comparison to a pre-landfall period, a post-landfall period decreases the likelihood to evacuate if all other variables remain constant. It is also same for office: evacuation, other: evacuation and evacuation: evacuation transitions (see Table 2.2). These results are quite expected as individuals are less likely to evacuate after the landfall.

The coefficient of the variable *time difference from landfall* (I^1) is insignificant for the evacuation: evacuation transition; but it is significant and has negative coefficients for other 3 transitions (i.e., home: evacuation, office: evacuation, other: evacuation). This means that if everything remains constant, with increase in time difference from the landfall (as time becomes closer to landfall or away from landfall) these transitions are less likely to occur. The input variable I^4 (*home location under mandatory evacuation order*) has positive coefficients for home: evacuation, office: evacuation, and other: evacuation but it has a negative coefficient for the evacuation: evacuation transition. A plausible explanation is that a user may evacuate directly from home/office (some user's home and office are same) or may perform some other activities

(distance < 200 km) and then evacuate. The positive coefficient of I^4 indicates that compared to the users from zones with no or voluntary evacuation order, users from mandatory evacuation order zones are more likely to evacuate. A negative coefficient of I^4 for evacuation: evacuation transition means that users who evacuate from mandatory evacuation zone are less likely to remain in the evacuation state. It might be because due to their concerns about the damage of their home caused by the hurricane. The input variable I^5 (*home location under voluntary evacuation order*) has positive coefficients for home: evacuation, office: evacuation transition and negative coefficients for other: evacuation and evacuation: evacuation transition. The positive coefficient of (I^5) indicates that compared to the users from no evacuation order, the users from voluntary evacuation order are more likely to evacuate given their current activity is home or office. The negative coefficient of I^5 indicates that given the current state is other or evacuation, compared to the users from zones with no evacuation order, users from voluntary evacuation zone are less likely to evacuate or continue to maintain evacuation state.

Table 2.2 Coefficients of the Transition Models for IO-HMM

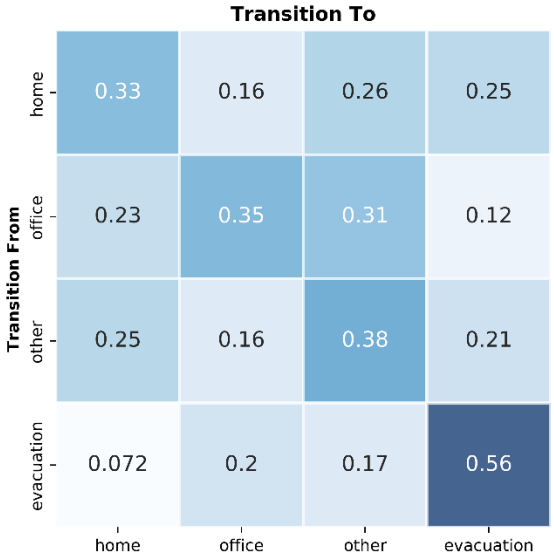
From Activity: To Activity	Intercept	Time difference from landfall (hour), I^1	Whether time is post landfall, (I^2)	Home location under mandatory evacuation order, I^4	Home location under voluntary evacuation order, I^5
Home: Home	0.034***	0.253***	0.0001***	0.521***	0.051
Home: Office	0.037***	-0.084	0.002***	-0.032	0.083
Home: Other	0.121***	0.521***	-0.001***	0.088*	0.067
Home: Evacuation	-0.191***	-0.691***	-0.0003	0.401***	0.201***
Office: Home	0.174***	0.061	0.001***	-0.062**	-0.293***
Office: Office	0.393***	0.450***	0.002***	0.588***	0.391***
Office: Other	0.282	0.364***	-0.001***	0.196***	0.069*
Office: Evacuation	-0.242***	-0.875***	-0.003***	-0.722***	0.167***
Other: Home	0.159***	0.023	0.001***	0.228***	0.065

Other: Office	-0.062***	-0.157	0.002***	0.111*	-0.088
Other: Other	0.103***	0.957***	-0.004**	0.187***	0.233***
Other: Evacuation	-0.165***	-0.734***	-0.0002	0.236***	-0.177*
Evacuation: Home	-0.099	0.103	0.010***	-0.072	0.137
Evacuation: Office	-0.132	-0.460	0.0008	0.049	0.237
Evacuation: Other	-0.053	0.025	-0.0009	-0.372***	0.147
Evacuation: Evacuation	0.283***	0.332	-0.010***	-0.394***	-0.523***
Note: ~p*<0.1; **~p<0.05; *~p<0.01;*					

Figure 2.6 shows the combined effect of different variables contributing to the transition probability from one activity to another. The color of each cell represents the probability of making a transition from the associated row activity to the associated column activity. The sum of each row equals to 1 indicating that from the current state/activity type, it will make transition to any of the four activity types. Figure 2.6 (a) and 2.6 (b) show the transition probabilities 100 hours before landfall, whereas Figures 2.6 (c) and 2.6 (d) show the transition probabilities 100 hours after the landfall. Overall, before the landfall, given a current state, it has higher probabilities to make a transition to evacuation state compared to the post-landfall period. Figure 2.6 (a) and 2.6 (b) show the differences in transition probabilities between voluntary evacuation order and mandatory evacuation order at the home location. Given the current state is a home activity, compared to the voluntary evacuation order, a mandatory evacuation order has a higher probability of evacuation (0.25) and a lower probability of transitioning to the office (0.16). In both cases, we see that given that the current state is a home activity, the probability to participate in other activity is high (0.30 and 0.26, respectively). Given the current state represents other activity, the probability to evacuate increases from 0.16 to 0.21 from zones with a voluntary evacuation order to zones with a mandatory evacuation order, respectively. Moreover, if the current state represents an office activity, the probability to evacuate decreases from 0.28 to 0.12 for a voluntary evacuation zone

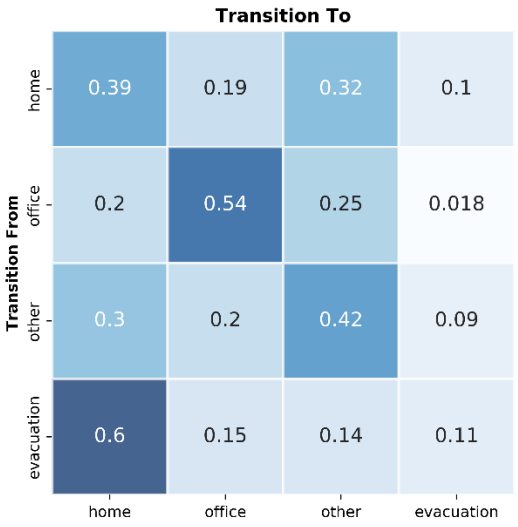
to a mandatory evacuation zone, respectively. Besides, the probabilities that an evacuated user will continue to remain evacuated for voluntary and mandatory evacuation zones are 0.45 and 0.56, respectively. Similarly, Figure 2.6 (c) and (d) show the transition probabilities after 100 hours of the landfall for users under voluntary and mandatory evacuation zones, respectively. An individual from a voluntary evacuation zone has a lower probability (0.076) to continuing to remain evacuated (see Figure 2.6c), compared to an individual from a mandatory evacuation zone (0.11) (see Figure 2.6d).

(a) 100 hours before landfall, home location under voluntary evacuation order



(b) 100 hours before landfall, home location under mandatory evacuation order

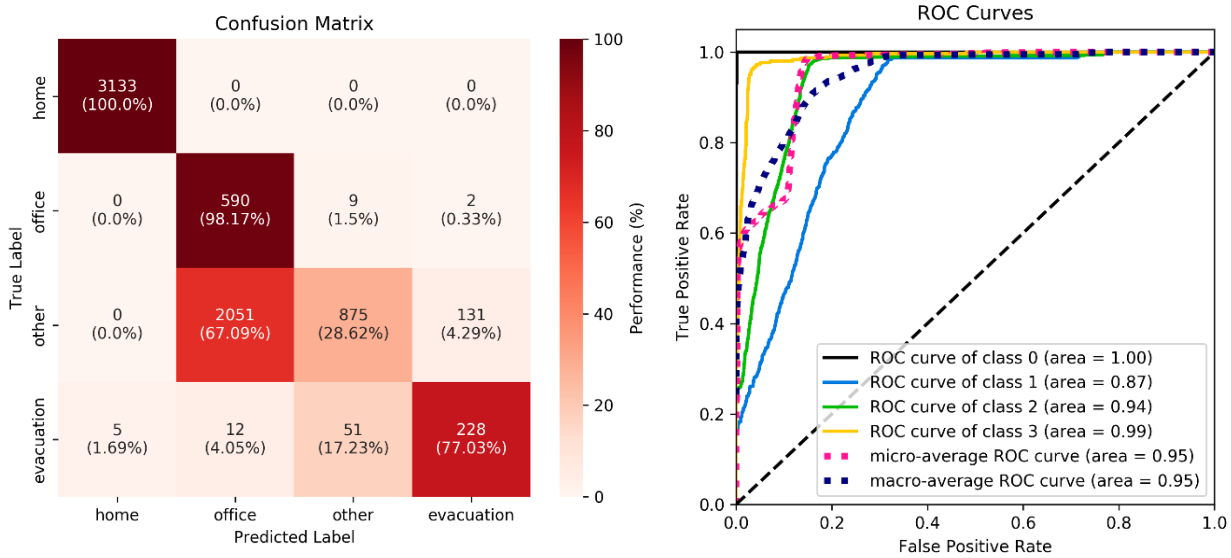
(c) 100 hours after landfall, home location under voluntary evacuation order



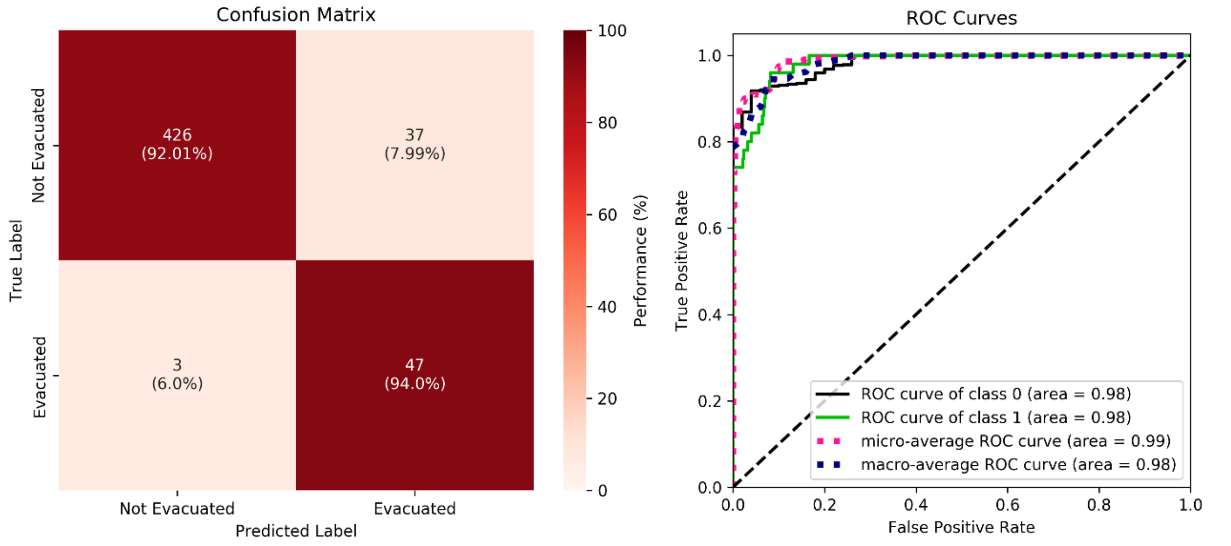
(d) 100 hours after landfall, home location under mandatory evacuation order

Figure 2.6 Activity Transition Matrices under different scenarios

There is not much difference in the probability of returning to home activity for voluntary and mandatory evacuation zone after 100 hours of hurricane landfall (0.60 and 0.59 probability that users will return to home given the current state is evacuation, for voluntary and mandatory order, respectively).



(a) for activity types (b) for activity types



(c) for identifying evacuation decisions (d) for identifying evacuation decisions

Figure 2.7 Classification Performance of IO-HMM. (a) and (b) represent the activity (home, office, other and evacuation) classification performance in terms of confusion matrix and ROC curve. (c) and (d) represent the evacuee (evacuated or not) identification performance in terms of confusion matrix and ROC curve.

Using the trained IO-HMM model, we predict the activity sequences for the test data (20% of the labelled dataset). Figure 2.7 shows the activity recognition performance of IO-HMM. The confusion matrix reports the numbers of predicted labels and the ratio of correctly predicted label to actual label. IO-HMM has 100 %, 98.17 %, 28.62% and 77.03% accuracy for recognizing home, office, other and evacuation activities, respectively (see Figure 2.7(a)). Using the standard HMM, we get 100%, 92.38%, 29.01%, and 62.51% accuracy for home, office, other and evacuation activity recognition, respectively. Thus, using a IO-HMM structure instead of a standard HMM structure improves accuracy.. Figure 2.7(b) shows the ROC curves which plot true positive rates vs. false positives rate under every possible classification threshold. For example, for home activity recognition, true positive rate answers the question when an actual activity is at home how often the model predicts it as a home activity (true home activity/all home activity). On the other hand, false positive rate for home activity recognition answers the question when actual activity is not at home how often the model predicts it as a home activity (false home activity/ all not home activity). In Figure 2.7(b) class 0, 1, 2, and 3 represent home, office, other, and evacuation activities, respectively. Area under the curve or AUC represents the classification performance where AUC is percentage of the whole box which is under the ROC curve (range 0 to 1). If any ROC curve is close the diagonal line or AUC =0.5, the model is not any better than random guessing. We can see that the model has the AUC values of 0.98, 0.87, 0.94, 0.99 for home, office, other, and evacuation activity, respectively.

We also report the performance of the IO-HMM model in identifying evacuation decision (if a user has evacuated or not) at an individual level. Using the test set, for each user, we convert the predicted activity sequence as a binary output by checking if any evacuation state is present in the predicted activity sequence or not. Then we compare the converted evacuation identification

result against our labelled data to estimate the model performance using confusion matrix and ROC curve. Figure 2.7(c) and 2.7(d) show the confusion matrix, ROC curve, respectively, for individual-level prediction. For identifying individual evacuation decision, the model has 92% and 94% accuracy for non-evacuated and evacuated users, respectively (see Figure 2.7(c)). The model has the same AUC value of 0.98 for both evacuated and non-evacuated users.

Figure 2.8 shows evacuation participation rates over time. From the labeled data, we separate evacuations in two categories: evacuation generated from zones with a mandatory evacuation order and evacuation generated from zones with a voluntary or no evacuation order. The evacuations from later zones are also known as shadow evacuation [119, 120]. We find that from our collected samples, around 65% evacuations are generated from the mandatory evacuation order zone and the remaining 35% are from a zone with either a voluntary or no evacuation order. Shadow evacuation causes additional traffic congestion and often hampers the evacuation of the actually threatened population [61]. Using the trained IO-HMM model, we predict the activity sequences of all the users (for both train and test data). We compare the predicted timing of evacuation state and the number of evacuated users with the labelled data. From Figure 2.8, we see that on aggregate the model identifies around 62% percentage of total evacuation as evacuation and around 38% as shadow evacuation. The model captures the overall trend of the evacuation timing and participation numbers.

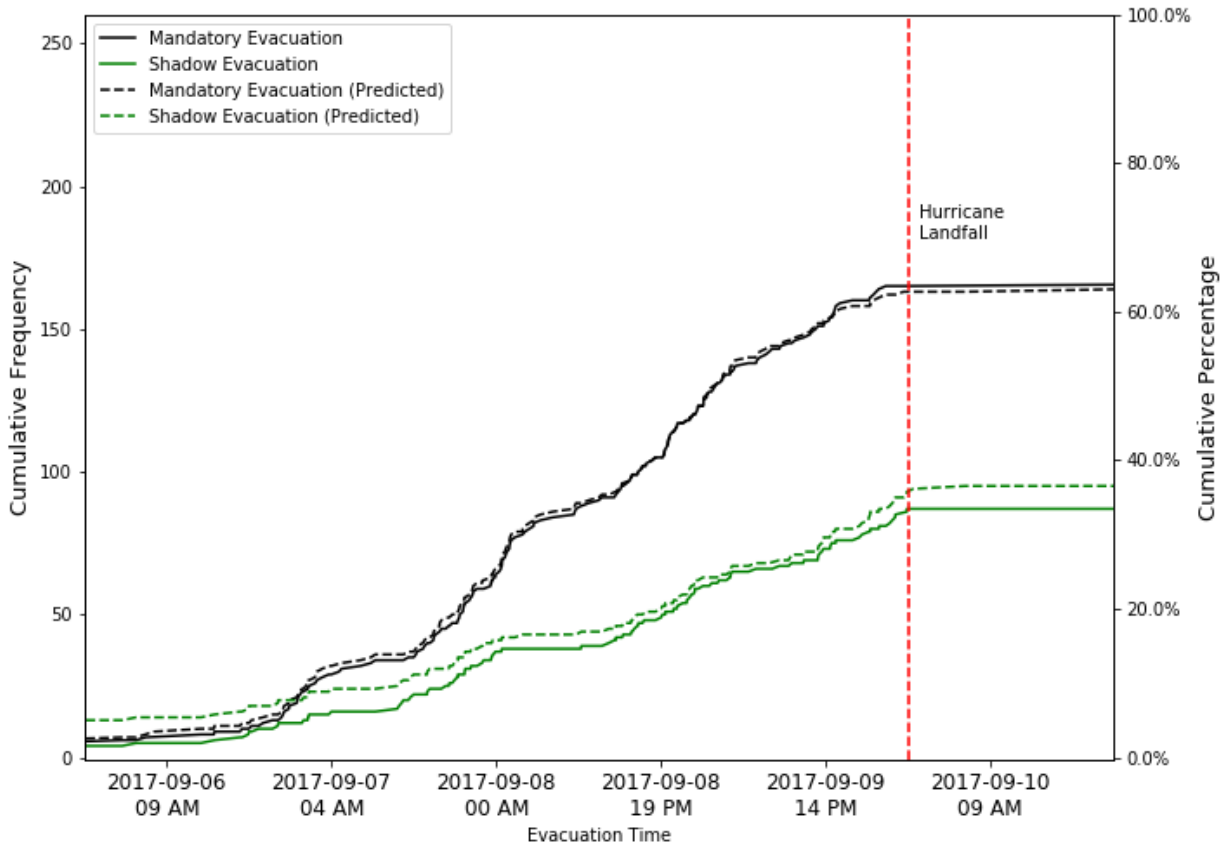


Figure 2.8 Cumulative evacuation frequencies and predicted evacuation frequencies across time.

2.8 Conclusions

To better capture the dynamics of individual-level evacuation behavior, longitudinal spatio-temporal data are needed covering both pre- and post-disaster periods. Traditional data collection approaches, such as household surveys, are static and conducted in a post-disaster period. This limits our ability to capture the dynamics of evacuation decision-making process such as determining the probability of evacuation given the states of the variables (e.g., evacuation order, projected landfall time) in the current and previous time-steps. With longitudinal data collected, we can determine the effects of the changes in variables over time on evacuation decisions. In

addition, since the data are collected in real time, we are able to capture the dynamics when the situation is evolving instead of at a post-disaster period.

In this study, we use Twitter data from Hurricane Irma to develop a model for inferring individual hurricane evacuation dynamics. We have collected evacuation data from Twitter covering all counties of Florida. Based on the tweets of active users during an evacuation period, we develop an input output hidden Markov model to infer what type of activities individuals participate, the locations and timing of those activities, when they evacuate, and where they evacuate to. We model individual participation in four activity types (home activity, office activity, other activity, and evacuation) during a hurricane.

The modeling approach provides rich insights on evacuation and other activity types during a hurricane both spatially and temporally. For instance, we have learned from real-time Twitter data to what extent individual social communication and evacuation distance depend on evacuation order type and time to landfall.

The results associated with the spatial variables (e.g., home location under a mandatory evacuation order, home location under voluntary evacuation order) indicate that if a user's home location is under a mandatory or voluntary evacuation order, he/she is likely to evacuate longer distance compared to the users under no evacuation order. We also find that users from a mandatory evacuation zone are likely to post more about evacuation during home activity and the users from a voluntary evacuation zone are more likely to post about evacuation during an office activity compared to the users from no evacuation order zone. From the activity transition dynamics, we find that: given the current activity is a home activity, the probability to evacuate increases for both mandatory and voluntary evacuation order; given the current activity is an office activity, the probability to evacuate increases for mandatory evacuation order and decreases for

voluntary evacuation order; and given the current activity is other activity, the probability to evacuate increases for mandatory evacuation order and decreases for voluntary evacuation order.

The results associated with the variables related to temporal dynamics show that evacuation distance does not depend on time difference from landfall but the evacuation related tweets (representing evacuation context) are likely to increase with decrease in time difference from landfall in pre-disaster period given all other variables remain constant. We also find that, before the landfall as the time difference decreases, the likelihood of evacuation increases. And after the landfall as the time difference from landfall increases the probability of returning to home increases. Thus, this study can capture the dynamics of evacuation behavior both spatially and temporally within a single modeling framework. Such insights for hurricane evacuation are critical for emergency management. For instance, identifying the evacuated and not evacuated population during a hurricane can make its preparation more effective and dynamic. Another benefit of our modeling framework is that we can generate the behavior of a synthetic population by simulating their activity dynamics with the parameters estimated in this study. Such simulated data from the model based on the total population of a region will allow us to determine evacuation demand in real-time.

This study has some limitations such as Twitter may have different penetration in different areas. Twitter users are not equally distributed across different age groups. Consequently, geotagged tweets may not represent the behavior of all population segments. We have assumed 200 km as threshold distance for evacuation and 20 km distance as to return thresholds. Thus, our approach cannot detect shorter distance evacuation such as relocating to higher ground or better-protected places or shelter within one's local geography. This is due to the limitations in the granularity of our data because some users have a city/county level location instead of a precise

GPS location. Other variants of HMM can be applied to get better accuracy. Also, to verify our results data from other sources are not used as they are not currently available.

In spite of the above limitations, this study adds to the growing literature on modeling the dynamics of evacuation behavior. In particular, it investigates the potential of using social media data for understanding evacuation dynamics. However, future research should focus on how to account for potential biases present in Twitter data. As social media data can be gathered in real time at large scale during a hurricane, our model can make evacuation traffic predictions and provide behavioral insights in real time. Since traditional survey data are costly and often conducted at a post-hurricane period, our method of using social media data can complement the traditional approaches of modeling evacuation behavior.

2.9 Supporting Information

2.9.1 Manual Checking of the Labeled Dataset

We created an interactive map to manually check whether a user evacuated or not. For each user, we visualized the home, office, evacuation destination (if any), the visited locations, and the tweets. We checked the tweets if there was any mention that the user was evacuating or leaving home during the evacuation period. As an example, Figure 2.9 S1 shows the snapshot of our manual checking process for a user. The locations are plotted with a 5-km precision to protect user privacy. The user, shown in Figure 2.9 S1, had home and office in Lee County, FL and evacuated to Birmingham city, Alabama. While evacuating, the user tweeted from Tampa, FL indicating that he/she was aware of hurricane Irma's changing path. We checked each user's home location, evacuation destination (if any), traveled distance, and tweet text to infer whether the user evacuated or not. Using this process, we checked 252 evacuated users and 2,319 non-evacuated users. The manual checking was performed by two individuals.

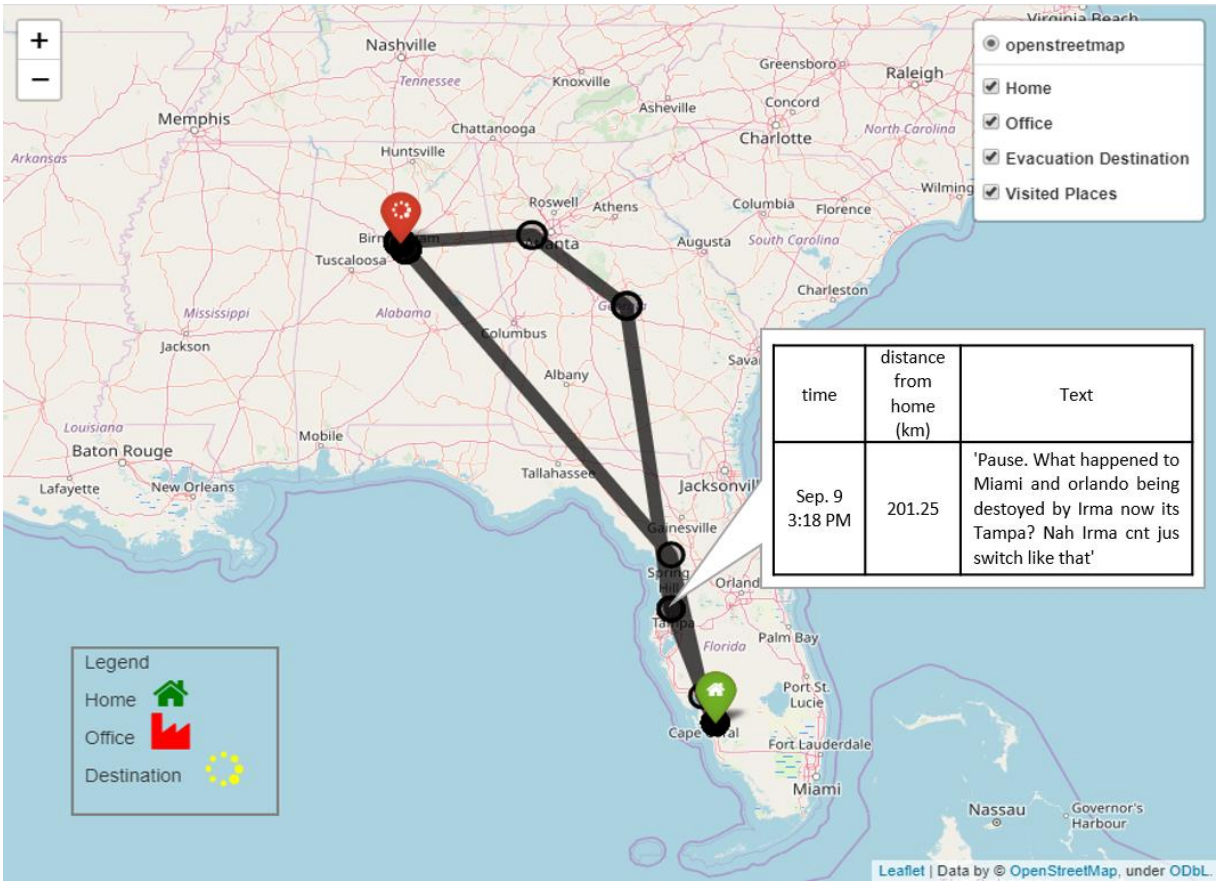


Figure 2.9 (S1) Demonstration of the Manual Checking Process. It shows a snapshot of the interactive visualization a user's home, evacuation destination, and visited places—containing tweet time, tweet text and distance from home.

2.9.2 Word2Vec Model

Word2vec is a predictive model developed by Mikolov et al.[121, 122]. It contains two distinct algorithms: Continuous Bag-of-Words (CBOW) and Skip-Gram. Skip-Gram predicts context word given a target word and CBOW predicts the target word given the context word. Details of word2vec model can be found in refs [121–123]. It is a very simple, scalable, fast to train model that can learned over billions of words of text that will produce exceedingly good word representations. Word2vec uses the theory of meaning to predict between each word and the context word. Word2Vec contains two distinct algorithms, Continuous Bag-of-Words (CBOW)

and Skip-Gram, where Skip-Gram predict context word given the target word and CBOW predict the target word given the context word. Figure 2.10 S2 shows the CBOW architecture.

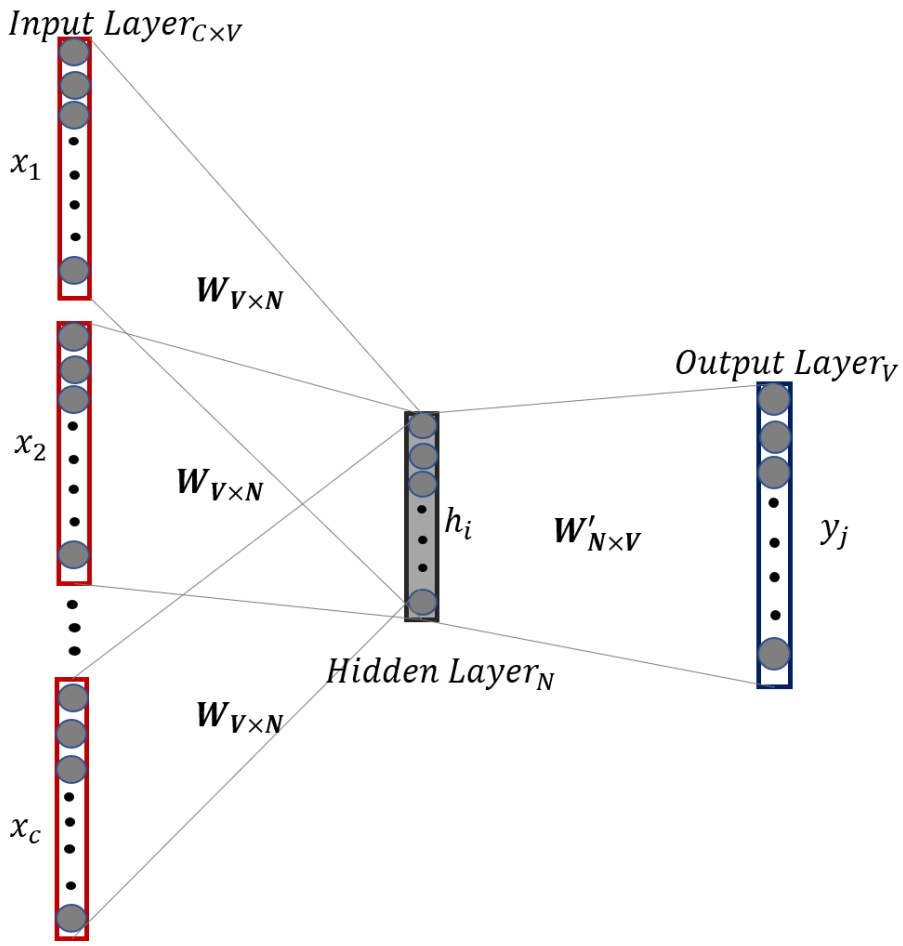


Figure 2.10 (S2) The CBOW architecture predicting the current word based on the context

In CBOW, for a window size C, the inputs are one-hot (size equal to vocabulary size, V) encoded context words $\{x_1, \ldots\ldots, x_c\}$. The hidden layer h is N-dimensional. The output/target word y_j for the context input words is also one hot encoded of size V. The input layer and hidden layer are connected by weights matrix W of dimension $V \times N$ and the hidden layer and output layer are

connected by another weight matrix W' of dimension $N \times V$. The workflow of CBOW can be described in three steps described below.

Forward Propagation

This section describes how the output is computed from the input given that the input and output weight matrixes are known. Hidden layer output is computed first from the input layer and weight matrix W. This is computed as shown in equation (2.7)

$$h = \frac{1}{C}\, W.(\sum_{i=1}^{C} x_i) \tag{2.7}$$

which is the weighted average of the input vectors and weight matrix W. Next, the input to each node of output layer is computed by the following

$$u_j = {v'_{w_j}}^{T}.h \tag{2.8}$$

Where v'_{w_j} is the j^{th} column of the output weight matrix W'. Finally, the outputs y_j of the output layer are computed by applying a soft-max function as shown in equation (2.9).

$$y_j = p\left(w_{y_j}\middle|w_1, \dots\dots, w_c\right) = \frac{\exp(u_j)}{\sum_{j=1}^{v}\exp(u_j)} \tag{2.9}$$

As the output is computed, the weight matrix W and W' can be learned from by back-propagating the errors. The process is discussed in the next section.

Learning the Weight Matrices

To learn the weight matrices, at first the W and W' are randomly initialized. By feeding the training examples sequentially and observing the predicted output, we get the error which is a function of difference between the actual and predicted output. It is also known as loss function.

The objective is to maximize the conditional probability of the output word given the input context, therefore our loss function will be the following:

$$E = -log\ p(w_0|w_I)$$

$$= -u_{j^*} - \log \sum_{j'=1}^{v} \exp\ (u_{j'})$$

$$= -v_{w_0}^T . h - \log \sum_{j'=1}^{v} \exp\left(v_{w_j'}^T . h\right) \quad (9)$$

Here j^* is the index of the actual output word. The next step is to update the weight matrices based on the gradient. The gradient of this error is computed with respect to both weight matrices and correct them in the direction of this gradient. This optimization procedure is known as stochastic gradient descent. Details of the optimization procedure can be found here [124].

Word2Vec Sample Results

We train the model with the corpus of hurricane related tweets collected from 4 hurricanes (Irma, Matthew, Harvey and Sandy). We use minimum word count=3 for preparing the vocabulary. We tran the model using a window size, C=32 for context words. Once the model is trained each word in the vocabulary will have a vector representation with its context words. The cosine similarity between two word vectors is computed using the equation 2.6. Figure 2.11 S3 shows the top 15 similar words for 'Evacuation', 'Evacuating' and 'Sheltering'. For example, similar words to evacuation

Evacuation **Evacuating** **Sheltering**

evac	0.906
evacuations	0.902
evacs	0.831
prisoninsider	0.824
evacuations.…	0.805
evacuate	0.777
mandatory	0.775
emics	0.768
evac*	0.764
evacuation…	0.762
curfews	0.759
patrols	0.743
409-283-2172	0.741
lizholmes7	0.74

evacuate	0.853
evacuated	0.848
leave	0.787
sheltering	0.784
flee	0.765
stranded	0.76
stuck	0.759
fleeing	0.755
hunkering	0.737
some	0.736
locked	0.731
evac	0.731
relocated	0.724
escape	0.721
trapped	0.72

evacuating	0.784
evacuated	0.772
relocating	0.766
shelter	0.758
housed	0.747
include…	0.738
displaced	0.732
relocated	0.723
rehabilitate	0.722
featureso	0.722
prepare	0.72
storypirates	0.715
hunkering	0.713
hunker	0.713
sos*	0.712

Figure 2.11 (S3) Top 15 Word2Vec Cosine Similarity Score of Evacuation, Evacuating, and Sheltering

contains evac, evacuations, evacs, evacuations…, evacuate etc. which may have been used as a short form of evacuation and also evac is emergency service provider name in Volusia county. Other similar words are mandatory, curfews, patrols and a cell number (4092832172) etc. Evacuation is very related with mandatory order for evacuation. And patrol, curfew is also related to evacuation because during state of emergency, state issue curfew and police patrol monitor the situation during hurricane evacuation. This cell number (409-283-2172) is the contact number of Tyler County - Sheriffs' Association of Texas which was very active during Harvey evacuation period. Thus, the result shows very good consistency in finding out the related/similar word

CHAPTER 3: PREDICTING TRAFFIC DEMAND DURING HURRICANE EVACUATION USING REAL-TIME DATA FROM TRANSPORTATION SYSTEMS AND SOCIAL MEDIA

3.1 Introduction

[2]During recent hurricanes (Matthew, Irma), massive evacuations took place in the entire Florida region especially in its coastal counties. Millions of people were under mandatory evacuation orders, creating severe congestion in major evacuation routes especially in the interstate highways (I-75 and I-95). To alleviate congestion, emergency management agencies can adopt strategies such as opening hard shoulder for traffic, contraflow operations, modified traffic control, route guidance, and staged evacuation etc. However, traffic prediction plays the most critical role to decide upon the nature and extent of such congestion management strategies. Existing works on traffic prediction mainly focus on short term (5 mins to 1 hour) prediction, which is not adequate for managing hurricane evacuations that last several days. During hurricanes, traditionally adopted short-term features such as present and past traffic conditions are not enough to make traffic predictions. Social media messages and geotagged information about the actions taken by the users can provide valuable signals for predicting evacuation traffic in the long term. The objective of this study is to investigate how real-time information from traffic and social media sensors can be used to better predict long-term traffic demand during evacuation.

We propose a machine learning approach for making long-term traffic prediction during evacuation. In particular, to predict traffic demand during a hurricane for different time horizons,

[2] Roy, KC, Hasan, S, Culotta, A, Eluru, N, Predicting Traffic Demand during Hurricane Evacuation Using Real-time Data from Transportation Systems and Social Media. Under 2nd review in Transportation Research Part C: Emerging Technologies

we propose a neural network model based on long-short term memory (LSTM-NN) architecture. We have used Twitter data from hurricanes Matthew and Irma and the corresponding traffic counts from the loop detectors in two major interstate highways (I-75 and I-95). We compare the results with a baseline forecast and other machine learning algorithms such as *K*-nearest neighbor regression (KNN regression), support vector regression (SVR), gradient boosting regression (GBR), and XGBoost regression (XGBR). Experimental results show that during hurricane evacuation, LSTM model captures the traffic demand irregularities better than the other models. In this work, we answer the following four research questions:

- *During hurricane evacuation, can we predict traffic demand for a longer time horizon utilizing real-time data from traffic sensors and social media?* We collect traffic data and Twitter data during two major hurricane evacuation periods. We use these two data sources for predicting traffic demand for a longer forecast horizon (≥1 hour).
- *How far in advance can traffic demand be predicted during evacuation using real-time data?* For that we apply the proposed models for different forecast horizons (1 hour to 30 hours) and compare the predictive performance of the models.
- *How well does the predictive model perform when one of the data sources is not available?* We apply the models for different combinations (only traffic data, only social media data, and combined) of the features and compare the predictive performance of the models.
- *How can we predict the uncertainties of the demand predictions during evacuation?* We implement a machine learning model to predict possible errors in prediction and give the prediction with 90% confidence interval.

3.2 Literature Review

Many previous studies investigated evacuation behavior during emergencies including hurricanes [125]. These studies mainly focused on understanding the factors relating to evacuation decisions [60, 126–129], mobilization time [130], departure time [67, 131] and destination choice [69, 132, 133]. Behavioral response [134] to a disaster depends on many factors [135] such as previous evacuation experience [136], receiving a warning [137], higher risk perception [136], strong social network [137], gender (female) [138] etc. increase the likelihood to evacuate. On the other hand, factors such as frequent hazard experience [139], longer residence duration [137], fear of looting [140] etc. increase the likelihood of not to evacuate.

Evacuation behavior also depends on the type of emergency events such as predictable events or evacuation with warning/notice (e.g. hurricane, flood), unpredictable events or no-notice evacuation (e.g. earthquake, chemical spills, terrorist attack)[141, 142], and short-notice evacuation (e.g. tsunami) [133]. For example, unlike hurricanes, tsunami evacuation destinations are likely to be within short distance (evacuation by foot is recommended)[143]. For destination choice of tsunami evacuation, Parady and Hato [133] proposed a spatially correlated logit model considering variables like distance, altitude difference, number of buildings, shelters, etc.; such spatial correlation is yet to be explored for hurricane evacuation.

However in many cases, the covariates used in the models are not available for demand prediction during an unfolding disaster [8, 68]. Wilmot and Mei compared five types of models (participation rates, logistic regression and 3 types of neural networks) for predicting evacuation demand [144]. Xu et. al proposed an ordered Probit model for predicting evacuation demand for a future event using data from North Carolina [68]. Studies have proposed ensemble based framework [81, 82], integrated modeling approach [64], sequential logit [129], nested logit model

[145], random parameter model [76], portfolio choice model [108] to understand and predict evacuation. Although this type of modeling approach captures individual level evacuation participation in greater detail, these approaches highly depend on surveys that are difficult to collect as a hurricane unfolds in real-time. In our study, we use real-time data for predicting the traffic demand during evacuation for longer forecasting horizon.

During a hurricane, traffic state abruptly changes depending on the time to landfall and hurricane intensity. Evacuation orders are issued considering the damaging effect of storm surge and the overall traffic impact. Evacuation process exerts significant challenges to transportation planning and operations processes [61, 146]. Litman described the planning (e.g., transportation) failures during hurricane Katrina and Rita [147]. During hurricane Katrina, only 60% of the projected vulnerable people were willing to or able to evacuate. In contrast, during hurricane Rita enormous response to evacuation orders created excessive traffic problems (e.g., 100-mile-long traffic jams, out of fuel etc.) and dozens of accidents or heat related deaths. Considering these experiences, evacuation orders were not issued during hurricane Harvey [148]. Incorporating real-time data in evacuation planning can make evacuation traffic management more flexible, pro-active, and effective.

With ubiquitous sensors and smartphone devices, many real-time data sources are available now. Traffic detectors installed in the road networks provide multi-resolution real-time data. These data sources have been used for traffic state prediction by many studies. Seo et al. provide a comprehensive review of existing methods of highway traffic state (flow, volume, speed etc.) estimation [149]. However, these studies [150–155] mainly focus on short term (5 min to 1 hour) traffic state prediction. Modeling approaches include historical average and smoothing techniques [156], auto-regressive moving average models [156], Kalman filter algorithms [156], non-

parametric regression [156], artificial neural networks (ANN) [156, 157] etc. However, during a hurricane, such short-term predictions are not adequate to adopt pro-active traffic management strategies. In addition, historical data and present traffic conditions are not enough to predict long-term traffic states because of other external factors such as unexpected events [158]. During a hurricane, traffic flow does not follow typical periodical patterns; rather it changes abruptly depending on many complex factors such as time to landfall, changes in hurricane path, evacuation orders etc.

Online social media is a major source of real-time data containing public opinion about real-world events. In disaster management, social media data have been used in different contexts such as understanding and detecting natural disasters [16–18], modeling human mobility [14, 19], monitoring epidemics [21], responding to crises [22–24], analyzing sentiment [25, 26], and so on. Social media users can also serve as social traffic sensors that traditional sensors cannot provide [33–35]. Moreover, traffic information from social media can supplement traditional physical sensors installed in road networks [36, 37]. Ming et al. have developed a social media (Twitter) based event detection and subway passenger flow prediction model under event occurrence [159]. He et al. [158] developed a regression based approach for long term traffic prediction using Twitter data. However, this study did not investigate how well the method would perform in case of emergencies such as hurricanes. Adding features based on tweet counts can improve long-term traffic volume prediction [158]. However, traffic pattern considered in these studies are either recurrent in nature or only have a peak for some hours. During hurricane evacuation, traffic pattern is more unpredictable and can be significantly different from one hurricane to another hurricane.

Existing approaches on traffic demand prediction model are not suitable in evacuation scenarios as these studies do not consider dynamic features such as time to landfall, evacuation

orders issued, and hurricane awareness that influence the temporal pattern of evacuation demand. In this chapter, we present an approach combining traffic sensor and Twitter data to predict traffic demand during hurricane evacuation for a longer forecast horizon.

3.3 Study Area and Data Description

In this study, for predicting traffic during evacuation, we have used both traffic volume and Twitter data. We collect traffic volume from two detectors: one in I-75 and the other one from I-95 interstate highway. We collect northbound volume data as we are interested in only the evacuation traffic moving from the affected regions. The detector at I-75 is located at I-75 north bound direction at mile marker 330.2 (see Figure 3.1) (detector id-9828). The detector at I-95 is located at north bound direction at zone id-10077, district 5, Florida at location I95-N US 92 (see Figure 3.1). These detectors are operated by the Florida Department of Transportation, and we have collected the data from Regional Integrated Transportation Information System (www.ritis.org). The data include traffic volume in 15 minutes intervals.

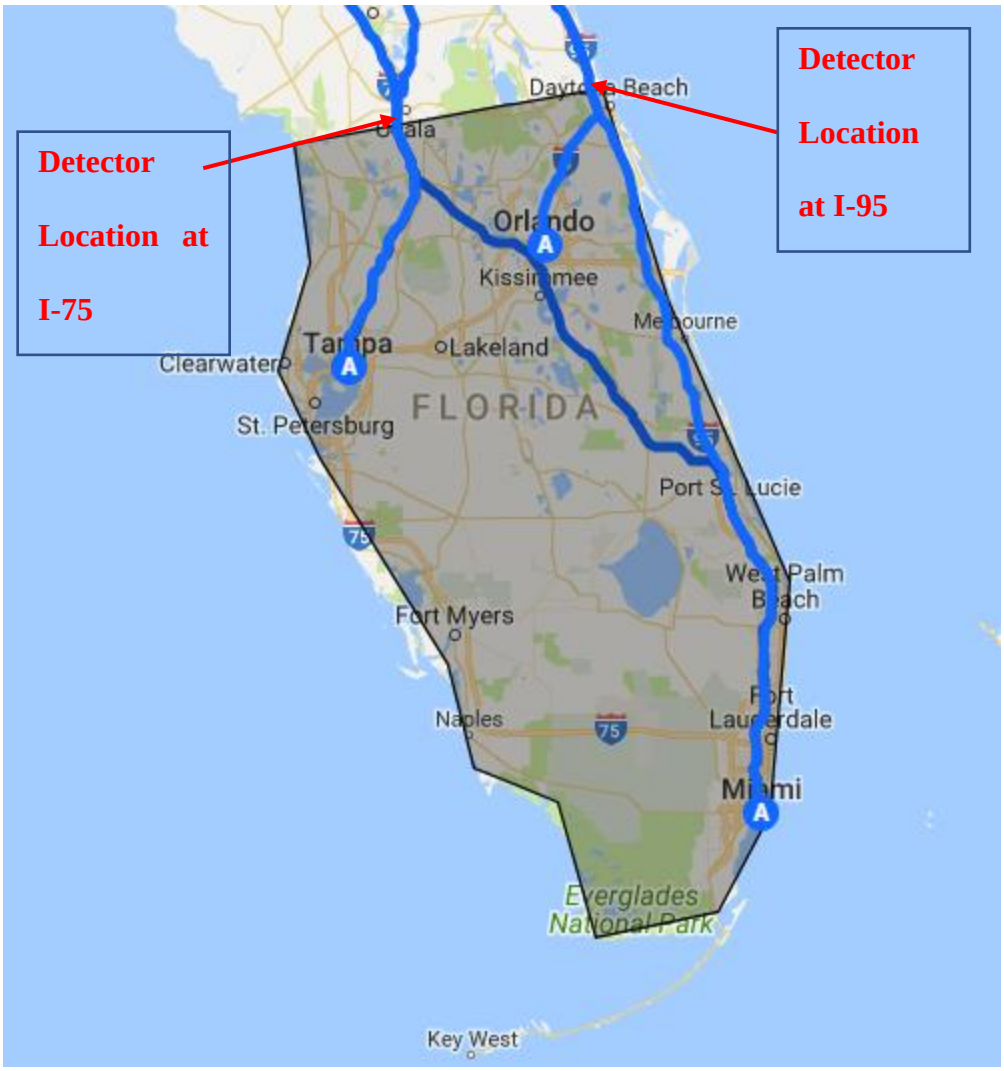


Figure 3.1 Detector Locations at I-75 and I-95 and the Study Area

For social media data, we have used Twitter data from hurricanes Matthew and Irma. We purchased hurricane Matthew data from Twitter. The data were purchased using keywords such as hurricane, matthew, hurricanematthew, huracan, huracanmatthew, huracan, storm, evacuation, evacuations, and FEMA. Matthew data contains 11.5 million tweets collected between September 25, 2016 to October 24, 2016. We collected hurricane Irma data using Twitter streaming API for a selected bounding box covering Florida, Georgia, North Carolina, and South Carolina. For

hurricane Irma, we collected around 1.8 million geotagged tweets from September 5, 2017 to September 14, 2017.

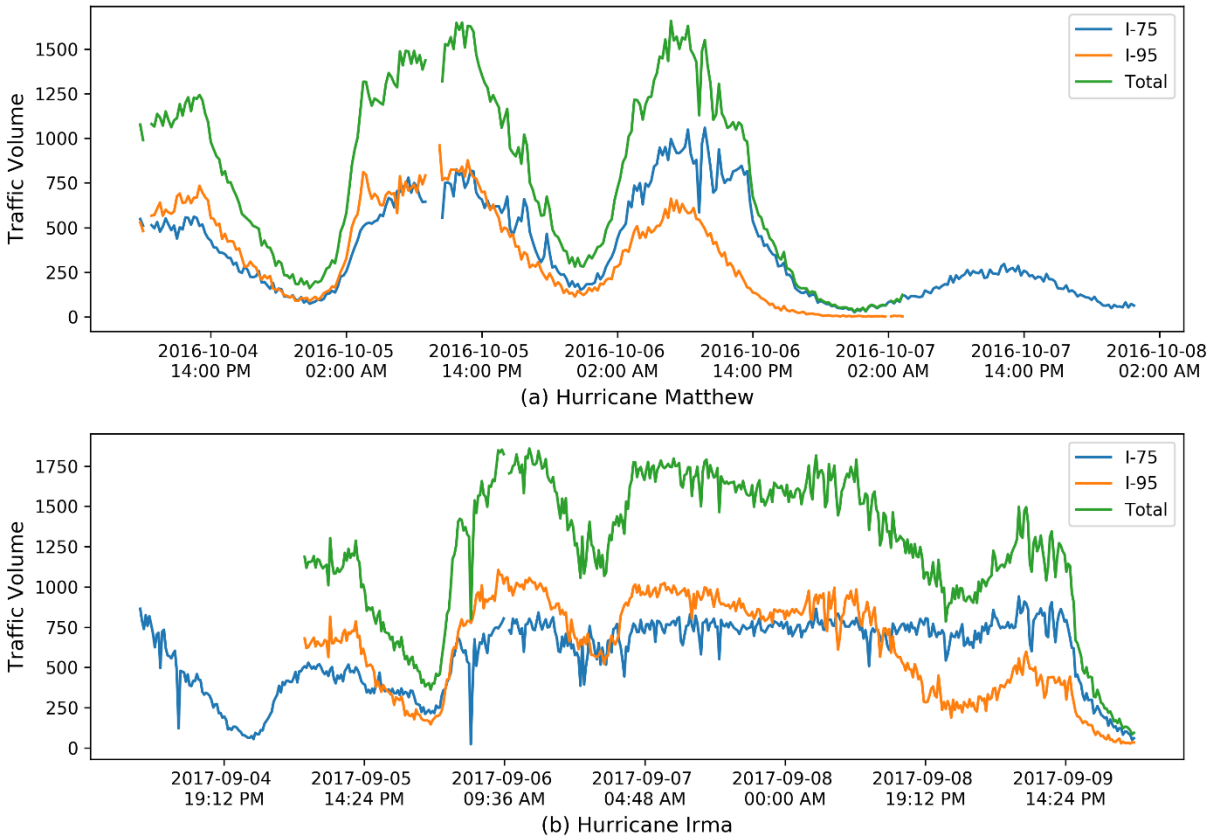


Figure 3.2 Traffic Volume for 15 Minutes Intervals at Interstate Highways during Hurricane Evacuation (a) Hurricane Matthew (b) Hurricane Irma

3.4 Data Preparation

Interstate highways I-75 and I-95 are the most popular routes during evacuations from Florida. We take the sum of the traffic volume of I-75 and I-95 to capture the overall traffic demand during evacuation from the associated regions. Figure 3.2 shows the traffic volume generated during hurricanes Matthew and Irma. During hurricane Matthew, I-95 traffic was higher than I-75 traffic because Matthew was expected to hit on the east coast. On the other hand, during hurricane Irma, at first traffic on I-95 was higher than I-75; but later (after September 8, 2017) it was the

opposite. This is reasonable as the projected path of hurricane Irma changed overnight on September 8, 2017. Initially Irma was expected to hit from the east coast, but later it changed its path and was predicted to hit from the west coast. Hurricane Matthew Twitter data are filtered for geotagged tweets within the study region bounded by the coordinates (25.072, -82.963; 29.352, -79.232). Similarly, hurricane Irma data are also filtered by the tweets coming from our study area. We also filter both data sets by evacuation related tweets having words such as 'evacuation', 'evac', 'sheltering', 'evacuating', 'evacuate' etc. We aggregated the tweets based on 15 minutes interval to be consistent with the traffic volume data. The traffic data have some gaps; since the missing data cover for a very small period, we linearly interpolate the missing data.

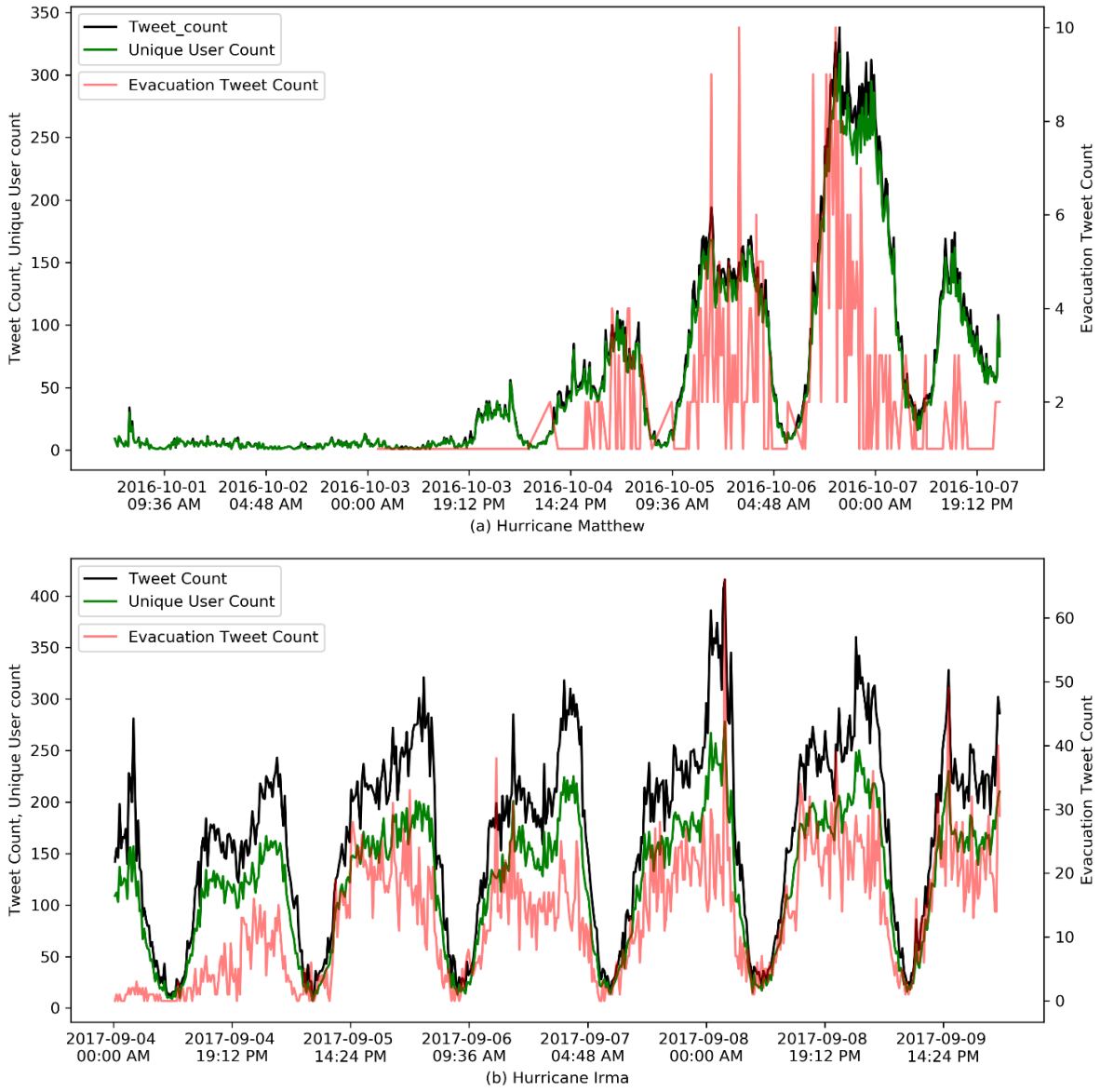


Figure 3.3 Twitter Features (a) for hurricane Matthew (b) for hurricane Irma

Hurricane Irma Twitter data have also some missing data, which we have recovered by collecting historical data, using REST API, for the active users found in the streaming data during hurricane evacuation. We standardize the data before fitting the model. Figure 3.3 shows the created features: tweet count, unique user count, evacuation tweet count at 15 minutes interval for

both hurricanes Matthew (Figure 3.3a) and Irma (Figure 3.3b). We have also created the following features: time difference (in hours) from landfall, hour of the day, number of counties ordering mandatory evacuation, number of counties ordering voluntary evacuation, total number of populations under voluntary order, total number of populations under mandatory order. We have used 2018 population data collected from https://www.florida-demographics.com/counties_by_population. We collect the issuance time of an evacuation order for a county from the official emergency management Twitter accounts of the corresponding county. Note that hurricane Matthew made its landfall on October 8, 2016 and hurricane Irma made landfall on September 10, 2017. In total, we retrieve 716 hourly observations, 263 are from Matthew and 453 are from Irma.

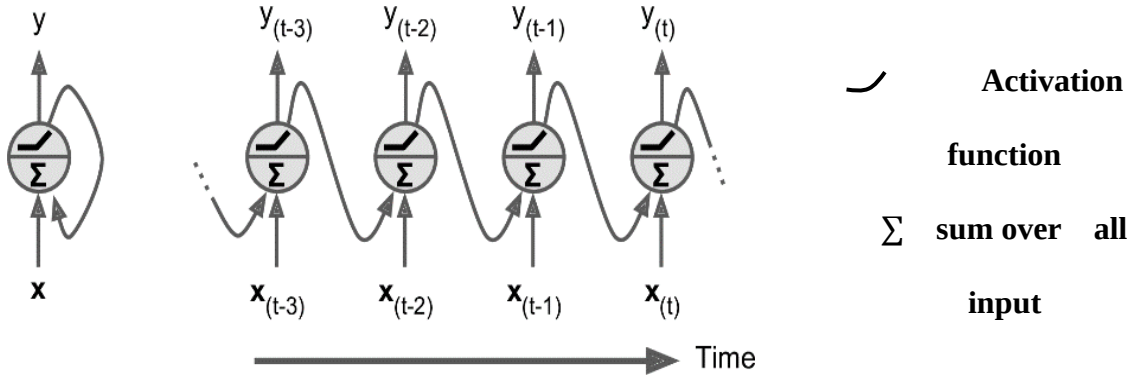
3.5 Modeling Approach

To predict traffic demand during a hurricane evacuation, we have developed a neural network approach. In particular, we use a long short-term memory neural network (LSTM-NN) architecture which is a special type of recurrent neural network (RNN). In machine learning, recurrent neural networks (RNN) are used for learning sequential trends. It has been used to solve many problems such as speech recognition [160], language modeling, image captioning etc. Unlike traditional neural networks, a recurrent neural network has loops in them (Figure 3.4(a) left) which allow to pass message to a successor. Figure 3.4(a) (right) shows a one neuron RNN unrolled over time. This chain-like nature reveals its potential to learn sequence both from current inputs and previous relevant information.

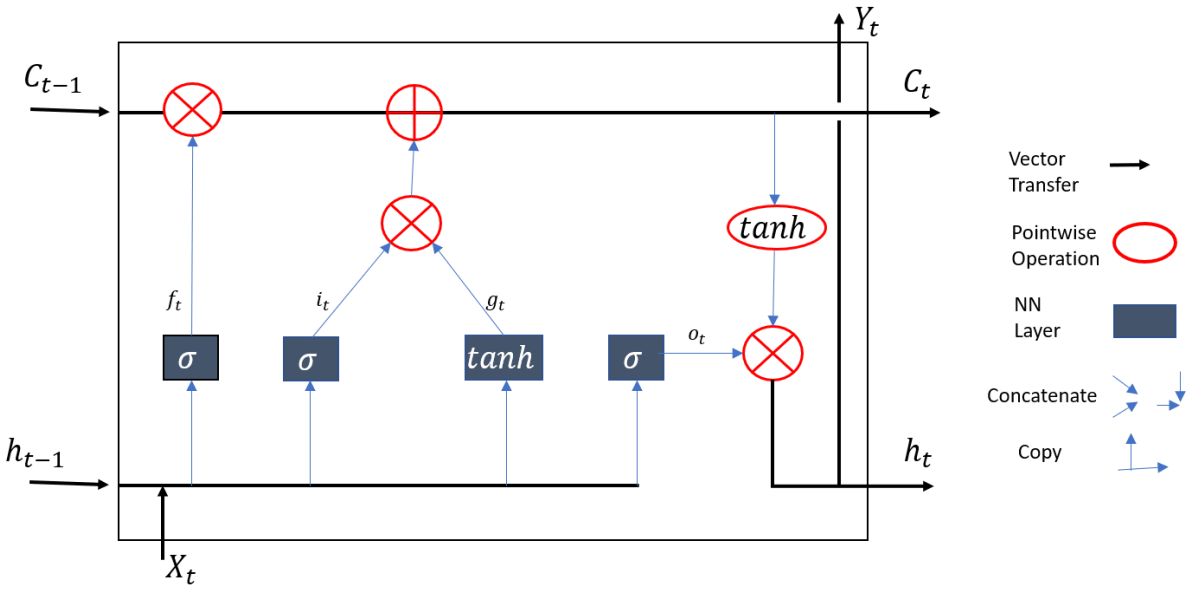
Although standard RNN performs well in general time series forecasting, it performs less in learning long-term dependencies due to vanishing/exploding gradient problem during

backpropagation [161–163]. LSTM, introduced by Hochreiter and Schmidhuber, resolves this problem by remembering information for long period of time [162]. Like RNN, LSTM also has the form of a chain of repeating modules of neural network. Unlike RNN's simple (e.g., a simple tanh layer) module, LSTM has four layers interacting in a very special way.

Figure 3.4(b) shows an LSTM cell with different components in it. Here σ represents a sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$ and $tanh$ represents a hyperbolic tangent function $\tanh(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$. The key difference between LSTM and RNN is the cell state (C_t) shown as a horizontal line in Figure 3.4(b). It runs through the entire chain with some minor linear interaction. Thus, it helps to keep track of long-term dependencies. It is also known as long-term state. LSTM allows to add or remove information to the cell state by some structures called gates. Gates are



(a) RNN architecture. Adopted from [164]



(b) LSTM cell

Figure 3.4 Architecture of RNN (a) a single neuron RNN unrolled trough time (b) a standard LSTM cell

composed of a sigmoid neural net and a pointwise multiplication operation (see Figure 3.4(b)). An LSTM has three such gates: forget gate, input gate and output gate. Sigmoid layer gives output numbers between zero to one where zero means nothing and one means everything. LSTM also uses the previous short-term state (h_{t-1}) and current input (X_t) and feed this into the above discussed layers. The first step is to decide what information to forget. For this, the forget gate (f_t) takes h_{t-1} and X_t and outputs numbers between 0 to 1 for each element in the cell state. Mathematically, the operation is shown below in equation (3.1).

$$f_t = \sigma(\mathrm{W_f}.[\mathrm{h_{t-1}}, x_t] + b_f) \tag{3.1}$$

where W_f , b_f are the weight matrices and bias for the corresponding forget gate neural network. The next step decides what new information to store in the cell states. It is performed in two parts: an input gate layer (i_t) that decides which values to update through its sigmoid layer

and a $tanh$ layer that converts the values into a vector (g_t) by its activation function. These two operations are shown below in equations (3.2) and (3.3).

$$i_t = \sigma(\mathrm{W}_i.[\mathrm{h}_{\mathrm{t}-1}, x_t] + b_i) \tag{3.2}$$

where, W_i , b_i are the weight matrices and bias for the corresponding input gate neural network.

$$g_t = \tanh(\mathrm{W}_g.[\mathrm{h}_{\mathrm{t}-1}, x_t] + b_g) \tag{3.3}$$

where, W_g , b_g are the weight matrices and bias for the corresponding $tanh$ layer.

The next step is to update the old cell state (C_{t-1}) into new state (C_t). The new cell state will be the combined result after forget gate and input gate operations. Equation (3.4) shows the updated cell state:

$$C_t = \mathrm{f}_\mathrm{t} * C_{t-1} + i_t * g_t \tag{3.4}$$

The last step is to predict the outputs from the current LSTM. The output is a filtered version of the current cell state (C_t). The previous state (h_{t-1}) and input (X_t) go through a sigmoid layer and the cell state go through a $tanh$ layer (to push the values to be between −1 and 1). Then multiplication of this two gives the output which is the decided part of the cell state. The operations are shown in equations (3.5) and (3.6):

$$o_t = \sigma(\mathrm{W}_o.[\mathrm{h}_{\mathrm{t}-1}, x_t] + b_o) \tag{3.5}$$

where, $W_o \; and \; b_o$ are the weight matrices and bias for the corresponding sigmoid layer.

$$h_t = \mathrm{o}_\mathrm{t} * \tanh\,(C_t) = Y_t \tag{3.6}$$

where, Y_t is the output at time t.

3.6 Model Development

The objective of this study is to predict hurricane traffic volume during evacuation for longer time horizon. We define the prediction problem as: *given the traffic or Twitter data or both* (X_t) *at time* t*, what is the traffic volume after* h *time intervals* $Y_{(t+h)}$*, where* h *represents the forecast horizon.*

We have used traffic sensor data and Twitter data as inputs to the proposed LSTM-NN model. We use the LSTM model because of its well-known performance in time series prediction. Previous studies found that LSTM model [151] or hybrid or fusion of LSTM model [165–168] outperformed other machine learning models in traffic state prediction. Moreover, LSTM provides more flexibilities (with respect to the number of parameters and regularization) than other models and the training of a model (underfit or overfit) makes a difference in its performance [169]. Studies suggest that with appropriate training mechanisms, a deep learning model may be trained with 100 – 1000 samples [169–171]. In this study, we have selected the epoch size (number of complete – both forward and backward – passes) depending on the forecast horizon and features so that the model is trained optimally. Moreover, we have used dropout as a regularizer to prevent overfitting.

The LSTM-NN input data (*X*) needs to be provided with specific dimensions of array where dimension of *X* indicates [samples, time steps, features]. Our features are multivariate as we are using 10 features (traffic volume, time difference from landfall, hour of the day, tweet count in the study area, evacuation related tweet count in the study area, unique user count, number of counties ordered mandatory evacuation, number of counties ordered voluntary evacuation, number of people under voluntary evacuation, number of people under mandatory evacuation) for a single time period. However, these features are used in different combinations—only sensor data, only

Twitter data, combination of both—to test how the model performs under different conditions of data availability. The time step dimension indicates how many time instances we are using to predict the output. For example, $[X_{t-1}, X_t]$ can be used to make prediction of Y_{t+h}.

In our experiments, we find that a single layer LSTM with 50 neurons for all the models performs reasonably well. Batch size is a parameter which represents the number of training example to be considered in one forward or backward pass. Studies show that larger batch size degrades the quality of the model [172]. We find that batch size = 4 performs well on our data for all the forecast horizon. To determine the time horizons for which the model can perform well, we iteratively run the model for time horizons from 1 hour to 30 hours. For each time horizon, we run the model 5 times with different initializations. Along the run, we find the best epoch size (one forward and one backward pass on all training samples) to ensure that the model does not overfit.

Table 3.1 presents the summary of the estimated parameters in our study. We implemented all the models in Python programing language. Unless otherwise specified in Table 3.1, we have used the default parameters of Keras [173] for LSTM and the default parameters of Scikit-learn [174] for the other models.

We compare prediction accuracy of the proposed LSTM-NN model with traditional machine learning algorithms such as *K*-nearest neighbor regression (KNN regression), support vector regression (SVR), gradient boosting regression (GBR), and XGBoost regression (XGBR) models. We iteratively select the best parameters for these algorithms using a grid search approach [174]. Generally, for the KNN algorithm, a large number of neighbors underfits the model and a small number of neighbors overfits it. SVR tends to overfit with the increase in polynomial degree. For GBR and XGBR, model complexity increases (or overfit) with the increase of parameter value of max depth and the number of estimators. More details of these parameters and implementation

can be found here [164, 173, 174]. We report the average performance over 10-fold cross validation trials.

Table 3.1 Summary of the model parameters

Model	Parameter setup (range of parameter values tried to find the best performance)	Summary of the best parameters for forecast horizon 1 to 30 (min, max, avg.) for numeric values {Frequency Distribution} for other type of parameters			
		Sensor	Twitter	Combined	Important
KNN	Number of Neighbors (1, 15)	(3, 14, 12.5)	(1, 14, 11.33)	(2, 14, 11.66)	(2, 14, 10.06)
	p (1, 2)	{1: 20, 2: 10}	{1: 26 2: 4}	{2: 17, 1: 13}	{1: 19, 2: 11}
SVR	C (1, 1000)	(201, 901, 757.66)	501 901 854.33	101 901 544.33	101 901 627.66
	Degree (1, 4)	{3: 13, 2:10, 1: 7}	{2: 17, 3: 13}	{1: 16, 2: 9, 3:5}	{1: 12, 2: 11, 3: 7}
GBR	Max Depth (2, 10)	(2, 9, 6.33)	(2, 9, 5.53)	(2, 9, 6.46)	(2, 9, 6.1)
	Number of Estimator (5, 15)	(7, 14, 12.7)	(9, 14, 13.06)	(9, 14, 13.43)	(8, 14, 13.43)
	Sub Sample (.1, 1)	(0.2, 1.0, 0.54	(0.2, 0.8, 0.34)	(0.2, 1.0, 0.47)	(0.2, 1.0, 0.50)
XGBR	Learning Rate (0.03, 0.08)	(0.03, 0.07, 0.06)	(0.03, 0.07, 0.05)	(0.03, 0.07, 0.05)	(0.03, 0.07, 0.05)
	Max Depth (5, 8)	{5: 19, 6: 6, 7: 5}	{5: 18, 7: 7, 6: 5}	{6: 11, 5: 11, 7: 8}	{5: 14, 7: 9, 6: 7}
	Number of Estimator (5, 500)	(50, 500, 260)	(30, 500, 183.66)	(50, 500, 455)	(50, 500, 380)
LSTM	Batch Size	4			
	Epoch Size (1, 3000)	(121, 2288, 952.17)	(4, 2993, 930.51)	(2, 2354, 347.52)	(5, 2999, 720.27)
	Number of LSTM cells	50			
	Dropout	0.50			
	Optimizer	‘adam’			
	Learning Rate	0.0001			

To evaluate model performance, we have also created a baseline forecast. In this baseline forecast, the traffic volume in the next time interval is simply predicted as equal to the current

traffic volume. With forecast horizon h, at a given time (t), if the current traffic volume is Y_t, traffic prediction for $(t+h)$ is equal to Y_t (i.e., $Y_{t+h} = Y_t$).

Unlike other regression problems, in time series forecasting, the order of the observation is important to learn the sequence. Thus, keeping the order of the sequence, we have used 80% of the data as training set and the rest as validation dataset. To evaluate the performance of the implemented models, we have used Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) as performance measures. We choose the best model considering the performance over all forecast horizons (i.e., the model that shows overall stable performance). The equations for performance measures are given below:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_{t+h} - \hat{y}_{t+h})^2}{n}} \tag{3.7}$$

$$\text{MAPE} = \sum_{t=1}^{n} \frac{|Y_{t+h} - \hat{y}_{t+h}|}{Y_{t+h}} \times 100\% \tag{3.8}$$

where, $Y_{t+h}$is the actual traffic volume and $\hat{y}_{t+h}$ is the predicted traffic volume for forecast horizon h and n is the number of test observations.

Next, we implement an approach to estimate the confidence interval of the predicted traffic volume. We assume that for a forecast horizon h, the predicted traffic volume $(\hat{y}_{t+h})$ follows a normal distribution, where the parameters, mean (μ_{t+h}) and standard deviation(σ_{t+h}), depend on the input variables(X) at time t. Here,

$$\mu_{t+h} = f(X_t) = predicted\ traffic\ volume\ (\hat{y}_{t+h})\ by\ the\ best\ model \tag{3.9}$$

To compute the standard deviation, we estimate separate models where the input is the same as the model for traffic volume prediction, and the output is the absolute error ($|Y_{t+h} - \hat{y}_{t+h}|$) for the estimated best model for traffic volume prediction.

$$\sigma_{t+h} = predicted\ standard\ deviation\ by\ the\ best\ error\ prediction\ model \quad (3.10)$$

Finally, we compute the confidence interval at 90% confidence level by the following:

$$Lower\ bound\ of\ \hat{y}_{t+h} = \mu_{t+h} - 1.65 * \sigma_{t+h}$$

$$Upper\ bound\ of\ \hat{y}_{t+h} = \mu_{t+h} + 1.65 * \sigma_{t+h}.$$

3.7 Results

We implement the LSTM-NN model for different forecasting horizons ranging from 1 hour to 30 hours. Also, we run the model separately considering 4 scenarios where only traffic sensor data, only Twitter data, combination of both, and only the top 4 important features are used. We consider two features (time difference from landfall and hour of the day) available for all four scenarios.

We calculate the feature importance using permutation importance [175]. We calculate the importance of a feature based on RMSE score. For each feature column, we shuffle the corresponding feature and compute the importance by checking how much RMSE has increased. Feature importance values for all the available features are shown in Figure 3.5 for different forecast horizons. It shows that for small time horizon (1 to 5 hours), traffic volume has the highest importance. As the forecast horizon increases (7 to 30 hours), importance value of time difference from landfall feature increases. This implies that in predicting traffic during evacuation for longer forecasting horizon time difference from the forecasted landfall time plays a very critical role. Interestingly, Twitter features have almost no importance for forecast horizon between (1-10 hours), but the importance value increases from forecast horizon 11-15 hours. This matches the

intuition that people tweet well before the actual evacuation. For example, these tweets– *"Preparing to evacuate knowing full well that I could come back to nothing is kinda terrifying. #HurricaneIrma", "Bags packed ready to evacuate if needed #HuracanIrma"*–indicate user intent to evacuate before their actual evacuation. In addition, we are considering Twitter activities in the entire study region; but traffic is measured on two specific points (see Figure 3.1). It takes time to travel to the exit point from the other points within the study area. Nonetheless, it means that the Twitter features are very important in predicting traffic volume around 11-15 hours ahead.

The test RMSE and test MAPE for of the models are shown in Figure 3.6. We run the models for different forecast horizons with different combination of features. Except for the Twitter only features, in all the combinations, 1-hour forecast horizon has the lowest RMSE and MAPE values. This is expected since we have the most recent information in this case for our prediction purpose. As the forecast horizon increases, RMSE and MAPE increase with some exceptions.

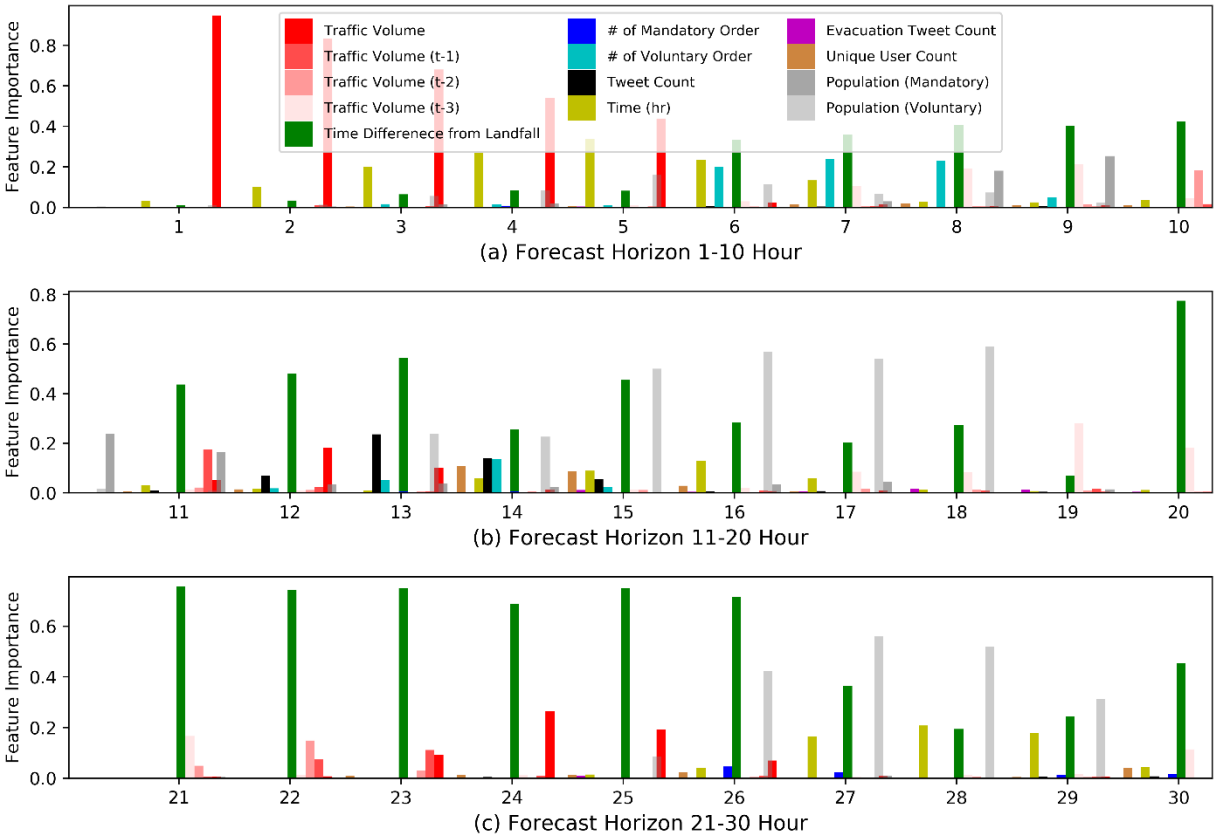
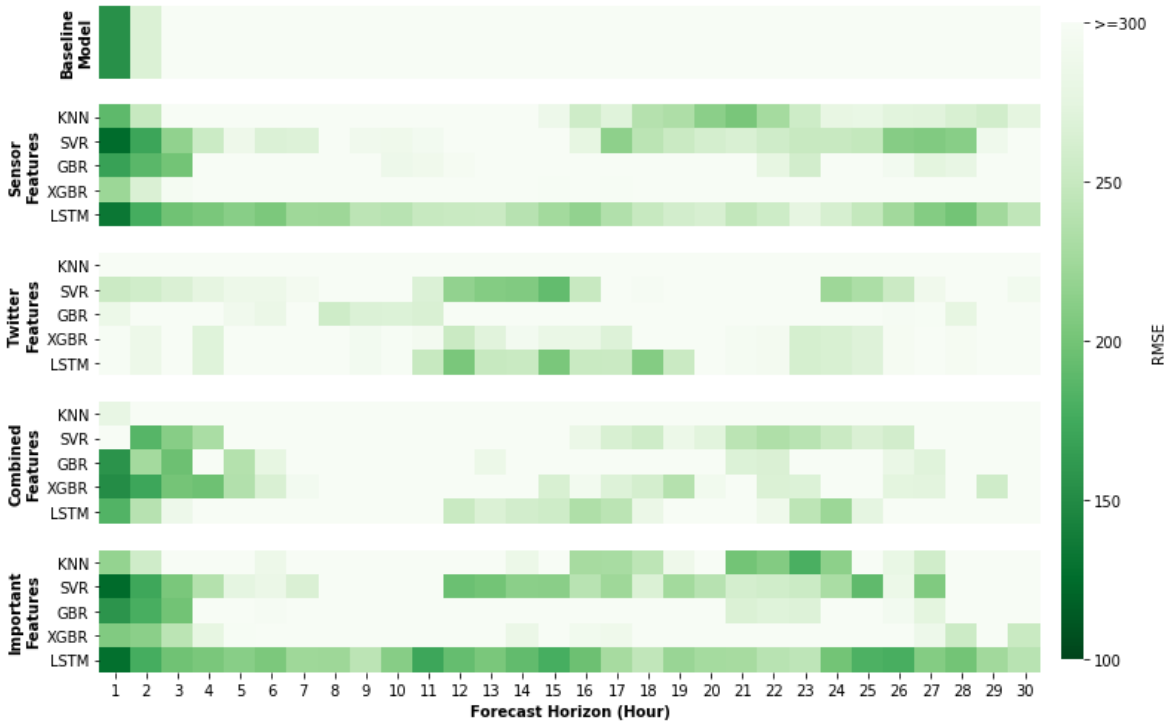
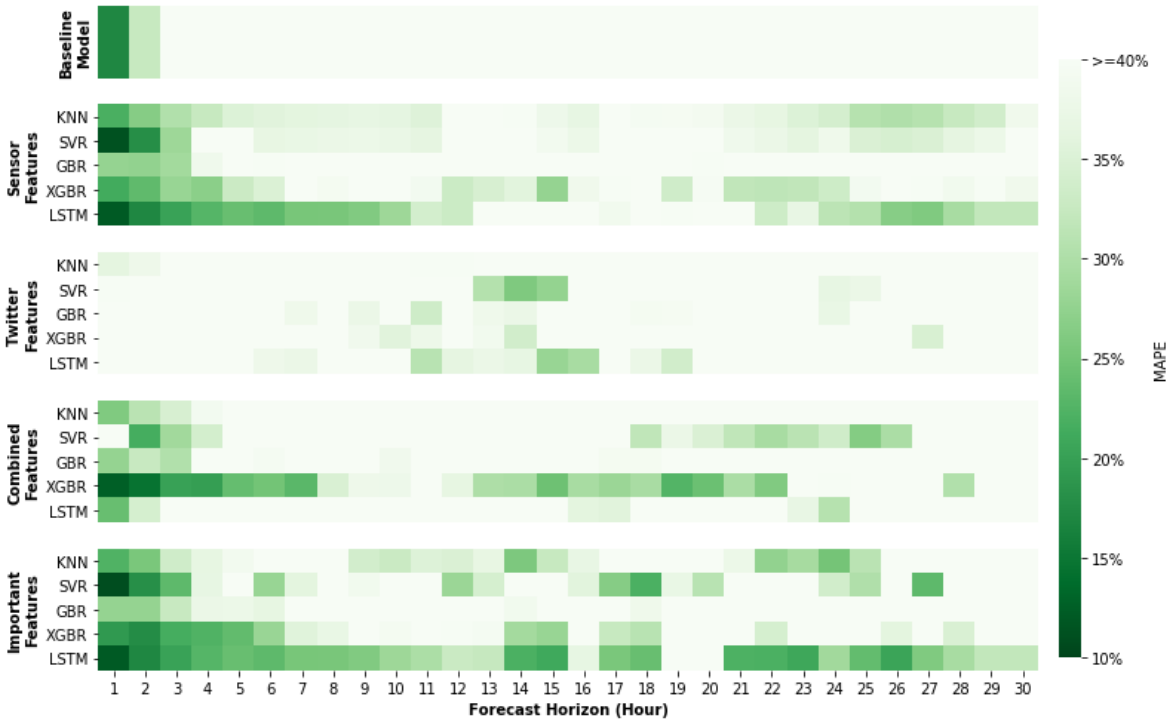


Figure 3.5 Feature Importance for Different Time Horizons


Baseline Model
Sensor Features
Twitter Features
Combined Features
Important Features
KNN
SVR
GBR
XGBR
LSTM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Forecast Horizon (Hour)
>=300
250
200
150
100
RMSE

(a)


Baseline Model
Sensor Features
Twitter Features
Combined Features
Important Features
KNN
SVR
GBR
XGBR
LSTM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Forecast Horizon (Hour)
>=40%
35%
30%
25%
20%
15%
10%
MAPE

(b)

Figure 3.6 Validation (a) RMSE (b) MAPE for different Forecast Horizons on Test Data. X-axis represents the forecast horizons, Y-axis shows the model names. The color within a cell represents the model performance in terms of RMSE in (a) and MAPE in (b). The greener the color the better is the performance.

For forecast horizons 1 hour to 12 hours, the model trained with only sensor data has performed better than the models trained with only Twitter features or combined features. This is consistent with the results related to feature importance where we found that traffic volume has higher importance for shorter forecast horizons. Models trained with only Twitter features perform well for forecast horizons 10 hours to 19 hours (see Figure 3.6), which is also consistent with the feature importance analysis. This is probably due to the fact that people post about their hurricane awareness or evacuation intent prior to the actual action. Also, the distances between the sensor locations and the location of Twitter users are not same for all the areas within the study region. Thus, it may take some time to realize the traffic impacts of those users stating evacuation intent in Twitter. The result indicates that, when traffic sensor data is not available, Twitter data can be used to predict traffic demand during evacuation from 10 hours to 20 hours forecast horizons. However, models trained on combined features, containing all the available features, do not perform well (see Figure 3.6). Adding unnecessary features degrades model performance in this case. For all the models, performances are better for the important features among the four (sensor, Twitter, combined, important features) feature types. Using only important features, models are performing consistently better than the sensor features for forecast horizons 11 hours to 23 hours.

The performances of all models are compared against a baseline forecast. For 1-hour and 2-hour forecasting horizons, all models trained with only Twitter data failed to outperform the baseline results and for the other feature combinations only the LSTM models and SVR models

outperform baseline forecast. We can see that overall LSTM-NN models perform better than the baseline and other models for all feature types (see Figure 3.6). The performances of the LSTM-NN models are more consistent across all forecast horizons compared to the other models. This shows the advantage of an LSTM-NN modeling framework to capture both short-term and long-term dependencies in predicting traffic during evacuation.

The LSTM-NN model performs best (RMSE=110, MAPE=13%) for 1-hour forecast horizon when trained with important features or only sensor features. LSTM-NN model trained with only Twitter data has the best result (RMSE=203, MAPE= 28%) for 15-hour forecasting horizon, which is better compared to the performance found for the models trained with combined features and only sensor features for the same forecast horizon. This indicates that when traffic sensor data are unavailable, Twitter data can be used to obtain reasonable prediction on future evacuation demand. Using top 4 important features (adding Twitter features with the sensor data) lowers the RMSE value to 160 and MAPE value to 25% for a 15-hour forecast horizon for the LSTM-NN model.

To further evaluate the prediction performance and the robustness of the models across hurricanes, we run two types of experiments. In the first type of experiment (Figure 3.7), we train models for different forecast horizons using full hurricane Matthew and part of Irma as training data and test the models using the remaining part of Irma data. In the second type of experiment (Figure 3.8), we train models for different forecast horizons using part of hurricane Matthew and part of Irma data as training data and the remaining parts of Matthew and Irma data as test data. In all these experiments, we use the LSTM models trained on important features only, derived previously. We report here only the results for 1-hour and 24-hour forecast horizons. In Figures 3.7 and 3.8, (a) and (b) represent Matthew and Irma data and (1) and (2) represent 1-hour and 24-

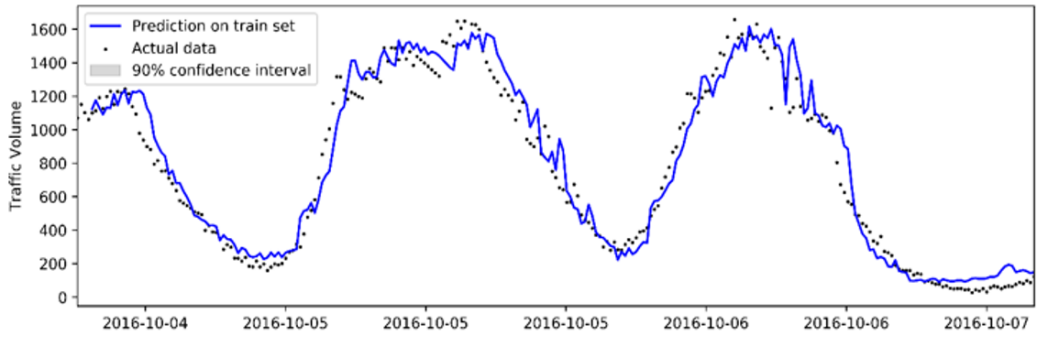
hour predictions, respectively. For a 24-hour forecast horizon, there is no prediction on the first 24 hours of data, because training data are not available for 24 hours before a given period.

Figures 3.77a.1 and 3.7b.1 together show the results of the LSTM model that has been trained over full hurricane Matthew data and a portion of hurricane Irma data and tested over the rest of the hurricane Irma data for 1-hour forecast horizon. Similarly, Figures 3.7a.2 and 3.7b.2 together show the results of the LSTM model that has been trained over full hurricane Matthew data and a portion of hurricane Irma data and tested over the rest of the hurricane Irma data for 24-hour forecast horizon.

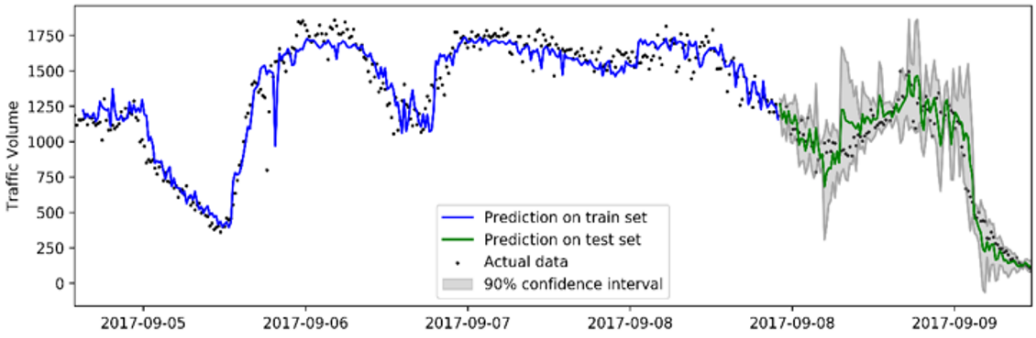
Figures 3.8a.1 and 3.8b.1 together show the results of the LSTM model that has been trained using part of hurricane Matthew and part of Irma data as training data and tested over the remaining parts of Matthew and Irma data for 1-hour forecast horizon. Similarly, Figures 3.8a.2 and 8b.2 together show the results of LSTM model that has been trained over a part of hurricane Matthew and part of Irma data and tested over the remaining parts of Matthew and Irma data for 24-hour forecast horizon.

Prediction on training data fits well for both hurricanes Matthew and Irma for 1-hour and 24-hour forecast horizons. On the other hand, prediction on the test data show that, prediction for 1-hour forecast horizon fits better than the prediction for 24-hour horizon, capturing the trend well enough. We also find that the model is predicting better for hurricane Irma test data than hurricane Matthew test data. This is because we have more training data available for hurricane Irma than hurricane Matthew. As such the results can be further improved by recording the trends of traffic volume over time for multiple hurricanes. Although prediction accuracy decreases with longer forecast horizon, the implemented model learns the overall trend (increasing or decreasing evacuation traffic) well enough.

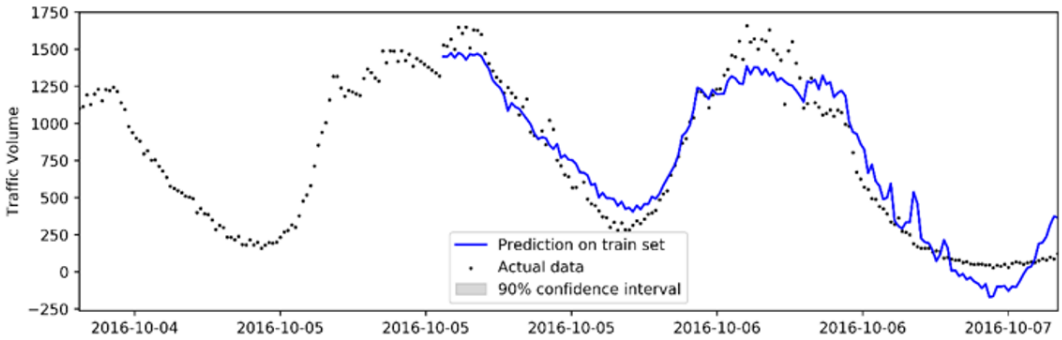
We have also shown the 90% confidence interval of the prediction on test set for 1-hour and 24-hour forecast horizons. We found that *k*-nearest neighbors (neighbors =3) perform best in predicting the absolute error. Predicted value by the best model (LSTM-NN) is always in between the predicted confidence interval. Moreover, the interval is greater when the demand prediction error is greater, and the interval is almost equal to zero when the LSTM-NN model make perfect traffic demand prediction (see Figure 3.7 and 3.8). Thus, the confidence interval prediction is working well in capturing the uncertainty in prediction by the LSTM-NN model. Evacuation demand prediction with its associated confidence interval will help interpret the prediction results more reliably.



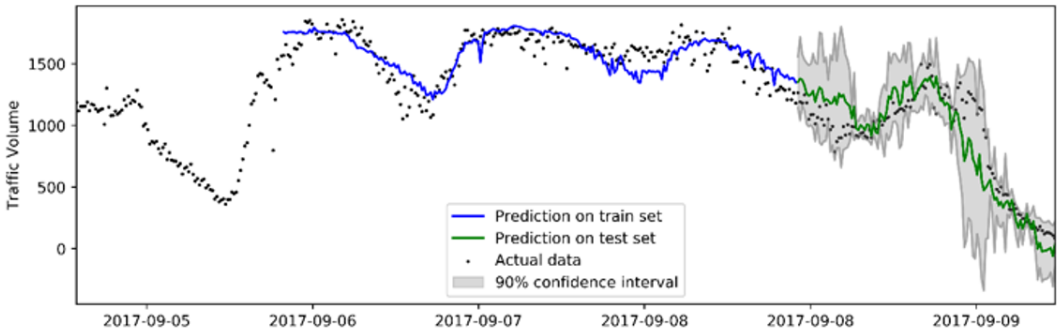
(a.1) 1-hour prediction (Matthew)



(b.1) 1-hour prediction (Irma)

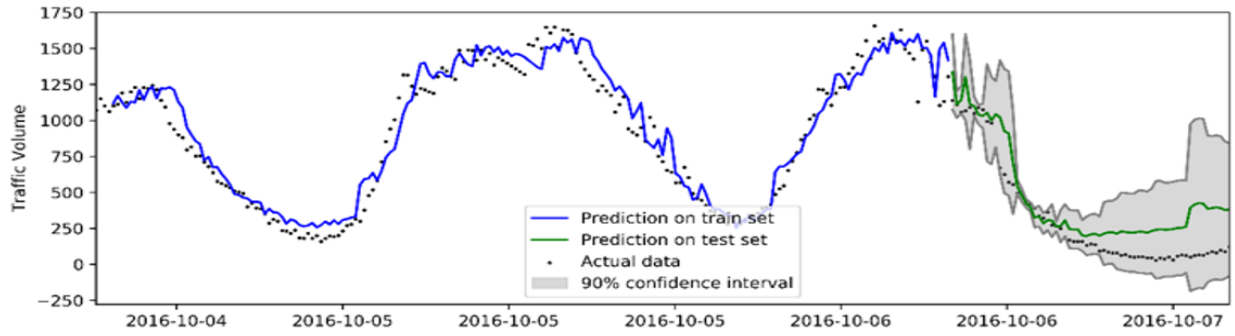


(a.2) 24-hour prediction (Matthew)

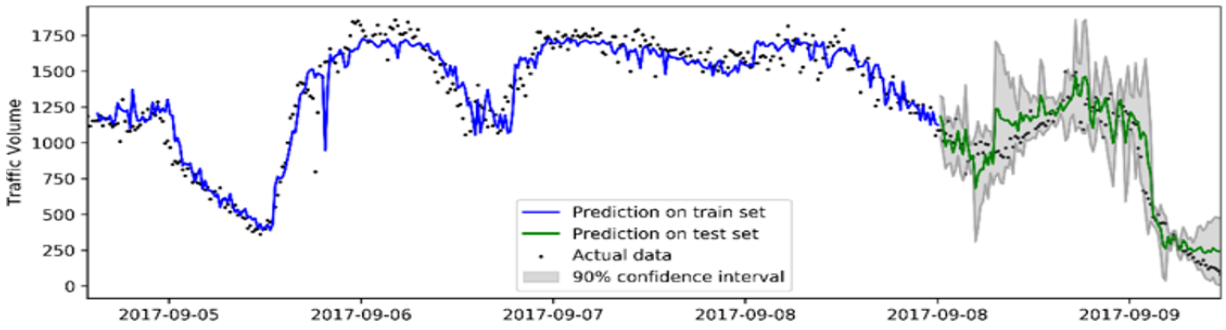


(b.2) 24-hour prediction (Irma)

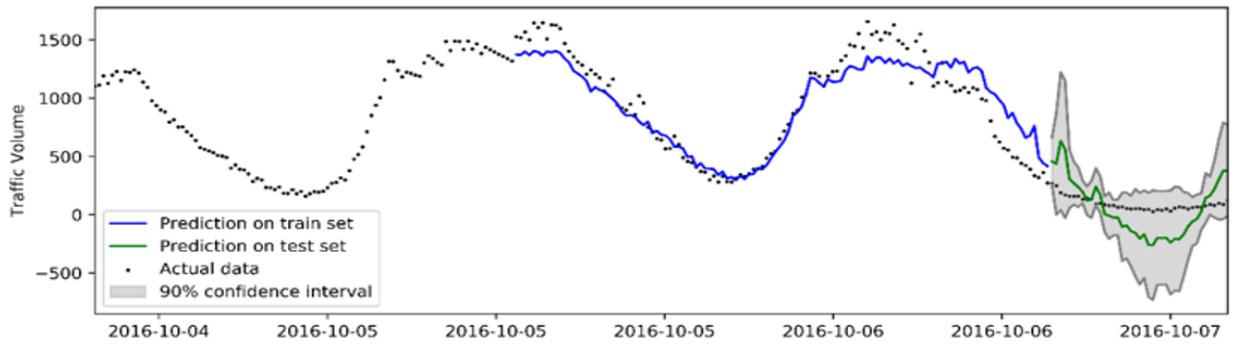
Figure 3.7 Prediction with 90% confidence interval on training and test data when test data contains only hurricane Irma data. Here (a.1) and (b.1) show 1-hr forecast for Matthew and Irma, respectively and (a.2) and (b.2) show 24-hour prediction for Matthew and Irma, respectively.



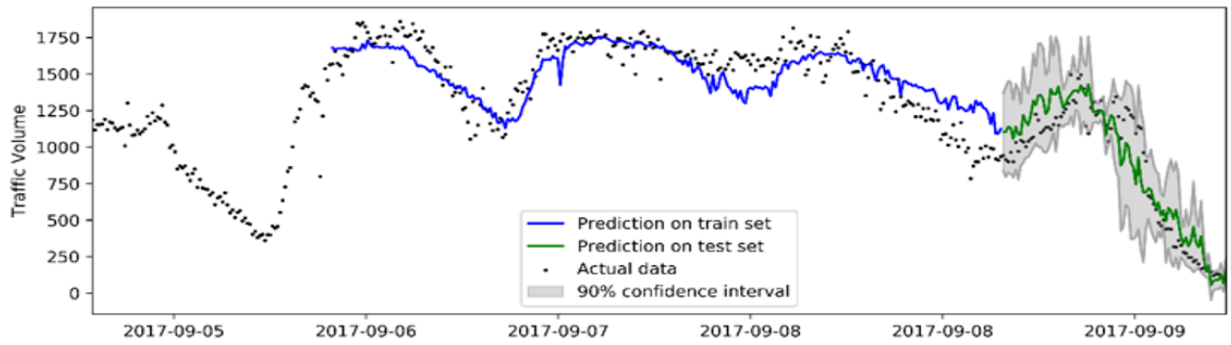
(a.1) 1-hour prediction (Matthew)



(b.1) 1-hour prediction (Irma)



(a.2) 24-hour prediction (Matthew)



(b.2) 24-hour prediction (Irma)

Figure 3.8 Prediction with 90% Confidence Interval on training and test data when test data contains both Matthew and Irma data. Here (a.1) and (b.1) show 1-hr forecast for Matthew and Irma respectively, and (a.2) and (b.2) show 24-hour prediction for Matthew and Irma respectively.

3.8 Limitations and Future Research Directions

Our study has some limitations. We have used traffic demand collected from traffic detectors and there are detectors at only two highways (I75 and I95) at the downstream boundary of the study area. We have simplified the problem by adding the traffic from I-75 and I-95 to determine the total traffic demand during evacuation. Although most evacuees during evacuations use one of these two highways, this assumption may not hold in some areas. However, our approach can be generalized for any number of highways (any size of study area) given the availability of the data. In addition, traffic sensor data suffer from missing information. Machine learning techniques [176] can be used to fill the information gaps in traffic sensors. We have used evacuation related tweets based on the presence of certain pre-selected keywords. Natural language processing models [177] can be developed to infer evacuation intent from social media posts. Furthermore, Twitter data suffer from demographic biases; population from certain areas may post more evacuation related tweets compared to other areas. Such biases should be corrected to rigorously predict evacuation traffic from tweets. Thus, further research addressing sensor selection and bias correction for Twitter data may improve our prediction accuracy in future.

Every hurricane is different from each other in many aspects, such as severity, hurricane path, and intensity. Thus, the generated evacuation traffic can be different from one hurricane to other. For example, evacuation traffic during hurricane Matthew and Irma shows different pattern (see Figure 3.2). Our approach presented in this study capture these two different patterns by

adopting some real-time features (sensors, social media posts etc.). Considering the unpredictable behavior of the hurricanes, hurricane specific model might perform better in predicting traffic during hurricane. While conducting our study, we find that many traffic sensors suffer disruptions during hurricane that makes it difficult to collect enough traffic sample to train separate models for each hurricane. Adopting domain adaption or transfer learning approach [178] to train models on historical hurricane data set and calibrate the model as new hurricane data is available should be explored in future research.

We have summed up the volumes of two major highways at the downstream cut-off points to get the total traffic volume at any given time irrespective of destination. Thus, evacuation destination choice (outside the study area) is less likely to have any effect on the traffic demand generated from the region. Our approach has missed the internal evacuation within the study area. However, because hurricane Irma was projected to affect a wide area (from East coast to West coast of Florida), such internal evacuation would be limited. Future study may adopt a spatially aware [179] deep learning technique considering relative location of the tweets and traffic sensors. A hybrid approach by combining a location aware Convolutional Neural Network [179] with an LSTM model (to capture the temporal effect) might be explored in future studies.

3.9 Conclusions

Traditional approaches for predicting hurricane evacuation demand use survey-based data and they work well upon a fixed set of assumptions, which may not be suitable for real-time traffic prediction. Information from real time data sources can make evacuation traffic management more dynamic, flexible, and proactive. In this study, we have used traffic sensor and Twitter data to predict traffic demand during evacuation for longer term forecasting horizons. We have applied a

machine learning model known as LSTM neural networks to predict traffic demand during evacuation for different forecast horizons ranging from 1 hour to 30 hours. We have applied the model for different combination of features (only traffic sensor data, only Twitter data, both sensor and Twitter data, only important features). Among the modeling approaches, LSTM-NN outperforms other models in terms of accuracy. Social media features show its best predictive power for 15 hours forecast horizon. Model trained on social media data can help make reasonable predictions of traffic during evacuation when sensor data are not available. We also implement a method to predict the confidence interval of the demand prediction made by the model. These approaches allow us to measure the reliability of the predicted traffic demand during evacuation.

With increasing population and the number of hurricanes in the coastal regions, efficient and demand responsive evacuation traffic management is warranted. Our study integrates data from multiple sources which are readily available to predict traffic demand during hurricanes. While more studies are needed to predict evacuation traffic at a network-wide level, this study serves as a key step towards building a pro-active and demand responsive evacuation traffic management system.

CHAPTER 4: A MULTI-LABEL CLASSIFICATION APPROACH TO IDENTIFY HURRICANE-INDUCED INFRASTRUCTURE DISRUPTIONS USING SOCIAL MEDIA DATA

4.1 Introduction

[3]Cities and communities all over the world largely depend on critical infrastructure systems/services such as electrical power, water distribution, communication services, and transportation networks. The growing interconnectedness and interdependency among these systems have changed the organizational and operational factors and increased the vulnerability in the face of unwanted disruptions. These systems provide critical services to a large population, and thus when disrupted they affect our quality of life, local and regional economy, and the overall community well-being. The need to quickly identify disaster-induced infrastructure disruptions is growing because of the increasing number of natural disasters such as hurricane Michael, Irma, Harvey, and Florence and their enormous impacts to affected communities.

For instance, hurricane Irma caused a substantial number of power outages in addition to transportation, communication, drinking water, and wastewater related disruptions. More than six million customers faced power outages during Irma. Storm related high winds and sustained storm surges cost approximately 3,300 megawatts of power generation [48]. Around 27.4% of cell phone towers in Florida were damaged due to hurricane Irma as reported by the Federal Communications Commission (FCC) [180]. Irma caused flooding to several areas throughout Florida, forcing health

[3] Roy, KC, Hasan, S, Mozumder, P. A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. *Comput Aided Civ Inf*. 2020; 1– 16. https://doi.org/10.1111/mice.12573

officials to issue unsafe drinking water and boiling water notices [181, 182]. Moreover, dozens of sewage systems were overflowed after the power went out, which further exacerbated the drinking water condition [181].

To ensure efficient operation and maintenance, it is important to gather real-time information about the performance and integrity of engineering systems. This is typically performed through a computational monitoring process that involves observation of a system, analysis of the obtained data, and prediction of future performance [183]. During a disaster, due to disruptions, the performance of critical infrastructures degrades rapidly—leading to cascading failures [56, 184]. In such extreme events, computational monitoring is required to assess the quickly changing condition of infrastructure systems and warn about an approaching failure or even a catastrophic event.

Considering the severity and forecasted path of an approaching hurricane emergency officials often declare evacuation orders for to save human life. Such evacuation orders are expected to propagate through multiple sources (traditional media, social networks, social media, etc.) to inform people living in the risk zone. In response to the evacuation order and forecasted risk, households take evacuation decisions, which depends on a complex and dynamic process varying over time and household locations [8–10]. The delivery of accurate and timely information is crucial to create situational awareness in the affected communities. But the effectiveness of such information depends on the perceived credibility of the information sources and households' response to such information, which depend on the socio-economic characteristics of a household [55, 185, 186].

In this study, we develop a multi-label classification approach to identify the co-occurrence and extent of multiple types of infrastructure disruptions. We also present a framework to create

dynamic disruption maps and case studies showing the developed approach based on Twitter data collected during hurricanes Irma and Michael. This study has the following contributions:

- We consider multiple types of infrastructure disruptions (e.g., power, transportation, water, wastewater, and other disruption) and their co-occurrences in a social media post, instead of considering a simple binary classification problem (i.e., whether a post is disruption related or not).
- To identify if a disruption related post reflects an actual disruption, we associate sentiments with disruption status—whether a post is reporting an actual disruption (*negative*), or disruption in general (*neutral*), or not affected by a disruption (*positive*).
- We propose a dynamic mapping framework for visualizing infrastructure disruptions by adopting a geo-parsing method that extracts location from tweet texts.

Instead of identifying disruption types and status in a single label, we identify disruption types and disruption status (through sentiment) separately. We adopt this approach since the neutral and positive sentiment about a disruption may also provide valuable information on the level of situational awareness about disruptions during a disaster.

4.2 Literature Review

According to the Department of Homeland Security, there are 16 critical infrastructure sectors [187]. Among these sectors, energy, communication, transportation, water/wastewater systems are the most vulnerable ones to a natural disaster. It is important to identify, characterize, and model infrastructure disruptions for a faster restoration and recovery operation [188–190]. Studies have focused on the recovery plans and damages due to extreme weather events [189, 191–194]. Several studies have proposed approaches to assess the reliability, resilience, vulnerability and

failure process of power, transportation, and water supply networks individually [195–201]. However, these critical infrastructures are inter-connected and inter-dependent [187, 202, 203]. Considering the increased connectedness and interdependencies among infrastructure systems, studies have proposed a holistic approach to assess the resilience to disruptions [3, 188, 204–206]. However, most of these studies are based on synthetic data or post-event data. Thus, they are not suitable for real-time decision-making.

Recently, real-time condition monitoring is becoming very popular in manufacturing, maintenance, and usage of many engineering systems [183] and civil engineering infrastructures [207]. Computational models have been developed for estimating the properties of constructional materials [208], detecting damages to building structures [209, 210], predicting construction costs [211] etc. Another potential approach for monitoring infrastructures is by collecting real-time data using smartphones, leading to citizen-centered and scalable monitoring systems in a disaster context [212].

During an ongoing disaster and post-disaster period, it is important to collect disruption data to take necessary actions as fast as possible. Due to the intensity and spread of a disaster, physical sensing techniques such as satellite, UAV (Unmanned Aerial Vehicle) etc. [47, 48] are not suitable. For example, after hurricane Irma, unmanned aerial drones, amphibious vehicles, airboats are used to perform damage assessment on inaccessible transmission and distribution lines [48]. A crowd-sourcing app that allows damage reporting might not be useful because of fewer participants. On the other hand, the ubiquitous use of social media on GPS enabled smartphone device, allows us to collect large-scale user generated data containing live and in situ events during a disaster [29]. Studies have already used social media data for crisis mapping [27–29]. However, real-time crisis mapping requires location information, but only around 1% to 4%

of social media (e.g., Twitter) data posts are geo-tagged [29, 213, 214]. Studies have proposed several location-extraction methods from content/textual data [29, 213, 214]. In addition, the power of social media to connect a large group of population has drawn significant attention towards using social media platforms for disaster management [215–217]. Studies have analyzed social media data for understanding human mobility and resilience during a disaster [14, 19]. Kryvasheyeu et al. proposed that social media users can be considered as early warning sensors in detecting and locating disasters [16]. Studies have also explored social media data to understand evacuation behavior [20, 128] and damage assessment [18, 30–32].

Damage assessment plays a vital role in resource allocation and coordination in disaster response and recovery efforts. Previous studies found that affected people provide damage related situational updates in social media [18, 30–32]. However, these studies do not consider the types of disruptions and are mainly suitable for post-disaster overall damage assessment. Most of these studies adopted simpler indicators of damage assessment such as frequency of disaster related tweets (based on keywords such as 'sandy', 'hurricane sandy', 'damage'). The limitation of using pre-defined keywords is that a large number of such tweets/texts may not contain any damage related information. Some studies [218] adopted supervised machine learning based classification approaches to resolve this limitation. These studies [54, 219] adopted support vector machine, naïve Bayes, decision tree classification algorithms to analyze damage related social media posts. However, these studies considered damage identification as a binary (damage related or not) classification problem, which may include posts that are not reporting an actual damage/disruption. In addition, deep learning models were used for image and text data [220, 221]. Image data are limited, computationally expensive, and cannot report disruptions in functionality such as power outage, communication disruptions etc.

The most relevant studies towards identifying an infrastructure disruption using social media posts are proposed by Fan et al. [49, 52]. The first study [49] has focused on summarizing the overall topics during a disaster given some predefined keywords, not suitable to identify disruptions from real-time data. In the second study [52], the authors have developed a graph-based method to identify situational information related to infrastructure disruptions by detecting time slices based on a threshold number of tweets. They compute content similarity within the detected time slices to get credible information. Some limitations of this approach include: it depends on keyword based filtering, which can miss out important information if appropriate keywords are not chosen; it requires the whole dataset as an input, which is not suitable for a real-time prediction; it considers the content posted only on the burst timeframe that might miss out some actual disruption related posts. Moreover, this study does not consider that a single post may have information about multiple types of disruptions and cannot distinguish if a particular post is reporting an actual disruption or not.

In summary, to the best of our knowledge, currently no study exists to identify the co-occurrence of multiple types of infrastructure disruptions using social media data. For this task, a multi-label classification approach [222] identifying multiple labels from a single input, can be useful.

In this study, we use a multi-label text classification approach to identify multiple disruption types and their status using social media data. To develop our multi-label disruption classification approach, we use eight well-known models on text classification. We present two case studies to identify disruptions using Twitter data from hurricanes Irma and Michael. Finally, we visualize the spatio-temporal dynamics of infrastructure disruptions in a map of the affected regions.

4.3 Data Preparation

In this study, we use Twitter data collected during hurricanes Irma and Michael for creating a dynamic disruption map of critical infrastructure disruptions. We use two different methods (Twitter streaming API and rest API) for data collection. A brief description of the data is provided in Table 4.1.

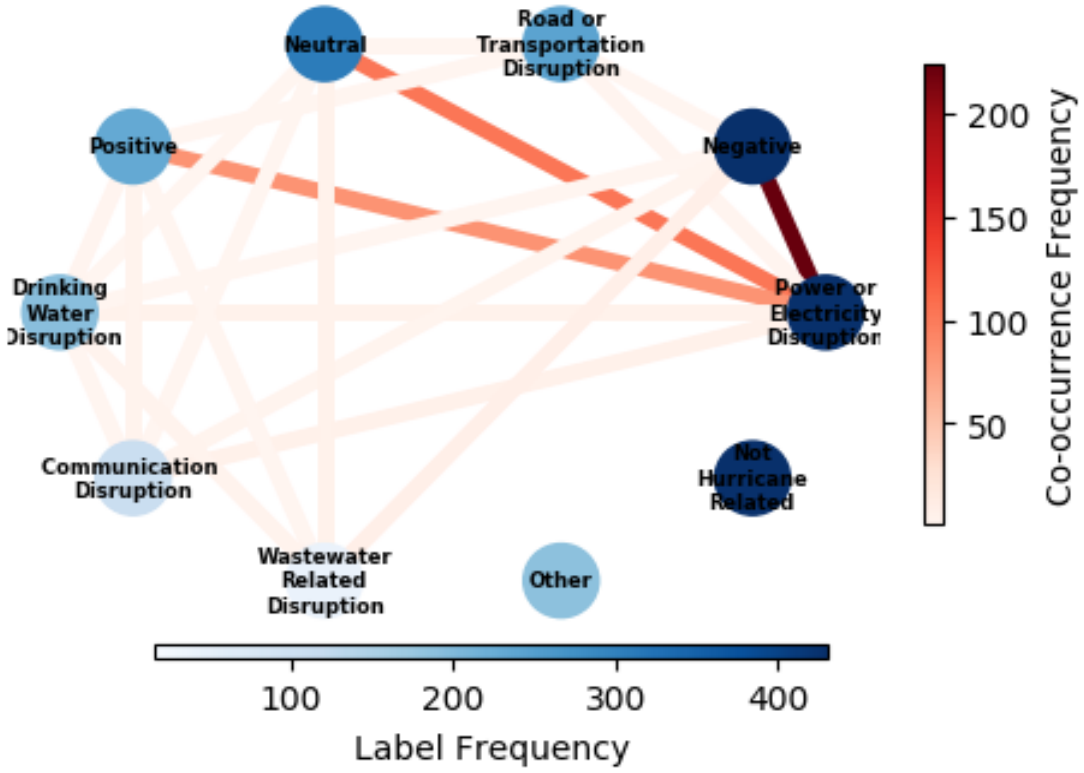


Figure 4.1 Distribution of Label Frequency and Label Co-occurrence Frequency

Table 4.1 Data Description

Hurricane Name	Regions (USA)	No. of Tweets	No. of Users
Irma (Streaming API)	FL, GA, SC	1,810,000	248,763

Irma (Rest API)		2,478,383	16,399
Michael (Streaming API)	FL, GA, SC, NC	3,534,524	1,289,204

Using the streaming API, we collected about 1.81 million tweets posted by 248,763 users between September 5, 2017 and September 14, 2017 during hurricane Irma. We collected the tweets using a bounding box covering Florida, Georgia, and South Carolina. To collect data for more time span and to fill some missing values contained in the steaming API data, we used Twitter's rest API to gather user-specific historical data. Twitter's rest API allows collecting the most recent 3,200 tweets of a given user. We collected user-specific data for 19,000 users, who were active for at least three days within the streaming data collection period. Similarly, we collected data for hurricane Michael using a bounding box covering Florida, Georgia, South Carolina, and North Carolina, containing 3.53 million tweets posted by 1.29 million users covering from October 8, 2018 to October 18, 2018.

To create an annotated disruption dataset, we manually labeled 1,127 tweets from hurricane Irma and 338 tweets (for testing purpose only) from hurricane Michael. The tweets were labeled by 5 human annotators. To ensure that we retrieve the right labels of the disruption types and sentiments, we only considered the labels when all 5 annotators agreed on it. Each tweet can have one or more labels out of the ten possible labels including: *not hurricane related, power/electricity disruption, communication disruption, road/transportation disruption, drinking water disruption, wastewater related disruption, other disruption, positive, negative, and neutral*. The first label indicates whether a tweet is hurricane related or not. The next five labels indicate five types of infrastructure disruptions. The label, *other disruption,* indicates a disruption that does not fall into

the five types of infrastructure disruptions considered here. The last three labels indicate the possible sentiment towards a disruption. We give below three examples of disruption related tweets:

- This tweet -"*Update I'm the only community in my area with power I feel really lucky right now but I hope everyone else is safe*"- mentions about *power/electricity disruption* but in a *positive* way. We would label such a tweet as *(power/electricity disruption, positive).*
- This tweet- "*we are in Clermont on Lake Minnehaha. We have no cable or power & cell service is spotty. When will be the worst here*"- mentions about both *power/electricity* and *communication disruptions*. We would label this tweet as *(power/electricity disruption, communication disruption, negative).*
- This tweet -"*im trying to eat and watch as much netflix as i can just incase my power go out*"- mentions about power/electricity disruption but does not indicate an actual disruption. We would label it as *(power/electricity disruption, neutral).*

Figure 4.1 shows the frequencies and co-occurrences of the labels in the annotated dataset. It shows that the annotated data contain many "*not hurricane related"* tweets. Among the tweets related to different types of disruptions, power/electricity related disruptions have the highest frequency. Among the sentiment related labels, *negative* sentiment has the highest frequency. On the other hand, power/electricity disruption and negative sentiment are the most frequently co-occurred labels in the annotated dataset.

4.4 Methodological Approach

The methodological approach adopted in this study has three main parts. The first part takes tweet texts as input and identifies disruptions and the sentiment towards the disruption. The second part

extracts the geo-location from the tweet's metadata or text. The third part visualizes the disruptions geographically in a dynamic map of disruptions. Figure 4.2 shows the steps and information flow among those steps. Each part of the framework is described below:

4.4.1 Disruption Identification

The objective of this step is to identify infrastructure disruptions and sentiments from a given text input, where more than one disruption type might be present. We use a supervised multi-label classification approach. The input texts collected from Twitter posts contain many noises, which may degrade classification performance. Therefore, we process the data before feeding it into the model. The sequential steps are shown in Figure 4.2 (left side).

4.4.2 Data Pre-processing

In this step, we discard the unnecessary tweets and remove noise from a tweet. Since a retweet (starting with RT in the texts) does not provide any new information in the dynamics of disruption, we discard retweets from the data to avoid false spike in the disruption count. To clean the tweet texts, we remove the stop words (e.g., 'a', 'an' and 'the'), short URLs, emoticons, user mentions (@Twitter user name), punctuations, and special characters (\@/#$ etc.). Finally, we tokenize (splitting texts into words) the texts and apply lemmatization (converting the words into noun, verb etc.) and stemming (converting words into root form) to the tokens.

4.4.3 Data Processing

In this step, we process the data for training models and predicting disruptions. In machine learning, training of a model refers to providing it with training data, which contains both inputs and correct answers, so that the algorithm can find the pattern to map the input features to the target/output features. We convert the preprocessed tokens as TF-IDF (Term Frequency-Inverse

Document Frequency), which measures the importance of a word in a document of a corpus (collection of documents). The details on TF-IDF can be found in this study [223]. The TF-IDF of a term/word (w) is calculated as follows:

$$TF\text{-}IDF(w) = TF(w) \times IDF(w) \tag{4.1}$$

where,

$$TF(w) = \frac{the\ number\ of\ times\ term\ w\ appears\ in\ a\ document}{the\ total\ number\ of\ terms\ in\ the\ document}$$

$$IDF(w) = log\frac{the\ total\ number\ of\ documents}{the\ number\ of\ documents\ with\ term\ w}$$

We create the TF-IDF using both unigram and bigram of words. We remove the features that appear in less than 2 documents. To remove the effect of total word counts in a document, we apply $l2$ normalization (sum of the squared value of TF-IDF =1 for a document). To prevent data leakage, we calculate the TF-IDF considering the tweets available in the training dataset. The output of the model may contain multiple disruptions; thus we convert the annotated labels into multi-label formats. We represent the multi-label output as a binary/one hot encoded matrix indicating the presence of disruption type and the sentiment label. In our study, we have 10 possible labels, so, each converted label is represented as 1×10 binary matrix where the value 1 represents the presence and the value 0 represents the absence of a particular label.

4.4.4 Model Selection

The objective of this step is to find the best model that maps an input tweet text to the binary matrix representing one or more types of infrastructure disruptions and sentiment. In our study,

we choose a multi-label classification approach for identifying disruptions and sentiments. This approach generalizes the multiclass classification, where a single input/tweet can be assigned to multiple types of disruptions. Let $L = \{\lambda_i\}$ be the set of labels containing disruption types and sentiment, where, $i = 1 \ldots \ldots . |L|$. In our case, $|L| = 10$. The objective of our disruption identification model, h is that: given the input tweet, X the model has to predict the disruption types and sentiments, $Y \subseteq L$.

$$h: X \to Y \tag{4.2}$$

We apply three methods that allow using the multiclass classification models for a multi-label classification task. The first method transforms a multi-label classification into multiple binary classification problems. This method is also known as binary relevance (BR) [222] that trains one binary classifier for each label independently. The equation for a binary classifier, h_{λ_i} for a label λ_i can be expressed as below:

$$h_{\lambda_i}: X \to \{\neg\lambda_i, \lambda_i\} \tag{4.3}$$

The BR method transforms the training data into $|L|$ datasets. The dataset D_{λ_i} for label λ_i contains all the original dataset labeled as λ_i if the original example contains λ_i, otherwise, as $\neg\lambda_i$. For an unseen sample, the combined model predicts all labels using the respective classifier. One of the disadvantages of the BR method is that it does not consider the correlation between labels.

The second method transforms the multi-label classification problem into a multi-class classification problem. This method is known as label powerset (LP) that considers each subset of L as a single label. Let, $P(L)$ be the powerset of L, which contains all possible subset of L. LP method considers each element of $P(L)$ as a single label. Now, in training LP learns one single label classifier h, where:

$$h: X \rightarrow P(L) \tag{4.4}$$

The LP method has advantages over the BR method, because it takes the label correlations into account. However, it requires high computation time if the size of $P(L)$ is very big and majority of the subsets have very few members. Also, the LP method tends to overfit (performs well on training data but performs poorly on test data), when the number of labeled samples of the generated subsets is low.

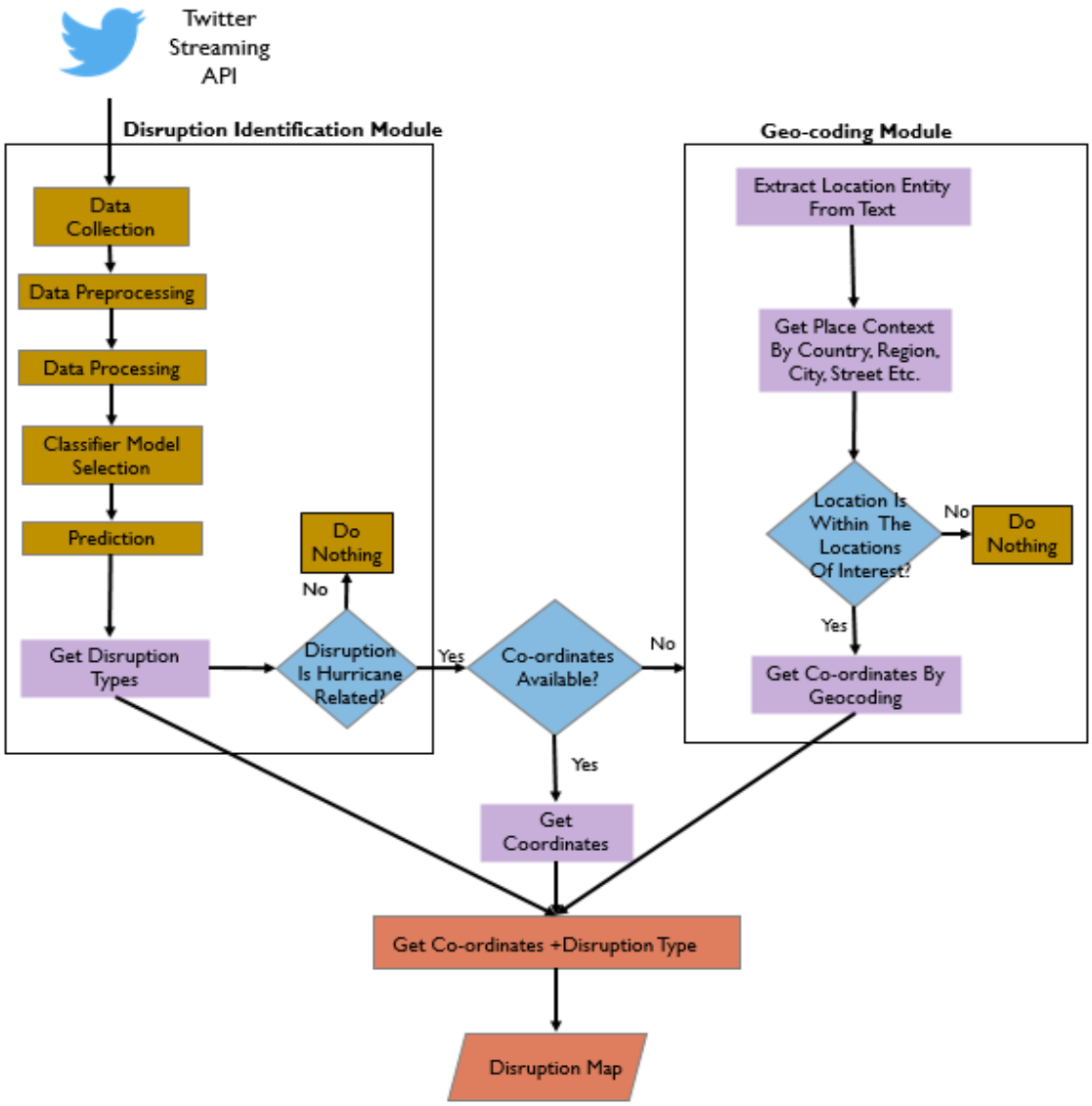


Figure 4.2 Methodological framework: disruption identification module (left); geo-parsing module (right); and visualization module (middle)

As the third method, we apply an ensemble technique, known as Random *k*-Labelsets (RAKEL) adopted from the study [224]. This method constructs an ensemble of LP classifiers, where each LP classifier is trained on a small random subset of labels. Instead of using $P(L)$, it creates *k*-labelset $Y \subseteq L$, where $k = |Y|$. If the set of all distinct *k*-labelset is L^k, then $|L^k| = \binom{|L|}{k}$. Given a user specified integer value for k and m, where, $1 \leq k \leq |L|$ and $1 \leq m \leq |L^k|$, the RAKEL algorithm iteratively constructs an ensemble of m numbers of LP classifiers. However, for $k = 1$ and $m = |k|$, RAKEL method becomes a binary classifier ensemble of BR method. On the other hand, for $k = |L|$, m becomes 1, and consequently, RAKEL method becomes a single label classifier of the LP method. Given a meaningful parameter of k ($2\ to\ |L| - 1$), at each iteration, $i = 1 \ldots . . m$, without replacement it randomly selects a *k*-labelset, Y_i from L^k and learns an LP classifier, h_i. Where,

$$h_i: X \to P(Y_i) \tag{4.5}$$

For a given input, the label prediction is accomplished by a voting scheme from the ensemble combination. The RAKEL method solves the overfitting problem of the LP method but loses some correlations as it considers a random subset of the labels (LP method considers all possible subsets). The full description of the RAKEL method can be found in this study [224].

In multi-label classification, a prediction cannot be assigned as a hard right or wrong value, because a prediction containing a subset of the actual classes should be considered better than a prediction that contains none of them. Thus, traditional performance metrics (e.g., precision, recall) are not suitable for evaluating our disruption identification model. We choose the best model based on three generally used performance metrics in multi-label classification: subset accuracy, micro F1 score, and hamming loss. Here, subset accuracy and hamming loss are

example-based metrics and micro F1 measure is a label-based metric. For each test sample, an example-based metric computes the difference between true and predicted class labels and then calculate the average over all test samples. Whereas, a label-based metric first computes the performance for each class label, and then calculates the average over all class labels. Assuming y as the set of true class labels, $\hat{y}$ as the predicted set of labels, L as the set of labels, y_l as the subset of y with label l, $\hat{y}_l$ the subset of $\hat{y}$ with label l, $n_{samples}$ the number of samples, the equations of these metrics are given below:

$$Subset\ Accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i \neq y_i) \tag{4.6}$$

$$Micro\ F_1\ Measure(y, \hat{y}) = 2 \times \frac{\frac{|y \cap \hat{y}|}{|y|} \times \frac{|y \cap \hat{y}|}{|\hat{y}|}}{\frac{|y \cap \hat{y}|}{|y|} + \frac{|y \cap \hat{y}|}{|\hat{y}|}} \tag{4.7}$$

$$Hamming\ Loss(y, \hat{y}) = \frac{1}{n_{samples} \times |L|} \sum_{i=0}^{n_{samples}-1} \sum_{l=0}^{L} 1(\hat{y}_l \neq y_l) \tag{4.8}$$

We further check the predictive performance of the model computing a confusion matrix for each label (representing disruption types and sentiment). Table 4.2 shows the components of a confusion matrix. The rows represent the actual labels and the columns represent the predicted labels where positive means the existence of a particular label and negative means the absence of a particular label. For a particular sample, if the actual label is negative, a negative prediction by the model is assigned as true negative and a positive prediction is assigned as false positive.

Similarly, if the actual label is positive, a positive prediction is assigned as true positive and a negative prediction is assigned as false negative.

Table 4.2 Confusion Matrix

		Predicted Label	
		Negative (0)	Positive (1)
Actual Label	Negative (0)	True Negative (TN)	False Positive (FP)
	Positive (1)	False Negative (FN)	True Positive (TP)

4.4.5 Disruption Location Extraction

The objective of this step is to extract the location of the disruptions that are identified by the previous step. Geo-tagged tweets provide location information either as a point type (exact latitude-longitude) or as a polygon type (bounding box). We use this location to indicate the location of a disruption either at a point resolution or a city/county resolution. However, geo-tagged tweets are only a few percentages (1% to 4%) of the total number of tweets. To address this limitation, we implement a location extraction method from tweet texts. This approach has several steps within it. Given a tweet text, the first step is to label each word (e.g., person's name, location, organization etc.), which is known as Named Entity Recognition (NER). We implement our NER model using the Natural Language Toolkit (NLTK), developed by [225]. The second step is to extract the location entity, words that are tagged as location, from the labeled words. In the third step, we match the extracted location with the county/city names of the affected regions.

Finally, if the extracted locations are matched, we collect the coordinates using the geo-coding API provided by Google Maps. The process of location extraction is shown in Figure 4.2.

4.4.6 Dynamic Disruption Mapping

This part of the methodology enables the visualization of the locations of disruptions with disruption types in a dynamic way. We visualize the exact disruption location, only if the location has the exact co-ordinate (location type: point or latitude-longitude). We choose a time interval (t) to count the number of disruptions within a geographical boundary (e.g., county) and then visualize the disruption intensity as a geographical heat map. We did not consider disruption severity in this study. But severity can be assumed to be correlated with the frequency of disruption related tweets from a given area; the higher the frequency of disruption related tweets the higher will be the severity level of disruptions. Hence, a dynamic disruption map can provide insights about the severity of infrastructure disruptions of an area based on the frequency of a specific or all disruption related posts generated from that area.

4.5 Results

Using Twitter data from real-world hurricanes, we present our results to identify infrastructure disruptions and visualize those disruptions in a dynamic map. To identify disruptions types and sentiment from text data, we use Binary Relevance, Label Powerset, and ensemble based multi-label classification approaches. We compare the performance of these approaches using eight existing models namely: Multinomial Naïve Bayes (MNB), Logistic Regression (LR), *K*-Nearest Neighborhood (KNN), Support Vector Machine (SVC), Random Forest (RF), Decision

Tree (DT), Multilayer Perceptron (MLP), and Deep Neural Network (DNN) methods. The details of these well-known methods can be found in these studies [226, 227]. We convert the annotated tweet text as TF-IDF and annotated label as binary matrix (multi-label format) by following the

Table 4.3 Keywords for Identifying Disruption Related Tweets

Disruption Types	Keywords
Power/Electricity Disruption	power, electricity, outage, (power, outage), (without, power)
Communication Disruption	internet, wi-fi, cell, (no, internet), (no, network)
Transportation Disruption	road, roads, traffic, transportation, turnpike, i-4, i-95, jam, closed, (traffic, signal), (road, closed)
Drinking Water Disruption	drinkingwater, drinking_water, bottledwater, bottled_water, (drinking, water), (bottled, water)
Wastewater Related Disruption	wastewater, waste_water, drainage, drainagewater, (waste, water), (drainage, water)

steps described in the data processing section. We use the TF-IDF as input and the binary matrix as output. For each model, we use 70% (788 tweets) of the annotated samples as training and the rest 30% (339 tweets) as test samples. We further validate our best model over 338 tweets from hurricane Michael to test model performance on the data from an unseen hurricane (i.e., for hurricane data which were never used for training the model).

We implement all the models in a personal computer using Python programming language and model parameters are selected using a grid search approach [228]. Moreover, we implement a baseline method that uses keyword matching and sentiment analysis to identify disruptions and

sentiment characteristics, respectively. Currently no benchmark method exists that can identify the co-occurrence of multiple types of disruptions from social media posts. Since a keyword based approach has been used in similar studies [31, 52], we choose to use this as a baseline method. The keywords used are listed in Table 4.3.

For sentiment identification, we use a pre-trained model adopted from this study [229]; this model has been trained on social media texts. We consider this combined (keyword matching and sentiment identification) approach as a baseline method to evaluate if the trained models perform better than this baseline method. Table 4.4 presents the performance of each model on hurricane Irma test dataset with respect to the selected performance metrics: subset accuracy, micro F1 measure, and hamming loss.

From the results, we can see that Logistic Regression classifier (LP method) has the best subset accuracy and micro F1 scores and Support Vector classifier (RAKEL method) has the best hamming loss score. The models (LR, KNN, SVC, MLP, and Deep DNN) perform better than the baseline method in all approaches (BR, LP, and Ensemble) (see Table 4.4). Among the three multi-label approaches, LP has the best performance; RAKEL is second; and BR method is the last in terms of the considered performance metrics. The reasons for this result are the following: (i) BR method considers the labels as mutually exclusive or the correlation between the disruptions is ignored; (ii) LP method considers the correlations between the labels/disruptions by considering all label combination; and (iii) RAKEL method falls between the BR and LP methods with respect to label correlations as it considers a random small subset of labels.

To select the best model, we further check the confusion matrix and choose Logistic Regression (LP method) classifier. Figure 4.3 shows the confusion matrix for the LR (LP) on the test samples from hurricane Irma. The selected best model (LR-LP) shows 74.93% increase (0.351 to 0.614)

in subset accuracy, 30.73% increase (0.550 to 0.719) in micro F1 measure, and 44.65% decrease (0.159 to 0.088) in hamming loss compared to the baseline method.

Table 4.4 Model Performance Values (Accuracy, Micro F1-measure, Hamming-loss) (A higher score of subset accuracy or micro F1 measure indicates better performance and a lower score of hamming loss indicates better performance)

Model Name	**Binary Relevance (BR)**	**Label Power set (LP)**	**Ensemble (RAKEL)**
Baseline (keyword search + sentiment)	*0.351, 0.55, 0.159*		
Multinomial Naïve Bayes (MNB)	0.218, 0.519, 0.145	0.472, 0.615, 0.14	0.268, 0.527, 0.151
Logistic Regression (LR)	0.463, 0.709, 0.090	**0.614**, **0.719**, 0.092	0.525, 0.702, 0.094
K-nearest Neighborhood (KNN)	0.490, 0.613, 0.130	0.525, 0.612, 0.125	0.510, 0.598, 0.126
Support Vector Classifier (SVC)	0.472, 0.707, 0.089	0.608, 0.699, 0.096	0.519, 0.709, **0.088**
Random Forest (RF)	0.124, 0.471, 0.170	0.54, 0.635, 0.116	0.357, 0.588, 0.109
Decision Tree (DT)	0.292, 0.628, 0.129	0.522, 0.634, 0.119	0.366, 0.617, 0.124

Multilayer Perceptron (MLP)	0.440, 0.662, 0.099	0.540, 0.615, 0.119	0.507, 0.635, 0.11
Deep Neural Network (DNN)	0.466, 0.342, 0.138	0.569, 0.684, 0.103	-

We also check the performance of our best model (LR-LP) for disruption and sentiment identification separately. We validate for hurricanes Irma and Michael, using 339 test data from hurricane Irma and 338 test data from hurricane Michael. Table 4.5 shows the performance on disruption identification.

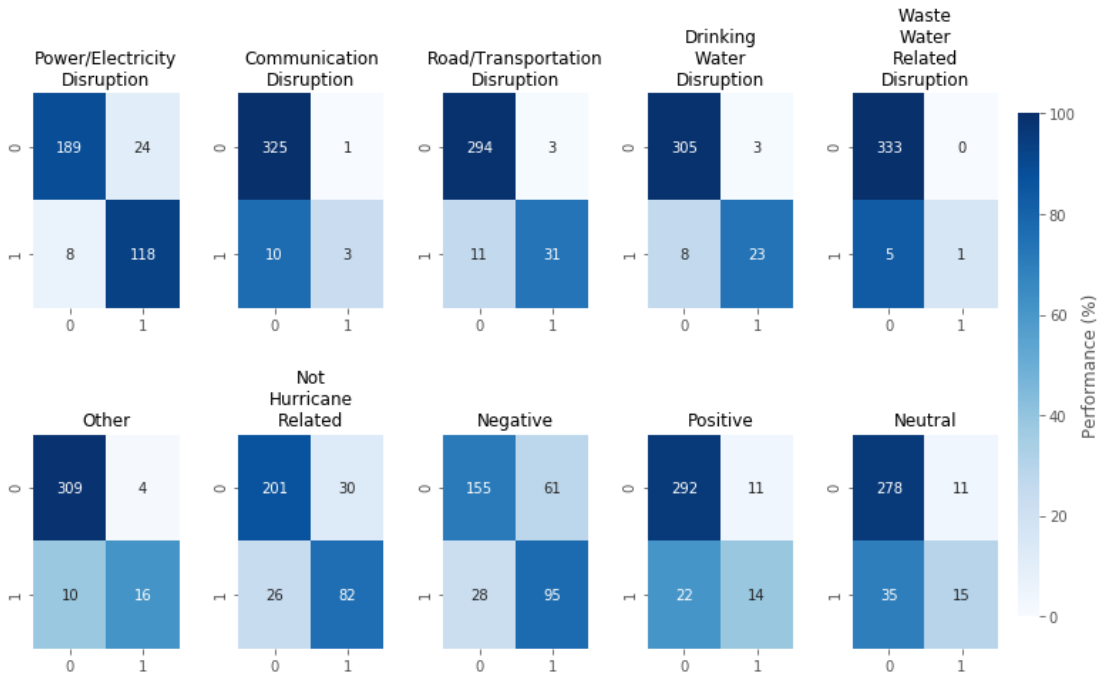


Figure 4.3 Confusion Matrix (In each panel, the x axis represents the predicted label and the y axis represents the actual label in the test set of hurricane Irma. For a particular label, the value 1 means the presence of this label whereas 0 means the absence of the label. The value within a cell represents the number of times a predicted label matched or mismatched with the actual label)

Except hamming loss for hurricane Michael, our model performed better for both hurricanes with respect to accuracy, micro F1, and hamming loss. The baseline method performed better in

Table 4.5 Performance Comparison of disruption identification

	Baseline	Model (LR-LP)
Hurricane	Accuracy, Micro F1-measure, Hamming-loss	
Irma	0.351, 0.55, 0.159	0.614, 0.719, 0.092
Michael	0.476, 0.656, 0.115	0.515, 0.658, 0.119

hurricane Michael test set than the Irma test data set. On the other hand, LR-LP model performed better in Irma data than the Michael data since the model is trained on Irma dataset.

Table 4.6 Performance comparison of sentiment model

	Baseline	Model (LR-LP)
Hurricane	Accuracy, Micro F1-measure, Hamming-loss	
Irma	0.383, 0.368 0.311	0.673, 0.596, 0.165
Michael	0.571, 0.501, 0.226	0.609, 0.656, 0. 175

Table 4.6 shows the performance of LR-LP model against the baseline sentiment model (adopted from [229]). The (LR-LP) model performed better than the baseline for both Irma and Michael datasets. The baseline method also performed better for Michael data than Irma data for

sentiment classification. In summary, our developed model (LR-LP) performed better than the baseline for both hurricanes Irma (hurricane data used to train the model) and Michael (unseen hurricane data representing a future hurricane).

To understand the features that help to correctly identify a disruption, we analyze the training samples that our model correctly predicted (i.e., true positive samples in Table 4.2). For each disruption type, we rank the words based on their average TF-IDF score. A higher score represents more importance of a word for a disruption type. Figure 4.4 shows the TF-IDF scores of the top ten words of each disruption type (shown as horizontal bars) and the TF-IDF scores of the same words calculated over all disruption types in the training set (shown as color intensity). We can see that overall words such as 'power', 'water', 'wifi', 'internet', 'traffic', 'drainage' etc. have higher TF-IDF scores (see the color intensity of the corresponding bars in Figure 4.4). It means that these words are highly important in the overall classification performance. On the other hand, 'power', 'cell', 'stop', 'water', 'drainage', 'close' are the highest ranked words for power/electricity disruption, communication disruption, road/transportation disruption, drinking water disruption, waste water related disruption, and other disruption, respectively. Some words (e.g., 'power', 'water', 'cell') are present in multiple disruption types, indicating that these words would help identify the co-occurrence of multiple disruption types. For example, the presence of 'cell' and 'signal' in the top 3 words of power/electricity and communication disruptions indicates the co-occurrence of these two types of disruptions. Regarding sentiment features, the word 'power' is common in all the three sentiments. The differences among the words present in these three sentiment classes are: (i) the negative (actual disruption) contains the words that are mostly present in the disruption types, (ii) the positive sentiment contains slang words such as 'hell yeah',

'yeah', 'ac loll', (iii) neutral sentiment contains situation and forecast related words such as 'update', 'best update', 'situation', 'chance wont', 'good chance' (see Figure 4.4).

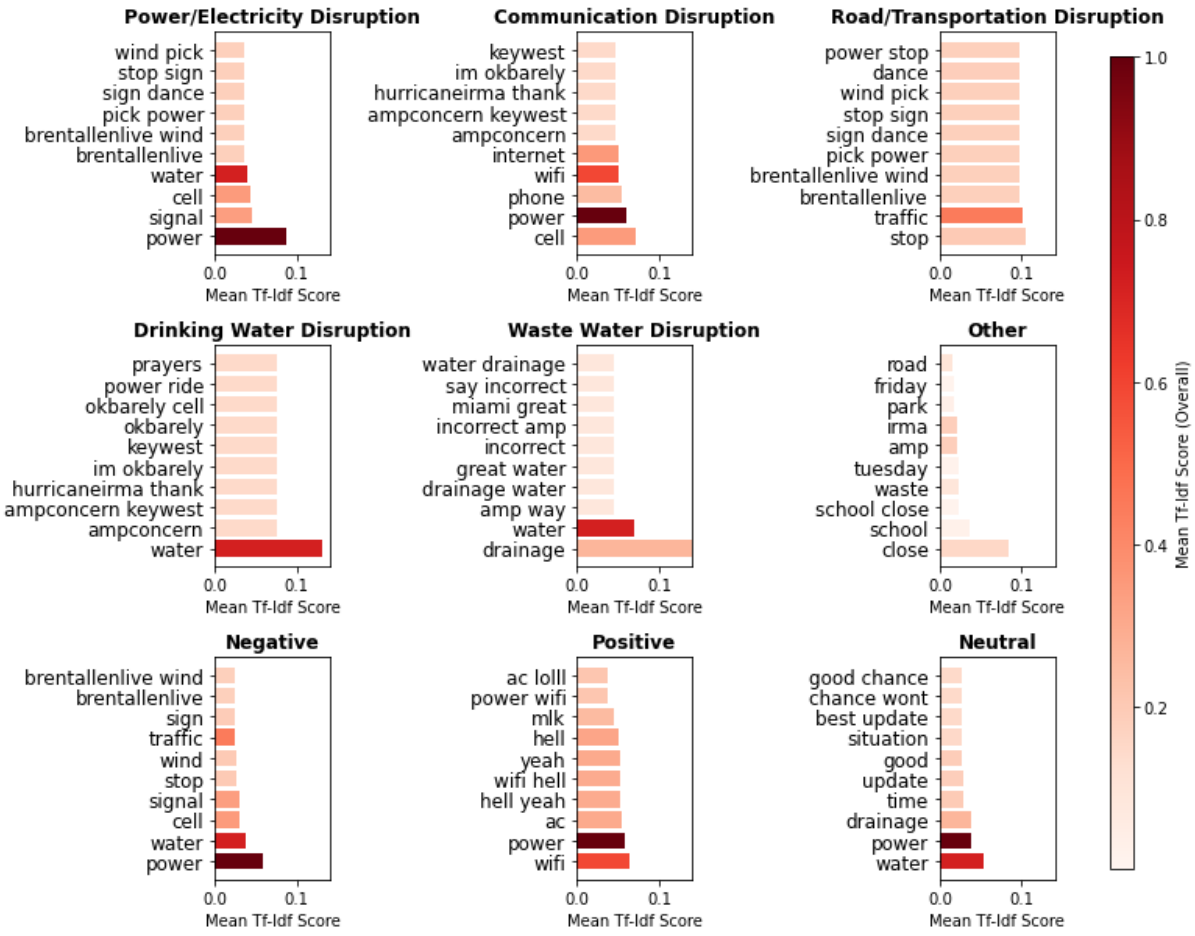


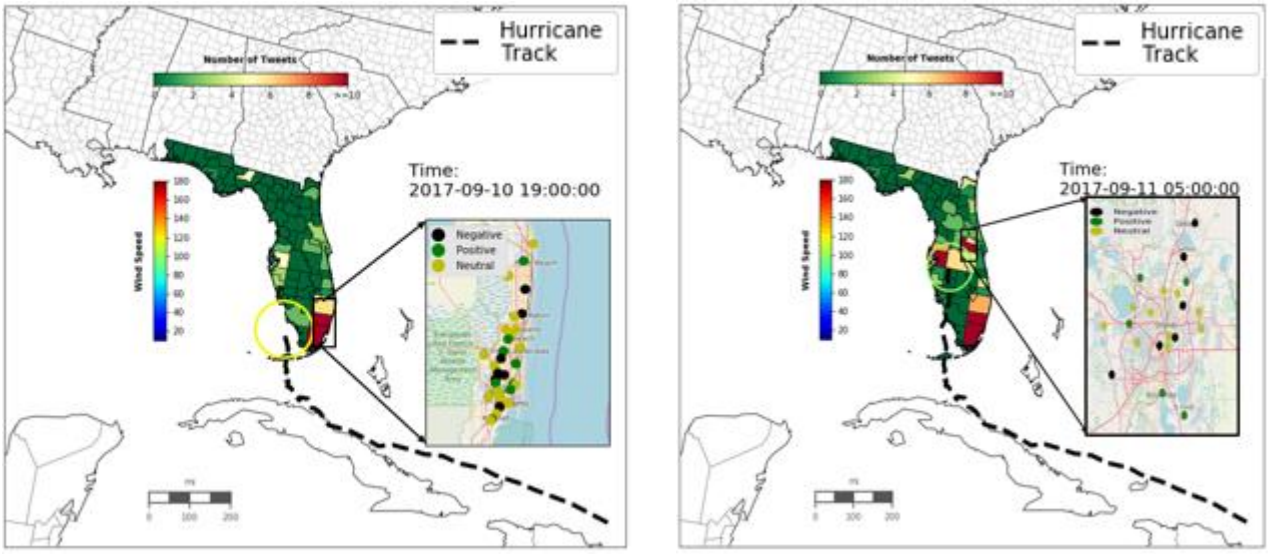
Figure 4.4 Important features for different disruption types. The X axis shows the mean TF-IDF score (calculated over individual disruption type) and the Y axis shows the words/features. The color of the bar indicates the mean TF-IDF score (calculated over all disruption types). The calculated scores and important features are based on the training dataset.

4.6 Case Studies: Hurricanes Irma and Michael

In this section, we present two case studies of our proposed approach, one for hurricane Irma and another for hurricane Michael. Our best model (LR-LP) predicts the disruption types and

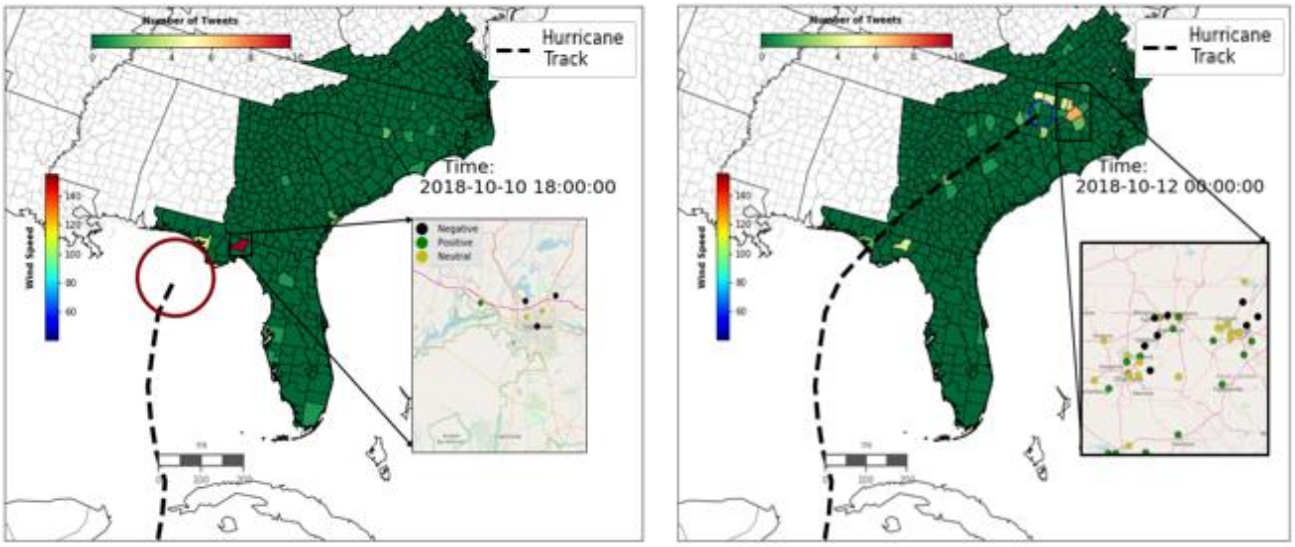
status over the input data described in Table 4.1. As shown in Figure 4.2, for a geotagged tweet, we obtain the disruption location from the tweet geo-location information. Otherwise, we extract the location from the tweet texts using the geocoding module. We match the extracted location with the city/county of a state and then obtain the coordinate using Google Maps API.

Finally, we plot the disruption types and status in a disruption map. To understand the hurricane context, we also present the hurricane track and wind speed data collected from the National



a.1 a.2

(a) Hurricane Irma



b.1 b.2

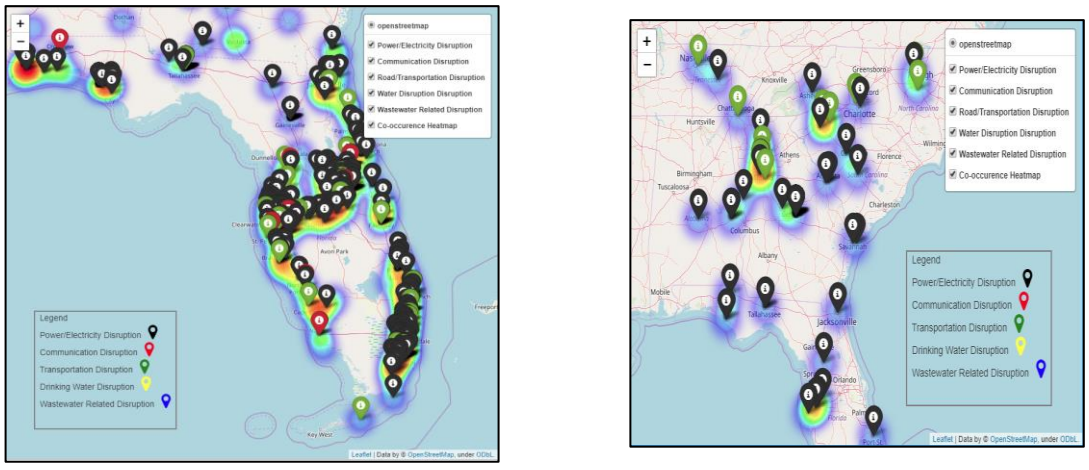
(b) Hurricane Michael

Figure 4.5 Dynamic Disruption Map for Power/Electricity Disruption. (a) Hurricane Irma, (b) Hurricane Michael.

Hurricane Center [230]. Two snapshots of the power/electricity disruption map from each hurricane are shown in Figure 4.5 (5a.1 and 5a.2 for hurricane Irma, 5b.1 and 5b.2 for hurricane Michael). We use a 3-hour time-interval for aggregating the tweets to create the county-level disruption heat map. The inset plot shows the locations of power/electricity disruptions. We show the location of hurricane center (shown as a circle at the beginning of the hurricane track line), wind speed (through the color of the circle), and disruption related tweets (geographic heat map) which will be updated dynamically as we receive data from Twitter stream. Figure 4.5a.1 shows a snapshot of Irma at around 7 PM on Sept. 10, 2017. It shows that majority of the power/electricity related posts were generated from Miami-Dade, Broward, Palm-Beach counties when Irma's center was near Collier county with a wind speed of around 120 mph. However, not

all the posts are about the actual power outage incident (disruptions are represented by black circles in the inset plot of Figure 4.5a.1), and a substantial number of these posts are expressing concerns about power outage or expressing that they still have power. The second snapshot (Figure 4.5a.2) shows that when the center of Irma was near Tampa, most of the disruption related tweets were posted from Orlando, Tampa, and Miami-Dade counties. A dynamic disruption map of Michael shows similar results. On October 10, 2018 around 6 PM (Figure 5b.1), when Michael was about to make its landfall near Tallahassee, most of the power/electricity disruption related tweets were coming from Tallahassee area. Figure 4.5b.2 shows the second snapshot of Michael around midnight of October 12, 2018 when the center of Michael was over North Carolina. It shows that most of power/electricity related disruptions were coming from Wake, Johnston, Durham and Orange counties of North Carolina.

Finally, we visualize the co-occurrence of multiple disruption types in an interactive map.



(a) Hurricane Irma

(Time 2017-09-11 05:00:00)

(b) Hurricane Michael

(Time 2018-10-12 06:00:00)

Figure 4.6 Disruption Co-occurrence Map (a) Hurricane Irma (b) Hurricane Michael

Figure 4.6 shows a snapshot of the co-occurrence map for hurricane Irma (Figure 4.6a) and Michael (Figure 4.6b). We plot this map using only the actual disruption samples (negative sentiment) aggregated over a 1-hour interval. This interactive map allows to explore the disruptions type separately as well as a combination of them. The co-occurrence heat map shows a relative intensity of the disruptions based on the co-occurrences of all the disruption types. For Irma, mostly co-occurred disruptions are power, communication, and transportation disruptions. On the other hand, for hurricane Michael (see Figure 4.6b) the most co-occurred disruptions are power and transportation disruptions.

In summary, we find that during hurricanes Irma and Michael affected people posted infrastructure related tweets. Those posts may represent actual infrastructure disruptions. A multi-label classification approach (a logistic regression model adopted over a label powerset) has been developed to predict both the disruption types and disruption status from such data. After locating the disruptions using a geocoding approach, a map can visualize the disruptions spatially and temporally. The training time of the model is about 7 sec, and it takes about 1 sec to process, predict, and visualize the data collected over one hour. Thus, this approach can be easily applied in a real-time setting.

4.7 Limitations and Future Research Directions

Our study has some limitations. For instance, the annotated dataset is small in comparison to the entire dataset. More annotated samples are likely to increase the accuracy of the model. Although the co-occurrence of multiple disruptions is considered, the approach cannot infer if a disruption is caused by another disruption. Incorporating causality as an input to the model may

improve its performance. Another limitation of our approach is that we have checked the accuracy of the method based on human-annotated tweets, which may not represent the total number of disruptions observed in the ground. To check the extent to which the reported disruptions match actual ones, ground truth data on disruptions occurring in different infrastructure systems are required. These datasets, often collected by infrastructure service providers including private companies and public agencies, may contain sensitive information. Collecting ground truth data on infrastructure disruptions from a variety of sources covering multiple states will be a very challenging task. Further studies are needed to verify what percentage of actual disruptions is reported in social media and to what extent these disruptions can be identified using the method developed in this study. In addition, our data cover hurricanes only. Future studies can transfer and validate our approach across other disasters such as wildfire, earthquake, snowfall, and thunderstorms.

In this study, we assume that a post with a negative sentiment is associated with an actual disruption, and a post with a neutral or positive sentiment is associated with no disruption. However, there could be a post with a positive sentiment, but associated with an actual disruption. These tweets are likely to be a small portion of the entire dataset. In our annotated dataset, we did not find such tweets. Future studies, based on natural language processing, can develop more advanced methods to capture the situations where even a positive tweet could be associated with a disruption.

When the communication network is disrupted, affected people may not have access to social media platforms. In such situations, our model cannot detect disruptions. In the geo-parsing method, we use exact matching process between the extracted location and county/city of the affected regions. Since our approach finds city/county names only, it cannot extract location if

street or any finer level location is mentioned in the text. In future studies, text-based location matching can be developed with finer resolution (e.g., street name), which may help in locating disruptions with more specific location information.

For training our models, we adopt a batch learning approach which requires retraining the model to incorporate new data from the data stream. Future studies can explore an incremental learning approach [231, 232] to dynamically train models on newly available data from the ongoing/future disasters [233]. Such an incremental learning approach is likely to increase the accuracy of the model as it utilizes data from an ongoing disaster.

To achieve a better classification accuracy, more complex classification methods such as probabilistic neural networks [234], dynamic neural networks [235], and hierarchy-based models [236, 237] can be considered. A probabilistic neural network is a fast, efficient, and flexible model to add/remove new training data and hence may be more suitable for real-time disruption prediction for an unseen disaster. Since textual data have a large feature space, a dynamic neural network might be useful in finding an optimal number of features to achieve better performance. Moreover, hierarchy-based models might be more suitable when there exists more hierarchy in the disruption types, especially considering disruptions from multiple disasters (hurricane, wildfire, snowstorm etc.). A hierarchy-based model can have classes for disaster type, disruption type, and disruption status. A hierarchical relationship can be created from disaster type to disruption type to disruption status (e.g., if a post is not disaster related it has no disruption type and disruption status).

4.8 Conclusions

This chapter presents an approach to identify infrastructure disruptions and a dynamic disruption mapping framework using social media data. While previous research focused mainly on identifying hurricane or damage related social media posts, we consider five types (power/electricity, communication, drinking water, and wastewater) of infrastructure disruptions, their co-occurrence, and their status (whether a post is reporting an actual disruption, disruption in general, or not affected by a disruption). The result shows that our multi-label classification approach (logistic regression adopted in a label powerset approach) performs better than a baseline method (based on keyword search and sentiment analysis). Moreover, we present a method, to visualize disruptions in a dynamic map. Identifying disruption types and disruption locations is vital for disaster recovery, response and relief operations. The developed approach of identifying the co-occurrence of multiple disruptions may help coordinate among infrastructure service providers and disaster management organizations.

CHAPTER 5: MODELING THE INFLUENCE OF MULTIPLE INFORMATION SOURCES ON RISK PERCEPTION DYNAMICS AND EVACUATION DECISION DURING HURRICANES: AN AGENT-BASED MODELING APPROACH

5.1 Introduction

In this study, we develop an agent-based model that combines hydrologic characteristics, socio-demographic characteristics, and multiple information sources to understand the risk perception and evacuation behavior of the households of an area. Previous studies suggested that people from Miami-Dade are more concerned about inland flooding from heavy rainfall than storm surge despite leaving nearby coastal area [238]. We use simulated run-off depth forecast considering the hydrological aspect of a region. Then we use the forecasted run-off depth to generate the initial risk perception of a household, which solves the requirement of initial seed generation. The novelty of our study is that we integrate detailed flood depth information, actual neighborhood (with the spatial distribution of households), and social network to model risk perception and evacuation decision of a household. We also model how these multiple information sources (flood forecast, neighbor observation, and opinion in social network) and household's trust on these information sources are linked to evacuation compliance and shadow evacuation. This study contributes to the literature by answering the following research questions:

- How does a household perceive the forecasted flood risk and makes evacuation decisions, and how does a households' trust on forecasts affect evacuation participation? We use a high resolution ($500\ meter \times 500\ meter$) rainfall-induced flood forecast data to model the flood risk and corresponding evacuation decision of a household.
- What is the effect of social networks on a household's evacuation decision, and how does a households' trust on opinion dynamics of social networks affect the overall evacuation

participation? We construct social networks between the households and model the influence of the opinion dynamics (circulating over the social networks) on the household's evacuation decision.

- How does observing neighbor activities affect a household's evacuation decision in an actual neighborhood setting, and how a household's trust on neighbors' observation affect the overall evacuation dynamics? Past studies have created agents' synthetic space or on aggregated space (census tract, blocks, etc.) that limits the understanding of actual spatial effect between the neighbors. We capture the actual neighborhood influence on evacuation decision by assigning agents at a building footprint level.

5.2 Literature Review

Studies on evacuation behavior can be divided into two main categories: (i) statistical models [130, 239, 240] that use empirical data collected by a survey to understand contributing factors of evacuation behavior; (ii) computational models [38–40] that use behavioral theories to simulate evacuation behavior. The statistical modeling approach includes different types of logit models, most frequently a binary logit model where evacuation decision is modeled a binary decision process (evacuate or not) [60, 108]. These studies have found that socioeconomic, demographic factors, social ties, etc., play an important role in evacuation decisions [60, 240]. The statistical modeling approach, however, cannot capture the dynamics of the collective evacuation behavior that evolves from the social interactions among households; because it is very challenging to collect such empirical data [41, 42]. A computational model such as an agent-based model allows incorporating the findings from the statistical models and other phenomena (risk propagation, dynamic forecast, social interaction) where a heterogeneous agent (household) interact with other agents (households) and take decisions, and hereby affect the collective

evacuation behavior [241]. Studies have found that the households reliability on weather forecast increases the likelihood to evacuate [55]; but the forecast might be less reliable and overblown for some households depending on the conveyed message [186, 238], and they might look for other information sources/channels such as community, peers, internet, etc. [242, 243].

Many studies have used agent-based models to understand evacuation behavior under the influence of social networks [38, 39, 244, 245]. Widener et al. found that social influence increases the evacuation participation using a random subset of population on a case study of Bay County, Florida [245]. Yang et al. [38] proposed a home-workspace based social network for Florida-key regions to simulate the effect of social networks on evacuation decisions. This study has the following limitations: it considered neighbors as a part of social networks and did not consider that households can observe their spatially close neighbors' activity (evacuating or not) without having a social link between them; Florida Key is a unique geographic location that might not be representative to other location. Moreover, all of these above mentioned ABM did not consider the credibility/trust on this peer pressure (or information from their connections).

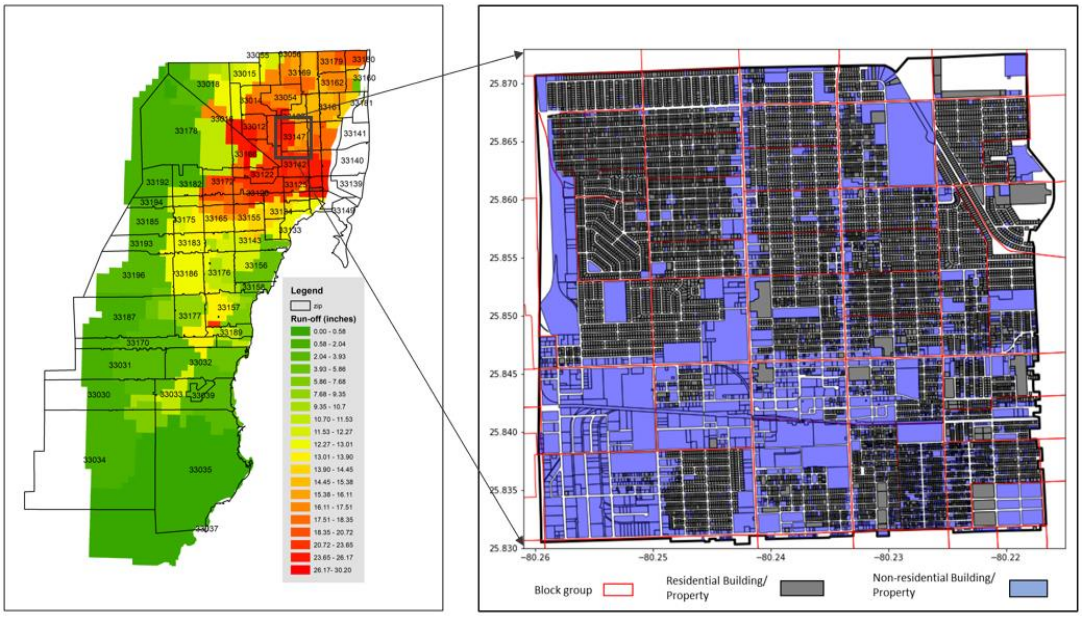
Du et al. [56] have used an agent-based model to simulate the effect of online social networks and neighbor observations separately on flood evacuation behavior considering the trust or weight on multiple information sources. However, this study is based on a hypothetical space and hazards; thus, representativeness of the results on actual social and neighbor effect is unknown.

In our study, we develop an agent-based model by incorporating flood risk and household's trust on multiple sources of influence (forecast, social networks, and neighbors' influence) to understand the dynamics of risk perception and evacuation behavior. We generate realistic run-off data of Miami-Dade County and create realistic households using the most recent census data. We

present our results on a case study created for the households located in a zip code of Miami-Dade County.

5.3 Data Description

To simulate the evacuation behavior in a real geographical area, we collect data from multiple sources such as flood forecast data, socio-demographic, building footprint, and property data. We generate synthetic runoff data using a process-based hydrologic model with ($500\ m \times 500\ m$) resolution. Please find the detailed procedures of the run-off depth forecast in the supporting information. We collect building footprint and property type from Miami-Dade county website [246]. The building or property data provide the geographical location of a building and type of uses (e.g. residential, commercial, unit count, floor count, etc.). We collect the 2018 5-year American Community Survey (ACS 5) [247] data for each block group of Miami-Dade county containing the following five variables: sex, age, income, race, and education.



(a) (b)

Figure 5.1 Study Area. (a) Run-off depth distribution on Zip codes over Miami-Dade County (b) Residential and non-residential property/building distribution over the block-group of Zip 33147.

Study Area

We choose zip code 33147 of Miami-Dade County as the area of interest for our simulation (see Figure 5.1). We choose this zip code because the runoff depth varies over the regions (block-groups) and it has socio-demographic diversity. Hence, this study area will capture the evacuation behavior on different level of flood risk and heterogeneity in risk perception over households. Figure 5.1(a) shows the run-off depth over Miami-Dade County – the run-off depth of zip code 33147 ranges from 13 to 26 inches. Figure 5.1(b) shows the spatial distribution of the building/property of zip 33147. It has 40 block groups where 32 are completely inside and 8 are partially inside the boundary of the zip code. Some block groups have a high density of residential building and some have a low density of residential buildings or households among these 40 block groups. Hence, the selected study area is well suited to capture the effects of flood risk and socio-demographic and neighborhood density on evacuation behavior.

5.4 Methodology

Our study has three main methodological parts: (i) creating synthetic households, (ii) creating social network among households, and (iii) modeling risk perception dynamics and evacuation behavior.

5.4.1 Creating synthetic households

We create 15,291 synthetic households in our study area. We assign the household characteristics (gender, employment, education, income, and race) according to block group level distribution of 2018 5-year American Community Survey data. We assign a single household to each unit of a

building according to the residential property data. Figure 5.2 shows the joint and marginal distributions of the characteristic of the synthetic households.

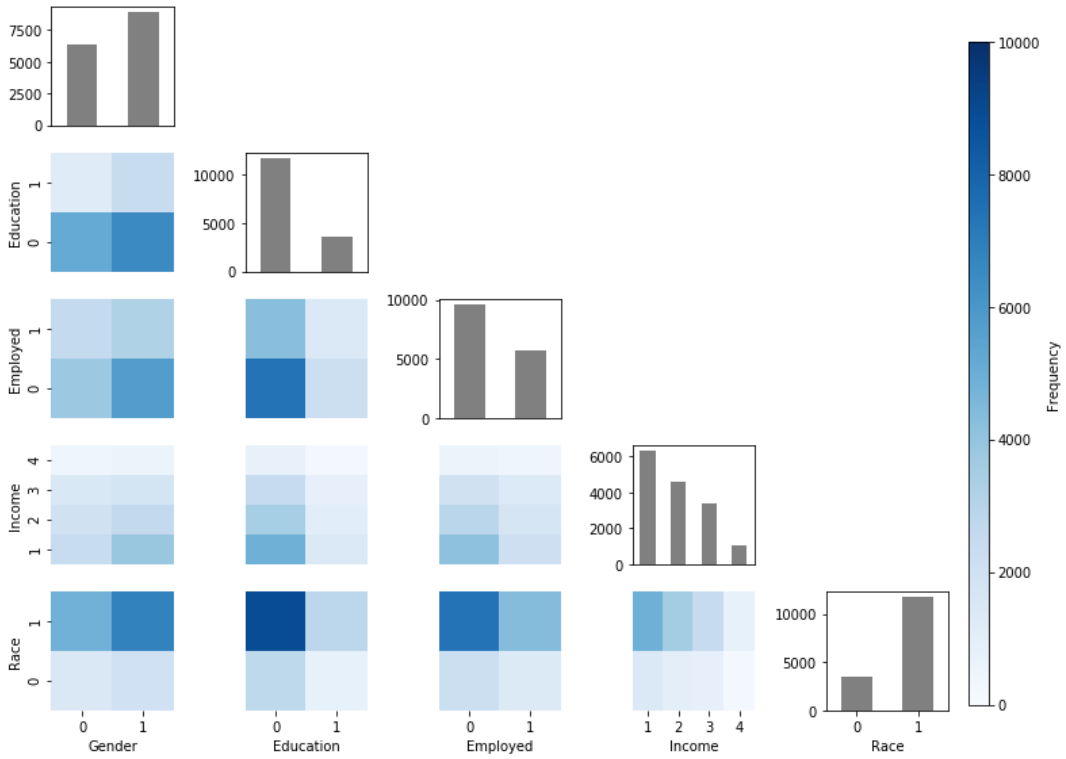


Figure 5.2 Joint and Marginal Distribution of the Synthetic Households. Here the variables are gender (0 = male, 1 = female); Education (1 = some college or higher, 0 = less than college), Employed (0 = unemployed, 1 = employed), Income (1= less than 25k, 2 = between 25k to 50k, 3 = 50k to 100k, 4 = greater than 100k), Race (0 = white, 1 = other).

Previous studies suggested that a household's evacuation tendency or willingness depends on its demographic and socio-economic variables. We estimate a household's evacuation tendency/willingness based on the household characteristics (as shown in Figure 5.2) using the binary logit model adopted from this study [248]. The binary logit model is given in equation 5.1.

$$Evacuation\ tendency, \pi_i = \frac{e^{\beta x_i}}{1+e^{\beta x_i}} \tag{5.1}$$

where, π_i is the evacuation tendency of household i, x_i are the characteristics variables of household i, β are the coefficients of the binary logit model adopted from [248].

5.4.2 Creating Social Network and Neighborhood

The evacuation decision of a household does not only depend on its characteristics but also on the surrounding neighborhood and social network [249]. Studies suggest that the connections among households are most likely to work as a small world network [38], proposed by Watts and Strogatz [250]. We create small world network between the synthetic households generated in the previous step. A household is more likely to connect with its nearby connections during an emergency. As our study area is a zip code with diagonal distance around 6.4 kilometers, all households are equally likely to be connected with each other. For each household, we create a small world undirected (i.e., communication can happen both ways) network with average degree, k_a of n. Where value of n can be 2 to 15 based on previous studies [38, 245]. We assign neighbors based on spatial proximity of the households ranging from 300 meters to 1 kilometer.

5.4.3 Risk Perception Dynamics and Evacuation Behavior

We develop a threshold-based model to simulate evacuation decision of the households, where each household (agent), j has a set of attributes: risk tolerance threshold, sensitivity towards landfall time, trust on different information sources, information search behavior, and learning attitude. The risk tolerance threshold of a household is assigned based on the evacuation tendency computed from the socio-economic characteristics of a household (see creating synthetic household section). We assign the risk tolerance based on the assumption that a household with low evacuation tendency has a high threshold of risk tolerance and vice versa [38]. Table 5.1 shows the assignment procedure of risk tolerance threshold.

Table 5.1 Assignment of Risk Tolerance Thresholds based on Evacuation Willingness/Tendency

Evacuation Tendency, (π)	Risk Tolerance Threshold (τ)	Percentage of Households
0 – 0.1	1.0	-
0.1 – 0.2	0.9	44.11
0.2 – 0.3	0.8	23.76
0.3 - 0.4	0.7	18.80
0.4 – 0.5	0.6	8.91
0.5 – 0.6	0.5	3.52
0.6 – 0.7	0.4	0.83
0.7 – 0.8	0.3	0.065
0.8 – 0.9	0.2	-
0.9 – 1.0	0.1	-

Following previous studies [56, 251], we model the risk perception of a household as a continuous variable. Each household will update its risk perception over time and will decide whether to evacuate or not. A household will evacuate if the risk perception is greater than the risk tolerance threshold. Let, $R_{j,t}$, τ_j are the risk perception of a household and risk tolerance of household j at time t, respectively. If $R_{j,t} > \tau_j$ household j will evacuate ($E_t = 1$), otherwise the household will stay ($E_t = 0$) at time t.

$$E_t = \begin{cases} 0 & if\ R_{j,t} > \tau_j \\ 1 & if\ R_{j,t} \leq \tau_j\ or\ E_{t-1} = 1 \end{cases} \quad (5.2)$$

At any given time t, a household forms its risk perception based on the influence I_t received from multiple sources of information and the perceived time difference from landfall (see Equation 5.3). Hurricane landfall time plays a crucial role in shaping household risk perception and each household perceives the risk differently with respect to the time difference from landfall [42] as follows:

$$R_{j,t} = I_{j,t} \times e^{-\frac{(t-C)^2}{2\sigma^2}} \quad (5.3)$$

where, C is the mean and σ is the standard deviation of the time difference from landfall when the risk perception is maximum. A household may not collect new information at every time step; thus, we model the risk perception dynamics as a stochastic process where the parameter $p_{j,t}$ ($update = 1$) determines whether a household j will search for new information or not. If a household chooses not to collect any new information, then $I_{j,t} = I_{j,t-1}$. At time t, if a household j choose to collect new information it will look for 3 information sources: 1) hazard risk forecast, $I_{j,t}^H$ 2) social network/media, $I_{j,t}^S$ 3) neighborhood activity, $I_{j,t}^N$.

$$I_{j,t}^{New} = \alpha_j I_{j,t}^H + \beta_j I_{j,t}^S + \gamma_j I_{j,t}^N \quad (5.4)$$

where, $\alpha_j, \beta_j, and\ \gamma_j$ are the trust (or weight) parameters on building household $j's$ influence on new information and $\alpha_j + \beta_j + \gamma_j = 1$. Because of the opinion adherence tendency, households do not completely abandon their past information influence [56, 252]. The parameter θ_j dictates what percentage of newly formed influence will be added to form the latest information influence I_t. Here, θ_j also serves as the learning rate of a household j in a Widrow-Hoff learning rule [56, 253].

$$I_{j,t} = I_{j,t-1} + \theta_j \Delta I_{j,t} = I_{j,t-1} + \theta_j (I_{j,t}^{New} - I_{j,t-1}) = \left(1 - \theta_j\right) I_{j,t-1} + \theta_j I_{j,t}^{New} \tag{5.5}$$

In our study, we represent the hazard risk in terms of flood risk due to heavy rainfall during a hurricane. Since our study area is less likely to face storm surge (not within the storm surge zone) and wind gust variation (because of small geographical area), we assume that a household will consider only the inland flooding to form its hazard related influence on risk perception. Here we introduce a threshold run-off depth τ_{rd} that will dictate whether emergency officials will declare evacuation or not. At time t, a household will form a hazard related risk by the amount of the ratio of the forecasted runoff depth at the home location at household j and τ_{rd}.

$$I_{j,t}^{H} = \frac{runoff\ depth\ at\ j's\ home\ location}{maximum\ permissible\ runoff\ depth\ (\tau_{rd})} \tag{5.6}$$

A household will also collect information from its social network. At time t, the collected information is modeled as a linear combination of the information obtained from its connected households following the setting of previous studies [56, 254]. A household may not collect information from all the connected households of its social network. Here we assume that a household is more likely to collect information from the connected households who live closer to that household.

$$I_{j,t}^{S} = \sum_{i=1}^{n} w_{i,j} I_{i,t-1} = \sum_{i=1}^{n} \frac{a_{ij,t}}{\sum_{i=1}^{n} a_{ij,t}} I_{i,t-1} \tag{5.7}$$

where $a_{ij,t}$ represents whether household j has read or collected $j's$ opinion ($a_{ij,t} = 1$) or not ($a_{ij,t} = 0$) and probability of $a_{ij,t} = 1$ depends on the distance, d_{ij} between $i\ and\ j$ and maximum considerable distance d_{max}.

$$p(a_{ij,t} = 1) = 1 - \frac{d_{ij}}{d_{max} + 1} \tag{5.8}$$

We simulate the information obtained from the neighbors as the observed action (whether a neighbor has evacuated or not) of the neighbors [56, 255]. We assume that the households living in proximity may not know each other – thus cannot share opinion – but, the households can observe their neighborhood whether they have evacuated or not. The obtained information of a household is the weighted average of the observed action of the neighborhoods (see Equation 5.9).

$$I_{j,t}^{N} = \sum_{i=1}^{n} \frac{b_{ij,t}}{\sum_{i=1}^{n} b_{ij,t}} E_{i,t-1} \tag{5.9}$$

where, $b_{ij,t}$ represents if household i is a neighbor ($b_{ij,t} = 1$) or not ($b_{ij,t} = 0$) based on the threshold distance τ_N.

$$b_{ij,t} = \begin{cases} 0 \ \ if\ d_{ij} > \tau_N \\ 1\ if\ d_{ij} \leq \tau_N \end{cases} \tag{5.10}$$

5.5 Results

In our experiment of risk perception dynamics and evacuation behavior, we run the simulation for 120 hours. The time is discrete in nature; at each simulation step we increase the time by 1 hour. In our simulation, we create an agent's (household's) risk tolerance threshold based on the socio-economic and demographic heterogeneity. We assign the other characteristics

(learning rate, trust on information source, information search behavior, etc.) of the households sampled from a distribution to represent the heterogeneity of the behavior. We present the influence of information sources as three separate scenarios as presented in Table 5.1.

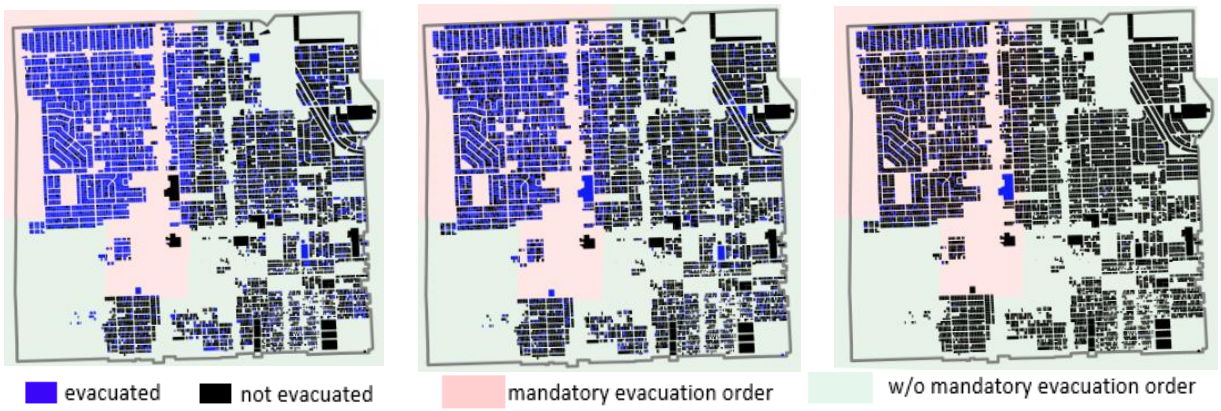
Table 5.2 Parameters of the ABM. Here $N(\mu, \sigma)$ represents that the corresponding value of the parameter is drawn from a normal distribution with mean μ and standard deviation σ.

Parameters		**Scenario 1**	**Scenario 2**	**Scenario 3**
Hazard Risk	τ_{rd} (inch)	22		
	α_j	1	$N(0.5, 0.1)$	$N(0.5, 0.1)$
Social Network	β_j	0	$N(0.5, 0.1)$	0
	d_{max}(kilometer)	3		
	K_{avg}	6		
Neighbor	γ_j	0	0	$N(0.5, 0.1)$
	τ_N (meter)	300		
Landfall Time	C_j	$N(45, 10)$		
	σ_j	$N(20, 10)$		
Update	p_j	$N(0.3, 0.1)$		
Learning Rate	θ_j	$N(0.5, 0.1)$		

In scenario 1, we assume that only hazard risk information is available for the households without any influence from social network and neighbor observation ($\alpha = 1, \beta = 0, \gamma = 0$). In scenario 2, households will have access to hazard risk forecast, $\alpha \in N(0.5, 0.1)$, and social network information, $\beta \in N(0.5, 0.1)$, without any neighbor observation, $\gamma = 0$. Similarly, in scenario 3, households will consider hazard forecast, $\alpha \in N(0.5, 0.1)$, and neighbor observation, $\gamma \in N(0.5, 0.1)$ without considering any information from social network, $\beta = 0$.

Figures 5.3(a), 3(b) and 3(c) show the results of the scenarios in terms of spatial distribution of evacuation compliance and shadow evacuation. Figure 5.3(d) shows the temporal variation of evacuation participation in the simulated scenarios. The temporal dynamics shows that the evacuation started early, and the participation rate is higher for scenario 1 followed by scenario 2. In scenario 3, evacuation participation is significantly low and evacuation started late compare to scenarios 1 and 2. In all three scenarios, we find that more households have evacuated from the mandatory evacuation zones which is consistent with previous studies [55, 60]. Among the three scenarios, scenario 1 has the highest evacuation participation rate (40.29%) including evacuation compliances and shadow evacuation (see Figure 5.3(a)). This result is consistent with previous studies that reliance on weather/forecast related information increases the likelihood of evacuation [55]. However, the reliance on this hazard information also increases the shadow evacuation which is less desirable in terms of managing evacuation traffic [61]. In our model setting, the households having less risk tolerance (or high evacuation tendency) are evacuating despite residing in the low risk zone. In that case, neighbor observation (who lives in proximity), may lower their overall risk perception, as we find from scenario 3 that trust (50% in this case) on neighbor observation significantly lowers the evacuation participation (see Figure 5.3(c)). We have found a decreased (compare to scenarios 1 and 2) evacuation participation for all threshold distances, τ_N between 300 meters to 1000 meters. This effect has also been observed in a previous study [56]. The influence of social network information decreases the overall evacuation participation (both compliance and shadow evacuation) compared to scenario 1 (see Figure 5.3(b)).

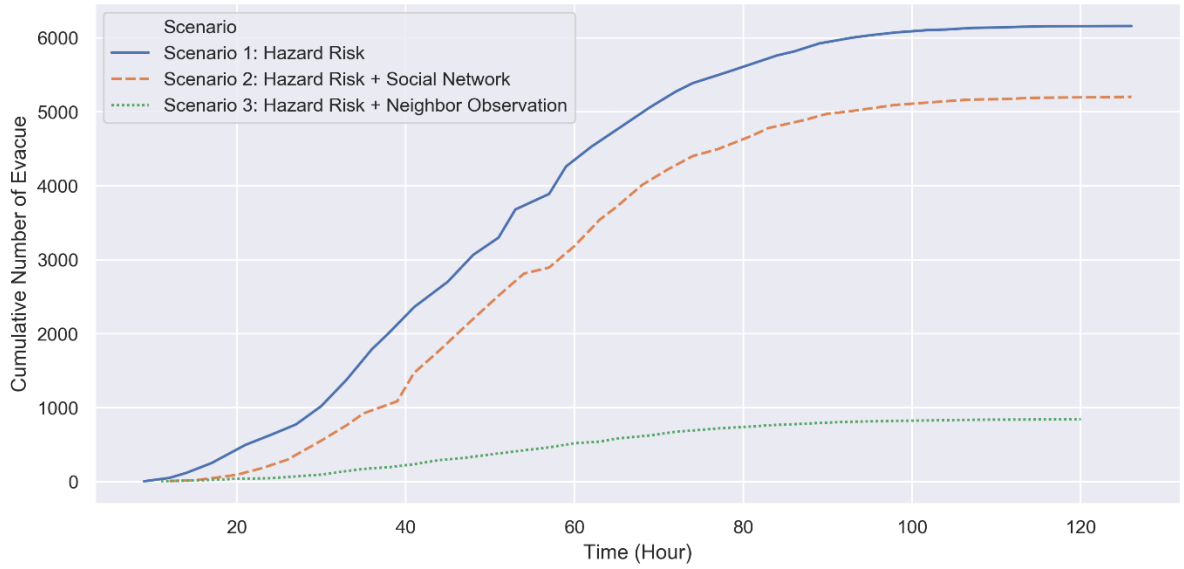
The scenarios shown in Figure 5.3 are the combinations of influences from two information sources at a time where the trust is equally divided between the two information sources. Next, we run



(a) Scenario 1
Compliance 86.97 %
Shadow Evacuation 19.59 %
Overall Evacuation 40.29 %

(b) Scenario 2
Compliance 73.89 %
Shadow Evacuation 16.35 %
Overall Evacuation 34.03 %

(c) Scenario 3
Compliance 16.21 %
Shadow Evacuation 0.78 %
Overall Evacuation 5.52 %

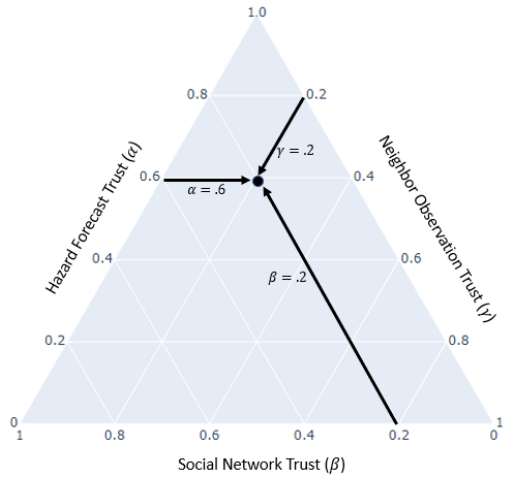


(d) Evacuation Participation over Time

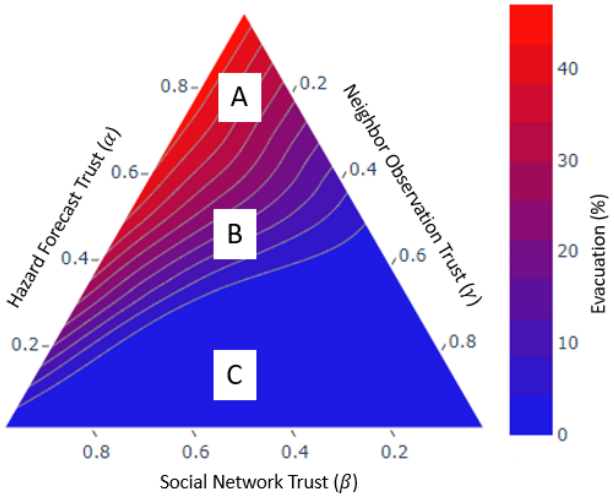
Figure 5.3 Effect of Information Sources on Evacuation Participation on Different Regions. (a) indicates the results of scenario 1, (b) indicates the results of scenario 2, (c) indicates the results of scenario 3, (d) shows the evacuation participation over time.

the simulation for different combinations of trust values of the three information sources and the results are shown in Figure 5.4. We find similar trend of the effects of the information sources on overall evacuation participation, evacuation compliance, and shadow evacuation (see figure 5.4b, 5.4c, 5.4d). The result implies that evacuation participation is likely to be higher when the

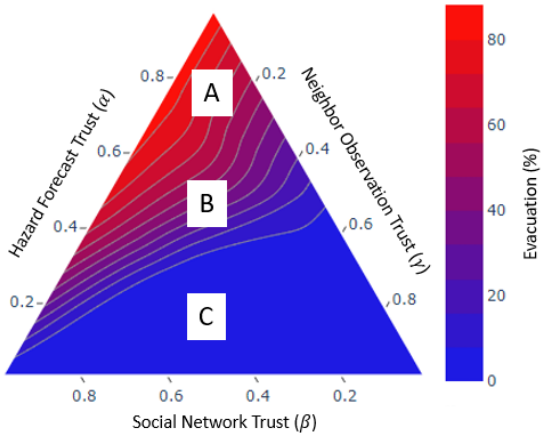
households trust is higher on the hazard forecast. This finding is in line with the findings of a real-world survey conducted in this study [186].



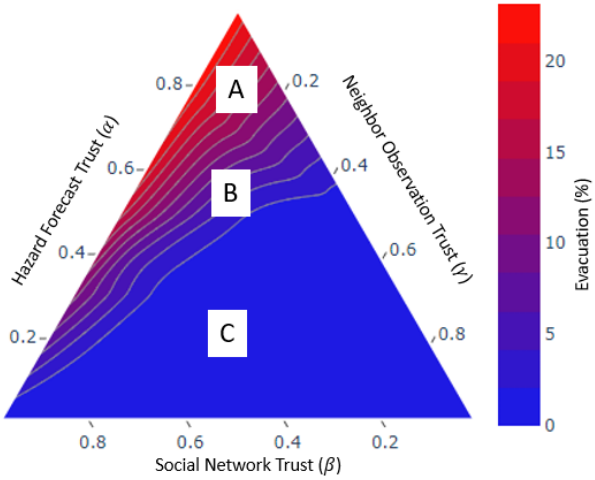
(a) Interpretation of the influence of the information sources



(b) Overall Evacuation



(c) Evacuation Compliance



(d) Shadow Evacuation

Figure 5.4 Combined Effect of Three Information Sources on Evacuation Participations. (a) interpretation of the axes, (b) overall evacuation (c) evacuation compliance or percentage of evacuee from mandatory evacuation zone (d) shadow evacuation, percentage of evacuee from w/o mandatory evacuation zone.

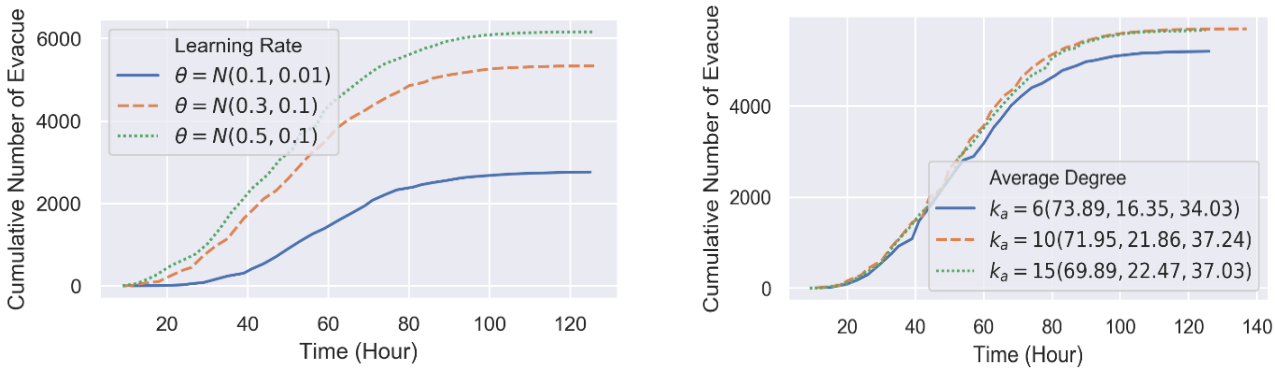
However, our study shows that when evacuation participation increases, it increases in both evacuation compliance and shadow evacuation (see Figure 5.4c and 5.4d). Evacuation participation is almost zero when households have very low (≤ 0.1) trust on hazard forecast. In our simulation we find that the influence of neighbor observation and social network depend on

other sources. For example, for a given trust value of hazard forecast, as the influence of neighbor observation (γ) increases the overall evacuation, compliance, and shadow evacuation decreases (zone A to B to C in figure 5.4). Similarly, for a given trust value of social network as the influence of neighbor observation (γ) increases the overall evacuation, compliance, and shadow evacuation decreases. A very high trust in neighbor's observation is likely to generate a very low evacuation rate/participation (e.g. for $\gamma \geq .4$ see zone C in figure 5.4). On the other hand, for a given trust value of hazard forecast, as the influence of social network (β) increases the overall evacuation, compliance, and shadow evacuation increases (see figure 5.4 from right to left). However, for a given trust value of neighbor's observation, evacuation rate is not same for all zones (zone A, B, and C in figure 5.4). For example, at zone A and above for a given trust value of neighbor's observation the trust value of social network has no effect (or same) in the evacuation rate/participation. But, around zone B, for a given trust value of neighbor's observation, increasing trust in social network increases the evacuation participation up to a certain trust value ($\beta \leq 0.5$), after that increasing social network trust decreases the evacuation participation.

The influence of social network also depends on the network size of a household [55, 245]. From the sensitivity analysis of network size (see Figure 5.5(b)) we find that evacuation participation rate increases with the increase in network size of a household up to a certain size (34% to 37.2 % for an increase in average degree from 6 to 10); after that, increase in network size does not affect evacuation participation (almost same result for network size 10 and 15, see Figure 5.5(b)).

Another effect of network size is that the increase in overall evacuation rate (with increase in network size) mainly happens for shadow evacuation (16.35% to 21.86% for an increase in network size from 6 to 10). Evacuation participation also depends on how a household updates its

risk perception when they get new information (see Figure 5.5a). The quicker a household updates its belief based on new information, the higher the evacuation participation is.



(a) Learning Rate (b) Average Degree

Figure 5.5 Evacuation Participation Rate for different Parameters. (a) shows the cumulative number of evacuees over time for different distribution of learning rate (b) shows the cumulative evacuation participation for different values of average degree of the social networks. The values within the braces shows the evacuation compliance, shadow evacuation, and overall evacuation participation rates, respectively.

In summary, our study on the households living in the zip code 33147 with a hypothetical hazard risk- implies that information sources greatly affect household evacuation participation. Provided hazard risk is forecasted accurately (higher trust on the forecast), the most desirable evacuation scenario is expected if households make their decisions based on the forecasted risk only. If households make their evacuation decisions using information from social networks, overall evacuation and shadow evacuation are likely to increase if they collect information from a big network (e.g. online social networks/social media). On the other hand, if households take evacuation decisions by observing their neighbors, evacuation participations will decrease significantly.

5.6 Limitations and Future Research Directions

Our study has some limitations. We have not considered the other aspects of hazard scenarios such as wind speed, storm surge, etc. The study area is comparatively small. Although we create the households with socio-demographic heterogeneity, some characteristics such as information search behavior, learning rate are randomly assigned. Future studies can collect survey data focusing on these aspects that will allow to assign information search and perception update behavior in a more realistic manner. We use the model parameters that was developed from a previous survey data, which may not be representative for our study area.

We use a simulated static run-off depth to represent the hazard risk of a household. But the hazard forecasts are dynamic and sometimes uncertain. Future studies may consider the dynamics of hazard forecast with the associated uncertainty to model the reliability/trust of the information sources.

We simulate the information collection from social networks based on small-world type random graphs. But the communication pattern in social networks, especially in an online social network, is still not well understood. Designing appropriate survey to collect this information from the affected regions might help find realistic network structure and properties. We considered that more neighbors evacuating will always add to the information favoring an evacuation decision. But some studies have shown that the fear of looting can decrease the evacuation intension after a certain threshold [249]. We have not considered the effect of the possibility of infrastructure disruption in our study. Studies suggest that after landfall evacuation is likely to happen if the critical infrastructure is damaged due to hurricane. Future studies can simulate these effects in a more complicated agent-based systems.

5.7 Conclusions

In this study, we simulate the influence of multiple information sources on households' risk perception dynamics and evacuation behavior. We create realistic households using the findings from existing literature and census data.

Our study shows that reliability on hazard forecast increases evacuation compliance. We also find that increasing trust on social network and neighborhood observation (or decreasing trust on hazard forecast) decreases the overall evacuation participation. The influence of social networks however depends on the network size of the households. While a bigger network increases the overall evacuation, it might increase shadow evacuation.

This study has implications for emergency management practices. Although past studies suggested that effect of social network increases the likelihood to evacuate, our study finds that the increase may happen in shadow evacuation. Thus, it is important to increase the flow of accurate forecast information available to a large extent. Now a days online social media play an important role in propagating information. Emergency organizations should increase their presence in social media with accurate or reliable information on hazard risk.

5.8 Supporting Information

Huq and Abdul-Aziz [256] has developed a process-based hydrologic model for the Florida Southeast Coasts Basin (encompassing Miami through Port St. Lucie) by utilizing the Stormwater Management Model (SWMM) of U.S. Environmental Protection Agency. The SWMM linked climate, land use, and surface and subsurface hydrologic processes in a watershed [257]. We modeled drainage networks as a series of nodes (typically representing large changes in hydraulic head) connected by links (e.g., open channels, pipes). The large basin (7117 km^2) was divided into

small subcatchments based on topography, land uses, hydrography, and urban developments. The land surface hydrologic features (e.g., watershed area, width, slope, imperviousness) were analyzed on an ArcGIS platform using digital elevation models (DEM), hydrography (e.g., National Hydrography Dataset), and land cover (e.g., National Land Cover Database). Operating rules of the existing water control structures in the drainage channels and rivers were also be incorporated into the model. The model ingested water table data as initial conditions and simulated groundwater level in each subbasin. We also incorporated data for basin rainfall and evapotranspiration as inputs, and sea level data as downstream boundary conditions (i.e., water level at the outfalls into bay). The model was successfully calibrated and validated with historical streamflow observations at 6 major rivers and canals during 2004-2013 (R^2 = 0.74-0.92). We ran the model 10, 20, 30, and 40 inch uniform (in time and space) rainfall scenarios over 24 hours to predict the corresponding spatially explicit scenarios of runoff depths in the basin, and extracted the runoff depth information for the Miami-Dade County. Our assumption was that the spatial explicit scenario of runoff depth would closely represent the associated inundation risk.

CHAPTER 6: CONCLUSIONS

In recent times, hurricanes Harvey, Irma, Maria, Michael are some of the costliest hurricanes that have disrupted the lives of millions of people across multiple states in the United States. Emergency evacuation is one of the most effective strategies for decades to reduce risk during natural disasters like hurricane, flooding, wildfire. But individual evacuation decision-making is a complex dynamic process, often studied using post-hurricane survey data. Alternatively, ubiquitous use of social media generates a massive amount of data that can be used to predict evacuation behavior in real time. In addition, real-time social media data can be used to monitor disaster-induced infrastructure disruptions that may reduce response and recovery time. In this dissertation, we present four studies developing methods to make disaster management more dynamic and responsive. These studies also allow us to better understand the socio-infrastructure systems during disasters using social media data and building behavioral models based on agent-based modeling approach. The objectives of this dissertation are as follows:

The first objective of this dissertation is to develop a model using social media data that can capture the dynamics of hurricane evacuation by answering what, when and how users participate in different activities during evacuation. The second objective of this dissertation is to develop a model that can predict evacuation traffic for a longer time horizon (> 1 hour) utilizing real-time data from traffic sensors and social media. The third objective of this dissertation is to develop a method to identify and monitor the co-occurrence of multiple types of infrastructure disruption during disaster from social media data. The final objective of this dissertation is to develop an agent-based model to understand the effect of multiple information sources (social network, neighborhood, weather forecasts, etc.) and its credibility in risk perception dynamics and evacuation decisions.

6.1 Summary of Major Results

This dissertation has shown significant potential of using social media data in improving disaster management especially by contributing to making disaster preparedness, response, and recovery phases of disaster management cycle more dynamic and responsive.

- In the second chapter of this dissertation, we present a method to infer individual evacuation behaviors (e.g., evacuation decision, timing, destination) from social media data. We develop an input output hidden Markov model (IO-HMM) to infer evacuation decisions from user tweets. Using input variables such as evacuation context, time to landfall, type of evacuation order, and the distance from home, the proposed model infers what activities are made by individuals, when they evacuate, and where they evacuate to. Our findings show that the proposed IO-HMM method can be useful for inferring evacuation behavior in real time from social media data. Since traditional surveys are infrequent, costly, and often performed at a post-hurricane period, the proposed approach can be very useful for predicting evacuation demand as a hurricane unfolds in real time.
- In chapter three, we use traffic sensor and Twitter data during hurricanes Matthew and Irma to predict traffic demand during evacuation for a longer forecasting horizon (greater than 1 hour). We present a machine learning approach using Long-Short Term Memory Neural Networks (LSTM-NN), trained over real-world traffic data during hurricane evacuation (hurricanes Irma and Matthew) using different combinations of input features and forecast horizons. Results show that the proposed model can predict traffic demand during evacuation well up to 24 hours ahead. Under hurricane evacuation, efficient traffic operations can maximize the use of transportation infrastructure, reducing evacuation time and stress due to massive congestion. Accurately predicting evacuation traffic is critical to

plan for effective traffic management strategies. Thus, the proposed LSTM-NN model can significantly benefit future evacuation traffic management.

- In chapter four, we present a multi-label classification approach to identify the co-occurrence of multiple types of infrastructure disruptions considering the sentiment towards a disruption—whether a post is reporting an actual disruption (negative), or a disruption in general (neutral), or not affected by a disruption (positive). In addition, we propose a dynamic mapping framework for visualizing infrastructure disruptions. We use a geo-parsing method that extracts location from the texts of a social media post. The proposed multi-label classification approach performs better than a baseline method (using simple keyword search and sentiment analysis). We also find that disruption related tweets, based on specific keywords, do not necessarily indicate an actual disruption. Many tweets represent general conversations, concerns about a potential disruption, and positive emotion for not being affected by any disruption. In addition, a dynamic disruption map has a potential in showing county and point/coordinate level disruptions. Identifying disruption types and their locations are vital for disaster recovery, response, and relief actions. By inferring the co-occurrence of multiple disruptions, the proposed approach may help coordinate among infrastructure service providers and disaster management organizations.
- In chapter five, we develop an agent-based model that integrates flood-related hazard risk, household socio-demographic factors, social network characteristics and decisions, and neighbors' decisions to understand how all these factors interplay in forming risk perception and evacuation decision. We simulate the effect of multiple information sources and their perceived credibility on the evacuation participation by separately exploring the

evacuation compliance and shadow evacuation. Our study indicates that an accurate forecast (higher credibility) of the hazard risk leads to higher evacuation compliance. The effect of a social network is highly dependent on the number of connections. While social networks might increase overall evacuation participation it might also increase shadow evacuation. Putting more trust in neighbor actions induces significantly lower evacuation rates. This study will guide emergency managers to design appropriate strategies for providing hazard forecasts or communicating overall risk during natural disasters like hurricanes and flooding.

6.2 Limitations and Future Research Directions

This dissertation is not without limitations. Some of the limitations are associated with the social media data in general such as difference in penetration rate in different areas, inequal distribution across different age groups, inadequate location precision of the social media data, etc. We summed up the traffic volume of only two highways to estimate the evacuation demand from our study area due to unavailability of enough traffic sensors in all the roads/highways. We do not consider individual's socio-demographic variables in our data-driven models since such data is not present in social media data. Future studies may combine both survey-based and social media data to study evacuation behavior.

In the agent-based model, we do not consider some hazard sources such as wind speed, and storm surges assuming that the study area is not nearby coastal area and small enough to ignore the difference in wind speed perceived by the households. Moreover, in the agent-based simulation we did not consider the effect of infrastructure disruption, post-landfall evacuation, and households' perception to fear of looting. In future studies a more extensive agent-based model can be developed that considers a wide varieties of socio-infrastructure attributes to understand

community resilience where risk perception and evacuation compliance is one of the aspect of resilience.

Despite some limitations, this dissertation adds to the growing literature on understanding our socio-infrastructure systems during a disaster and presents modeling approaches and framework to understand and manage disasters more dynamically and pro-actively.

REFERENCES

1. NOAA National Centers for Environmental Information (NCEI) (2020) U.S. Billion-Dollar Weather and Climate Disasters (2020). https://www.ncdc.noaa.gov/billions/
2. Tousignant Lauren (2017) The cost of natural disasters nearly doubled in 2017. NEWYORK POST
3. Hasan S, Foliente G (2015) Modeling infrastructure system interdependencies and socioeconomic impacts of failure in extreme events: emerging R&D challenges. Nat Hazards 78:2143–2168
4. Re S (2013) Mind the risk: a global ranking of cities under threat from natural disasters. Swiss Re
5. Guha-sapir D, Hoyois P, Below R (2017) Annual Disaster Statistical Review 2016: The numbers and trends. Cent Res Epidemiol Disasters. doi: 10.1093/rof/rfs003
6. Khan H, Vasilescu LG, Khan A, others (2008) Disaster management cycle-a theoretical approach. J Manag Mark 6:43–50
7. FEMA Emergency Management in the United States. https://training.fema.gov/emiweb/downloads/is111_unit 4.pdf
8. Murray-Tuite P, Ge YG, Zobel C, et al (2019) Critical Time, Space, and Decision-Making Agent Considerations in Human-Centered Interdisciplinary Hurricane-Related Research. Risk Anal. doi: 10.1111/risa.13380
9. Baker EJ (1991) Hurricane evacuation behavior. Int J Mass Emerg Disasters 9:287–310
10. Gladwin H, Lazo JK, Morrow BH, et al (2007) Social science research needs for the hurricane forecast and warning system. Nat Hazards Rev 8:87–95
11. Bharosa N, Lee J, Janssen M (2010) Challenges and obstacles in sharing and coordinating

information during multi-agency disaster response: Propositions from field exercises. Inf Syst Front 12:49–65

12. Roy KC, Hasan S, Sadri AM, Cebrian M (2020) Understanding the efficiency of social media based crisis communication during hurricane Sandy. Int J Inf Manage 102060
13. Xiao Y, Huang Q, Wu K (2015) Understanding social media data for disaster management. Nat Hazards 79:1663–1679 . doi: 10.1007/s11069-015-1918-0
14. Wang Q, Taylor JE (2014) Quantifying human mobility perturbation and resilience in hurricane sandy. PLoS One 9:1–5
15. Sadri AM, Hasan S, Ukkusuri S V., Cebrian M (2017) Understanding Information Spreading in Social Media during Hurricane Sandy: User Activity and Network Properties. arXiv Prepr arXiv170603019
16. Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P CM (2015) Performance of Social Network Sensors During Hurricane Sandy. PLoS one 102 e0117288 10:
17. Garg M, Kumar M (2016) Review on event detection techniques in social multimedia. Online Inf Rev 40:347–361
18. Guan X, Chen C (2014) Using social media data to understand and assess disasters. Nat hazards 74:837–850
19. Roy KC, Cebrian M, Hasan S (2019) Quantifying human mobility resilience to extreme events using geo-located social media data. EPJ Data Sci 8:18
20. Martín Y, Li Z, Cutter SL (2017) Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. PLoS One 12:1–22 . doi: 10.1371/journal.pone.0181701
21. Schmidt CW (2012) Trending now: using social media to predict and track disease

outbreaks. Environ Health Perspect 120:a30

22. Latonero M, Shklovski I (2013) Emergency Management, Twitter, and Social Media Evangelism. Using Soc Inf Technol Disaster Cris Manag 3:196–212

23. Ukkusuri S, Zhan X, Sadri A, Ye Q (2014) Use of Social Media Data to Explore Crisis Informatics. Transp Res Rec J Transp Res Board 2459:110–118

24. Sadri AM, Hasan S, Ukkusuri S V, Cebrian M (2020) Exploring network properties of social media interactions and activities during Hurricane Sandy. Transp Res Interdiscip Perspect 6:100143 . doi: https://doi.org/10.1016/j.trip.2020.100143

25. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: LREc. pp 1320–1326

26. Roy KC, Hasan S, Mozumder P (2020) A multilabel classification approach to identify hurricane-induced infrastructure disruptions using social media data. Comput Civ Infrastruct Eng 1–16 . doi: 10.1111/mice.12573

27. Birregah B, Top T, Perez C, et al (2012) Multi-layer crisis mapping: a social media-based approach. In: 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises. pp 379–384

28. Gao H, Barbier G, Goolsby R (2011) Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell Syst 26:10–14

29. Middleton SE, Middleton L, Modafferi S (2013) Real-time crisis mapping of natural disasters using social media. IEEE Intell Syst 29:9–17

30. Kryvasheyeu Y, Chen H, Obradovich N, et al (2016) Rapid assessment of disaster damage using social media activity. Sci Adv 23 e1500779

31. Yuan F, Liu R (2018) Feasibility study of using crowdsourcing to identify critical affected

areas for rapid damage assessment: Hurricane Matthew case study. Int J disaster risk Reduct 28:758–767

32. Deng Q, Liu Y, Zhang H, et al (2016) A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan. Nat Hazards 84:1241–1256
33. Lv Y, Chen Y, Zhang X, et al (2017) Social Media based Transportation Research : the State of the Work and the Networking. 4:19–26
34. Gu Y, Qian ZS, Chen F (2016) From Twitter to detector: Real-time traffic incident detection using social media data. Transp Res part C Emerg Technol 67:321–342
35. Zhang Z, He Q, Gao J, Ni M (2018) A deep learning approach for detecting traffic accidents from social media data. Transp Res Part C Emerg Technol 86:580–596 . doi: 10.1016/j.trc.2017.11.027
36. Kurkcu A, Morgul EF, Ozbay K (2015) Extended implementation method for virtual sensors: web-based real-time transportation data collection and analysis for incident management. Transp Res Rec J Transp Res Board 27–37
37. Zhang S, Tang J, Wang H, Wang Y (2015) Enhancing traffic incident detection by using spatial point pattern analysis on social media. Transp Res Rec J Transp Res Board 69–77
38. Yang Y, Mao L, Metcalf SS (2019) Diffusion of hurricane evacuation behavior through a home-workplace social network: A spatially explicit agent-based simulation model. Comput Environ Urban Syst 74:13–22 . doi: https://doi.org/10.1016/j.compenvurbsys.2018.11.010
39. Madireddy M, Kumara S, Medeiros DJ, Shankar VN (2015) Leveraging social networks for efficient hurricane evacuation. Transp Res Part B Methodol 77:199–212
40. Crooks AT, Wise S (2013) GIS and agent-based models for humanitarian assistance.

Comput Environ Urban Syst 41:100–111

41. Morss RE, Demuth JL, Lazrus H, et al (2018) Hazardous Weather Prediction and Communication in the Modern Information Environment. Bull Am Meteorol Soc 98:2653–2674 . doi: 10.1175/BAMS-D-16-0058.1

42. Watts J, Morss RE, Barton CM, Demuth JL (2019) Conceptualizing and implementing an agent-based model of information flow and decision making during hurricane threats. Environ Model Softw 122:104524

43. Chaniotakis E, Antoniou C, Pereira FC (2017) Enhancing resilience to disasters using social media. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). pp 699–703

44. Kumar D, Ukkusuri S V (2018) Utilizing Geo-tagged Tweets to Understand Evacuation Dynamics during Emergencies: A case study of Hurricane Sandy. In: Companion of the The Web Conference 2018 on The Web Conference 2018. pp 1613–1620

45. Metaxa-Kakavouli D, Maas P, Aldrich DP (2018) How Social Ties Influence Hurricane Evacuation Behavior. Proc ACM Human-Computer Interact 2:122

46. Mitsova D, Escaleras M, Sapat A, et al (2019) The Effects of Infrastructure Service Disruptions and Socio-Economic Vulnerability on Hurricane Recovery. Sustainability 11:516

47. Jongman B, Wagemaker J, Romero B, de Perez E (2015) Early flood detection for rapid humanitarian response: harnessing near real-time satellite and Twitter signals. ISPRS Int J Geo-Information 4:2246–2266

48. NERC (2018) Hurricane Irma Event Analysis Report. www.nerc.com

49. Fan C, Mostafavi A, Gupta A, Zhang C (2018) A system analytics framework for detecting

infrastructure-related topics in disasters using social sensing. In: Workshop of the European Group for Intelligent Computing in Engineering. pp 74–91

50. Huang Q, Xiao Y (2015) Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery. ISPRS Int J Geo-Information 4:1549–1568

51. Li Z, Wang C, Emrich CT, Guo D (2018) A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. Cartogr Geogr Inf Sci 45:97–110

52. Fan C, Mostafavi A (2019) A graph-based method for social sensing of infrastructure disruptions in disasters. Comput Civ Infrastruct Eng

53. Oyeniyi D (2017) How Hurricane Harvey Changed Social Media Disaster Relief. TexasMonthly

54. Yuan F, Liu R (2019) Identifying Damage-Related Social Media Data during Hurricane Matthew: A Machine Learning Approach. Comput Civ Eng 2019

55. Sadri AM, Ukkusuri S V., Gladwin H, et al (2017) The Role of Social Networks and Information Sources on Hurricane Evacuation Decision Making. Nat Hazards Rev 18:04017005 . doi: 10.1061/(ASCE)NH.1527-6996.0000244.

56. Du E, Cai X, Sun Z, Minsker B (2017) Exploring the role of social media and individual behaviors in flood evacuation processes: An agent-based modeling approach. Water Resour Res 53:9164–9180

57. Baker EJ (1979) Predicting response to hurricane warnings: A reanalysis of data from four studies. Mass emergencies 4:9–24

58. Roy KC, Ahmed MA, Hasan S, Sadri AM (2020) Dynamics of Crisis Communications in

Social Media: Spatio-temporal and Text-based Comparative Analyses of Twitter Data from Hurricanes Irma and Michael. In: In Amanda Hughes, Fiona McNeill, & Christopher W. Zobel (Eds.), ISCRAM 2020 Conference Proceedings – 17th International Conference on Information Systems for Crisis Response and Management. Blacksburg, VA (USA): Virginia Tech, pp 812–824

59. Whitehead JC, Edwards B, Van Willigen M, et al (2000) Heading for higher ground: factors affecting real and hypothetical hurricane evacuation behavior. Glob Environ Chang Part B Environ Hazards 2:133–142

60. Hasan S, Ukkusuri S, Gladwin H, Murray-Tuite P (2011) Behavioral Model to Understand Household-Level Hurricane Evacuation Decision Making. J Transp Eng 137:341–348

61. Murray-Tuite P, Wolshon B (2013) Evacuation transportation modeling: An overview of research, development, and practice. Transp Res Part C Emerg Technol 27:25–45

62. Gladwin CH, Gladwin H, Peacock WG (2001) Modeling hurricane evacuation decisions with ethnographic methods. Int J Mass Emerg Disasters 19:117–143

63. Kang JE, Lindell MK, Prater CS (2007) Hurricane Evacuation Expectations and Actual Behavior in Hurricane Lili 1. J Appl Soc Psychol 37:887–903

64. Yang K, Davidson RA, Blanton B, et al (2019) Hurricane evacuations in the face of uncertainty: Use of integrated models to support robust, adaptive, and repeated decision-making. Int J Disaster Risk Reduct 36:101093

65. Hasan S, Mesa-Arango R, Ukkusuri S (2013) A random-parameter hazard-based model to understand household evacuation timing behavior. Transp Res Part C Emerg Technol 27:108–116

66. Lindell MK (2008) EMBLEM2: An empirically based large scale evacuation time estimate

model. Transp Res part A policy Pract 42:140–154

67. Rambha T, Nozick L, Davidson R (2019) Modeling Departure Time Decisions During Hurricanes Using a Dynamic Discrete Choice Framework. In: Transportation Research Board 98th Annual Meeting

68. Xu K, Davidson RA, Nozick LK, et al (2016) Hurricane evacuation demand models with a focus on use for prediction in future events. Transp Res Part A Policy Pract 87:90–101 . doi: 10.1016/j.tra.2016.02.012

69. Mesa-arango R, Hasan S, Ukkusuri S V, et al (2013) Household-Level Model for Hurricane Evacuation Destination Type Choice Using Hurricane Ivan Data. Nat Hazards Rev 14:11–20

70. Sadri AM, Ukkusuri S V., Murray-Tuite P, Gladwin H (2014) Analysis of hurricane evacuee mode choice behavior. Transp Res Part C Emerg Technol 48:37–46

71. Sadri AM, Ukkusuri S V, Murray-Tuite P, Gladwin H (2015) Hurricane evacuation route choice of major bridges in Miami Beach, Florida. Transp Res Rec J Transp Res Board 164–173

72. Lin D-Y, Eluru N, Waller S, Bhat C (2009) Evacuation planning using the integrated system of activity-based modeling and dynamic traffic assignment. Transp Res Rec J Transp Res Board 69–77

73. Ukkusuri S V, Hasan S, Luong B, et al (2017) A-RESCUE: An Agent based regional evacuation simulator coupled with user enriched behavior. Networks Spat Econ 17:197–223

74. Yin W, Murray-Tuite P, Ukkusuri S V, Gladwin H (2014) An agent-based modeling system for travel demand simulation for hurricane evacuation. Transp Res part C Emerg Technol

42:44–59

75. Fu H, Wilmot C (2004) Sequential logit dynamic travel demand model for hurricane evacuation. Transp Res Rec J Transp Res Board 19–26

76. Sarwar MT, Anastasopoulos PC, Ukkusuri S V, et al (2018) A statistical analysis of the dynamics of household hurricane-evacuation decisions. Transportation (Amst) 45:51–70

77. Hasan S, Mesa-Arango R, Ukkusuri S, Murray-Tuite P (2011) Transferability of hurricane evacuation choice model: Joint model estimation combining multiple data sources. J Transp Eng 138:548–556

78. Lee KL, Meyer RJ, Bradlow ET (2009) Analyzing risk response dynamics on the web: The case of Hurricane Katrina. Risk Anal An Int J 29:1779–1792

79. Yabe T, Tsubouchi K, Shimizu T, et al (2019) Predicting Evacuation Decisions using Representations of Individuals' Pre-Disaster Web Search Behavior. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp 2707–2717

80. Meyer RJ, Baker J, Broad K, et al (2014) The dynamics of hurricane risk perception: Real-time evidence from the 2012 Atlantic hurricane season. Bull Am Meteorol Soc 95:1389–1404

81. Davidson RA, Nozick LK, Wachtendorf T, et al (2020) An Integrated Scenario Ensemble-Based Framework for Hurricane Evacuation Modeling: Part 1—Decision Support System. Risk Anal 40:97--116

82. Blanton B, Dresback K, Colle B, et al (2020) An Integrated Scenario Ensemble-Based Framework for Hurricane Evacuation Modeling: Part 2—Hazard Modeling. Risk Anal 40:117–133

83. Roy KC, Hasan S (2019) Modeling the Dynamics of Hurricane Evacuation Decisions from Real-time Twitter Data. In: In Proceedings of Transportation Research Board 98th Annual Meeting

84. Roy K (2018) Understanding crisis communication and mobility resilience during disasters from social media

85. Rahman R, Roy KC, Abdel-Aty M, Hasan S (2019) Sharing Real-time Traffic Information with Travelers using Twitter: An Analysis of Effectiveness and Information Content. Front Built Environ 5:83

86. Kryvasheyeu Y, Chen H (2015) Performance of Social Network Sensors During Hurricane Sandy. PLoS One 10:e0117288

87. Sadri AM, Hasan S, Ukkusuri S V, Cebrian M (2017) Crisis Communication Patterns in Social Media during Hurricane Sandy. Transp Res Rec 0361198118773896

88. Lachlan KA, Spence PR, Lin X, et al (2016) Social media and crisis management: CERC, search strategies, and Twitter content. Comput Human Behav 54:647–652

89. Beiró MG, Panisson A, Tizzoni M, Cattuto C (2016) Predicting human mobility through the assimilation of social media traces into mobility models. EPJ Data Sci 5:

90. Yanjie Duan, Yisheng Lv, Fei-Yue Wang (2016) Travel time prediction with LSTM neural network. 2016 IEEE 19th Int Conf Intell Transp Syst 1053–1058 . doi: 10.1109/ITSC.2016.7795686

91. Pan B, Zheng Y, Wilkie D, Shahabi C (2013) Crowd sensing of traffic anomalies based on human mobility and social media. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp 344–353

92. Dong X, Mavroeidis D, Calabrese F, Frossard P (2015) Multiscale event detection in social

media. Data Min Knowl Discov 29:1374–1405

93. Ghahramani Z, Jordan MI (1996) Factorial hidden Markov models. In: Advances in Neural Information Processing Systems. pp 472–478

94. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77:257–286 . doi: 10.1109/5.18626

95. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. J Mol Biol 305:567–580 . doi: 10.1006/jmbi.2000.4315

96. Eagle N, Pentland A (2006) Reality mining: Sensing complex social systems. Pers Ubiquitous Comput 10:255–268 . doi: 10.1007/s00779-005-0046-3

97. Yin M, Sheehan M, Feygin S, et al (2017) A Generative Model of Urban Activities from Cellular Data. IEEE Trans Intell Transp Syst 1–15 . doi: 10.1109/TITS.2017.2695438

98. Ye J, Zhu Z, Cheng H (2013) What's Your Next Move : User Activity Prediction in Location-based Social Networks. Sdm 171–179 . doi: 10.1137/1.9781611972832.19

99. Hasan S, Ukkusuri S V. (2017) Reconstructing Activity Location Sequences From Incomplete Check-In Data: A Semi-Markov Continuous-Time Bayesian Network Model. IEEE Trans Intell Transp Syst 1–12 . doi: 10.1109/TITS.2017.2700481

100. Duong T V., Bui HH, Phung DQ, Venkatesh S (2005) Activity recognition and abnormality detection with the switching hidden semi-Markov model. Proc - 2005 IEEE Comput Soc Conf Comput Vis Pattern Recognition, CVPR 2005 I:838–845 . doi: 10.1109/CVPR.2005.61

101. Bengio Y, Frasconi P (1995) An Input Output HMM Architecture. Neural Inf Process Syst 427–434 . doi: 10.1093/europace/euq350

102. González AM, Roque AMS, Garc\'\ia-González J (2005) Modeling and forecasting electricity prices with input/output hidden Markov models. IEEE Trans Power Syst 20:13–24

103. Marcel S, Bernier O, Viallet J-E, Collobert D (2000) Hand gesture recognition using input-output hidden markov models. In: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. pp 456–461

104. Balkić Z, Šoštarić D, Horvat G (2012) GeoHash and UUID identifier for multi-agent systems. In: KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications. pp 290–298

105. Cheng G, Wilmot CG, Baker EJ (2008) A destination choice model for hurricane evacuation. In: Proceedings of the 87th Annual Meeting Transportation Research Board, Washington, DC, USA. pp 13–17

106. Han SY, Tsou M-H, Knaap E, et al (2019) How Do Cities Flow in an Emergency? Tracing Human Mobility Patterns during a Natural Disaster with Big Data and Geospatial Data Science. Urban Sci 3:51

107. Wong S, Shaheen S, Walker J (2018) Understanding evacuee behavior: A case study of hurricane Irma

108. Wong SD, Pel AJ, Shaheen SA, Chorus CG (2020) Fleeing from hurricane Irma: Empirical analysis of evacuation behavior using discrete choice theory. Transp Res Part D Transp Environ 79:102227

109. Marshal A (2017) 4 Maps That Show the Gigantic Hurricane Irma Evacuation. In: wired

110. Luz Lazo LA (2017) Airlines scramble and roads fill as residents and visitors rush to get out of Florida ahead of Irma. In: Washington Post

111. FLKEYSNEWS (2017) Millions of Floridians who fled Irma are eager to get home. Patience will be necessary
112. ABC News (2017) Hurricane Irma begins to reach Florida as millions of residents evacuate ahead of monster storm. In: ABC NEWS. https://www.abc.net.au/news/2017-09-10/hurricane-irma-begins-to-impact-florida-as-residents-evacuate/8889076
113. McLachlan G, Krishnan T (2007) The EM algorithm and extensions. John Wiley & Sons
114. D. E. Rumelhart, G. E. Hinton RJW (1986) Learning internal representations by backpropagating errors. Nature 323:
115. Sahlgren M (2008) The distributional hypodissertation. Ital J Disabil Stud 1–18
116. Baroni M, Dinu G (2014) Don ' t count , predict ! A systematic comparison of context-counting vs . context-predicting semantic vectors. Proc 52nd Annu Meet Assoc Comput Linguist 1:
117. Huang A (2008) Similarity measures for text document clustering. Proc Sixth New Zeal 49–56
118. Pham EO, Emrich CT, Li Z, et al (2020) Evacuation Departure Timing during Hurricane Matthew. Weather Clim Soc 12:235–248
119. Sorensen J, Vogt B (2006) Interactive emergency evacuation guidebook. Chem Stock Emerg Prep Program Dep Homel Secur Washington, DC
120. Zeigler DJ, Brunn SD, Johnson Jr JH (1981) Evacuation from a nuclear technological disaster. Geogr Rev 1–16
121. Mikolov T, Corrado G, Chen K, Dean J (2013) Efficient Estimation ofWord Representations in Vector Space Tomas. arXiv Prepr 1–12
122. Mikolov T, Chen K, Corrado G, Dean J (2013) Distributed Representations of Words and

Phrases and their Compositionality. Adv Neural Inf Process Syst 3111--3119

123. Meyer D (2016) How exactly does word2vec work ? 1–18

124. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010. Springer, pp 177–186

125. Lindell MK, Murray-Tuite P, Wolshon B, Baker EJ (2019) Large-Scale Evacuation: The Analysis, Modeling, and Management of Emergency Relocation from Harzardous Areas. CRC Press

126. Hasan S, Mesa-Arango R, Ukkusuri S (2013) A random-parameter hazard-based model to understand household evacuation timing behavior. Transp Res Part C Emerg Technol 27:108–116

127. Huang S-K, Lindell MK, Prater CS (2016) Who leaves and who stays? A review and statistical meta-analysis of hurricane evacuation studies. Environ Behav 48:991–1029

128. Fry J, Binner JM (2015) Elementary modelling and behavioural analysis for emergency evacuations using social media. Eur J Oper Res 249:1014–1023

129. Gudishala R, Wilmot C (2013) Predictive Quality of a Time-Dependent Sequential Logit Evacuation Demand Model. Transp Res Rec J Transp Res Board 2376:38–44 . doi: 10.3141/2376-05

130. Sadri AM, Ukkusuri S V., Murray-Tuite P (2013) A random parameter ordered probit model to understand the mobilization time during hurricane evacuation. Transp Res Part C Emerg Technol 32:21–30

131. Pel AJ, Bliemer MCJ, Hoogendoorn SP (2012) A review on travel behaviour modelling in dynamic traffic simulation models for evacuations. Transportation (Amst) 39:97–123 . doi: 10.1007/s11116-011-9320-6

132. Wilmot CG, Modali N, Chen B (2006) Modeling Hurricane Evacuation Traffic: Testing the Gravity and Intervening Opportunity Models as Models of Destination Choice in Hurricane Evacuation

133. Parady GT, Hato E (2016) Accounting for spatial correlation in tsunami evacuation destination choice: a case study of the Great East Japan Earthquake. Nat Hazards 84:797–807

134. Lindell MK, Arlikatti S, Huang S-K (2019) Immediate behavioral response to the June 17, 2013 flash floods in Uttarakhand, North India. Int J disaster risk Reduct 34:129–146

135. Robinson RM, Foytik P, Jordan C (2017) Review and Analysis of User Inputs to Online Evacuation Modeling Tool. In: Transportation Research Board 96th Annual Meeting. p 9

136. Arlikatti S, Lindell MK, Prater CS, Zhang Y (2006) Risk area accuracy and hurricane evacuation expectations of coastal residents. Environ Behav 38:226–247

137. Baker EJ (1979) Predicting response to hurricane warnings-reanalysis of data from 4 Studies. Mass emergencies 4:9--24

138. Fothergill A (1996) Gender, risk, and disaster. Int J Mass Emerg Disasters 14:33–56

139. Anderson JW (1968) Cultural adaptation to threatened disaster. Hum Organ 27:298–307

140. Quarantelli EL (1990) The Warning Process and Evacuation Behaviour: The Research Evidence. http://udspace.udel.edu/handle/19716/520 1–14

141. Golshani N, Shabanpour R, Mohammadian A, et al (2019) Evacuation decision behavior for no-notice emergency events. Transp Res Part D Transp Environ 77:364–377 . doi: https://doi.org/10.1016/j.trd.2019.01.025

142. Golshani N, Shabanpour R, Mohammadian A, et al (2020) Modeling evacuation demand during no-notice emergency events: Tour formation behavior. Transp Res Part C Emerg

Technol 118:102713 . doi: https://doi.org/10.1016/j.trc.2020.102713

143. Kubisch S, Guth J, Keller S, et al (2020) The contribution of tsunami evacuation analysis to evacuation planning in Chile: Applying a multi-perspective research design. Int J Disaster Risk Reduct 45:101462 . doi: https://doi.org/10.1016/j.ijdrr.2019.101462

144. Wilmot CG, Mei B (2004) Comparison of alternative trip generation models for hurricane evacuation. Nat hazards Rev 5:170–178

145. Gudishala R, Wilmot C (2012) Comparison of Time-Dependent Sequential Logit and Nested Logit for Modeling Hurricane Evacuation Demand. Transp Res Rec 2312:134–140 . doi: 10.3141/2312-14

146. Parr SA, Wolshon B, Murray-Tuite P (2016) Unconventional intersection control strategies for urban evacuation. Transp Res Rec 2599:52–62

147. Litman T (2006) Lessons from Katrina and Rita: What major disasters can teach transportation planners. J Transp Eng 132:11–18 . doi: doi.org/10.1061/(ASCE)0733-947X(2006)132:1(11)

148. Mosher D (2017) The crucial reason Houston officials didn't order evacuations before Harvey made landfall. https://www.mysanantonio.com/technology/businessinsider/article/Why-evacuating-major-cities-before-a-hurricane-12073727.php

149. Seo T, Bayen AM, Kusakabe T, Asakura Y (2017) Traffic state estimation on highway: A comprehensive survey. Annu Rev Control 43:128–151 . doi: 10.1016/j.arcontrol.2017.03.005

150. Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. Transp Res Part C Emerg Technol 79:1–17 . doi: 10.1016/j.trc.2017.02.024

151. Ma X, Tao Z, Wang Y, et al (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp Res Part C Emerg Technol 54:187–197 . doi: 10.1016/j.trc.2015.03.014

152. Oh S, Byon YJ, Jang K, Yeo H (2017) Short-term travel-time prediction on highway: A review on model-based approach. KSCE J Civ Eng 1–13 . doi: 10.1007/s12205-017-0535-8

153. Meng M, Shao C, Wong Y, et al (2015) A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques. J Cent South Univ 22:779–786 . doi: 10.1007/s11771-015-2582-y

154. Barua S, Das A, Roy KC (2015) Estimation of traffic arrival pattern at signalized intersection using ARIMA model. Int J Comput Appl 975:8887

155. Roy K, Barua S, Das A (2015) A Study on Feasible Traffic Operation Alternatives at Signalized Intersection in Dhaka City. In: International Conference on Recent Innovation in Civil Engineering for Sustainable Development; Department of Civil Engineering, DUET: Gazipur, Bangladesh. pp 656–660

156. Smith BL, Demetsky MJ (1997) Traffic flow forecasting: comparison of modeling approaches. J Transp Eng 123:261–266

157. Rahman R, Hasan S (2018) Short-Term Traffic Speed Prediction for Freeways During Hurricane Evacuation: A Deep Learning Approach. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 1291–1296

158. He J, Shen W, Divakaruni P, et al (2013) Improving traffic prediction with tweet semantics. IJCAI Int Jt Conf Artif Intell 1387–1393

159. Ni M, He Q, Gao J (2017) Forecasting the Subway Passenger Flow under Event

Occurrences with Social Media. IEEE Trans Intell Transp Syst 18:1623–1632 . doi: 10.1109/TITS.2016.2611644

160. Graves A, Mohamed A, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. pp 6645–6649

161. Allen-Zhu Z, Li Y, Song Z (2019) On the convergence rate of training recurrent neural networks. In: Advances in Neural Information Processing Systems. pp 1310–1318

162. Hochreiter S, Urgen Schmidhuber J (1997) Long Short-Term Memory. Neural Comput 9:1735–1780 . doi: 10.1162/neco.1997.9.8.1735

163. Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: Continual prediction with LSTM. In: 9th International Conference on Artificial Neural Networks: ICANN '99. IET

164. Geron A (2019) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media

165. Gu Y, Lu W, Qin L, et al (2019) Short-term prediction of lane-level traffic speeds: A fusion deep learning model. Transp Res part C Emerg Technol 106:1–16

166. Lee M-C, Lin J-C (2020) DALC: Distributed Automatic LSTM Customization for Fine-Grained Traffic Speed Prediction. In: Barolli L, Amato F, Moscato F, et al (eds) Advanced Information Networking and Applications. Springer International Publishing, Cham, pp 164–175

167. Bogaerts T, Masegosa AD, Angarita-Zapata JS, et al (2020) A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. Transp Res Part C Emerg Technol 112:62–77 . doi: https://doi.org/10.1016/j.trc.2020.01.010

168. Rahman R, Hasan S (2020) Real-time signal queue length prediction using long short-term

memory neural network. Neural Comput Appl 1–14 . doi: 10.1007/s00521-020-05196-9

169. Beam AL (2017) You can probably use deep learning even if your data isn't that big. https://beamandrew.github.io/deeplearning/2017/06/04/deep_learning_works.html

170. Pasupa K, Sunhem W (2016) A comparison between shallow and deep architecture classifiers on small dataset. In: 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE). pp 1–6

171. Zhang Y, Ling C (2018) A strategy to apply machine learning to small datasets in materials science. Npj Comput Mater 4:1–8

172. Keskar NS, Mudigere D, Nocedal J, et al (2016) On large-batch training for deep learning: Generalization gap and sharp minima. arXiv Prepr arXiv160904836

173. Chollet F, others (2015) Keras. https://keras.io

174. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res 12:2825–2830

175. Altmann A, Toloşi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. Bioinformatics 26:1340–1347

176. Alemazkoor N, Wang S, Meidani H (2018) A Recursive Data-driven Model for Traffic Flow Predictions for Locations with Faulty Sensors. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp 1646–1651

177. Verma S, Vieweg S, Corvey WJ, et al (2011) Natural language processing to the rescue? extracting" situational awareness" tweets during mass emergency. In: Fifth international AAAI conference on weblogs and social media

178. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22:1345–1359

179. Ghafoorian M, Karssemeijer N, Heskes T, et al (2017) Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Sci Rep 7:1–12

180. FCC (2017) Communications Status Report for Areas Impacted by Hurricane Irma

181. HYMAX (2019) The effect of hurricane Irma on water supply. In: krauszusa.com. https://krauszusa.com/effect-hurricane-irma-water-supply

182. (2017) The Effect of Hurricane Irma on Water Supply. KRAUSZCOM

183. Dyskin A V, Basarir H, Doherty J, et al (2018) Computational monitoring in real time: review of methods and applications. Geomech Geophys Geo-Energy Geo-Resources 4:235–271

184. Kadri F, Birregah B, Châtelet E (2014) The impact of natural disasters on critical infrastructures: A domino effect-based study. J Homel Secur Emerg Manag 11:217–241

185. Di Baldassarre G, Kemerink JS, Kooy M, Brandimarte L (2014) Floods and societies: The spatial distribution of water-related disaster risk and its dynamics. Wiley Interdiscip Rev Water 1:133–139

186. Morss RE, Demuth JL, Lazo JK, et al (2016) Understanding public hurricane evacuation decisions and responses to forecast and warning messages. Weather Forecast 31:395–417

187. Homeland Security (2019) Critical Infrastructure Sectors. In: www.dhs.gov. https://www.dhs.gov/cisa/critical-infrastructure-sectors. Accessed 8 Jan 2019

188. Sriram LMK, Ulak MB, Ozguven EE, Arghandeh R (2019) Multi-Network Vulnerability Causal Model for Infrastructure Co-Resilience. IEEE Access 7:35344–35358

189. Fang Y-P, Sansavini G (2019) Optimum post-disruption restoration under uncertainty for enhancing critical infrastructure resilience. Reliab Eng Syst Saf 185:1–11

190. Hoque MM, Roy KC, Shah MI (2014) Safety Investigation and Assessment of High Risk Road Sections in Bangladesh. In: 2nd International Conference on Advances in Civil Engineering 2014 (ICACE-2014)

191. Bryson K-MN, Millar H, Joseph A, Mobolurin A (2002) Using formal MS/OR modeling to support disaster recovery planning. Eur J Oper Res 141:679–688

192. Lambert JH, Patterson CE (2002) Prioritization of schedule dependencies in hurricane recovery of transportation agency. J Infrastruct Syst 8:103–111

193. Sörensen S, Webster JD, Roggman LA (2002) Adult attachment and preparing to provide care for older relatives. Attach Hum Dev 4:84–106

194. Rosenzweig C, Solecki W (2014) Hurricane Sandy and adaptation pathways in New York: Lessons from a first-responder city. Glob Environ Chang 28:395–408

195. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science (80-) 286:509–512

196. Jenelius E, Mattsson L-G (2012) Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study. Transp Res part A policy Pract 46:746–760

197. Sumalee A, Kurauchi F (2006) Network capacity reliability analysis considering traffic regulation after a major disaster. Networks Spat Econ 6:205–219

198. Buldyrev S V, Parshani R, Paul G, et al (2010) Catastrophic cascade of failures in interdependent networks. Nature 464:1025

199. Ulak MB, Kocatepe A, Sriram LMK, et al (2018) Assessment of the hurricane-induced power outages from a demographic, socioeconomic, and transportation perspective. Nat hazards 92:1489–1508

200. Pietrucha-Urbanik K, Tchórzewska-Cieślak B (2018) Approaches to failure risk analysis of the water distribution network with regard to the safety of consumers. Water 10:1679

201. Ouyang M, Fang Y (2017) A mathematical framework to optimize critical infrastructure resilience against intentional attacks. Comput Civ Infrastruct Eng 32:909–929

202. Martani C, Jin Y, Soga K, Scholtes S (2016) Design with uncertainty: the role of future options for infrastructure integration. Comput Civ Infrastruct Eng 31:733–748

203. Alinizzi M, Chen S, Labi S, Kandil A (2018) A Methodology to Account for One-Way Infrastructure Interdependency in Preservation Activity Scheduling. Comput Civ Infrastruct Eng 33:905–925

204. Rinaldi SM, Peerenboom JP, Kelly TK (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Syst Mag 21:11–25

205. Pant R, Thacker S, Hall JW, et al (2018) Critical infrastructure impact assessment due to flood exposure. J Flood Risk Manag 11:22–33

206. Lu L, Wang X, Ouyang Y, et al (2018) Vulnerability of interdependent urban infrastructure networks: Equilibrium after failure propagation and cascading impacts. Comput Civ Infrastruct Eng 33:300–315

207. Chang PC, Flatau A, Liu SC (2003) Health monitoring of civil infrastructure. Struct Heal Monit 2:257–267

208. Rafiei MH, Khushefati WH, Demirboga R, Adeli H (2017) Supervised Deep Restricted Boltzmann Machine for Estimation of Concrete. ACI Mater J 114:

209. Rafiei MH, Adeli H (2017) A novel machine learning-based algorithm to detect damage in high-rise building structures. Struct Des Tall Spec Build 26:e1400

210. Rafiei MH, Adeli H (2018) A novel unsupervised deep learning model for global and local

health condition assessment of structures. Eng Struct 156:598–607

211. Rafiei MH, Adeli H (2018) Novel machine-learning model for estimating construction costs considering economic variables and indexes. J Constr Eng Manag 144:4018106

212. Alavi AH, Buttlar WG (2019) An overview of smartphone technology for citizen-centered, real-time and scalable civil infrastructure monitoring. Futur Gener Comput Syst 93:651–672

213. Cheng Z, Caverlee J, Lee K (2010) You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp 759–768

214. Li C, Sun A (2014) Fine-grained location extraction from tweets with temporal awareness. In: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. pp 43–52

215. Tang Z, Zhang L, Xu F, Vo H (2015) Examining the role of social media in California's drought risk management in 2014. Nat Hazards 79:171–193

216. Keim ME, Noji E (2010) Emergent use of social media : A new age of opportunity for disaster resilience. Am J Disaster Med 6:47–54

217. Sadri AM, Hasan S, Ukkusuri S V (2019) Joint Inference of User Community and Interest Patterns in Social Interaction Networks. Soc Netw Anal Min 9:11

218. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: A review of classification techniques. Emerg Artif Intell Appl Comput Eng 160:3–24

219. Cresci S, Cimino A, Dell'Orletta F, Tesconi M (2015) Crisis mapping during natural disasters via text analysis of social media messages. In: International Conference on Web Information Systems Engineering. pp 250–258

220. Mouzannar H, Rizk Y, Awad M (2018) Damage Identification in Social Media Posts using Multimodal Deep Learning. In: ISCRAM

221. Nguyen DT, Ofli F, Imran M, Mitra P (2017) Damage assessment from social media imagery data during disasters. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. pp 569–576

222. Sorower MS (2010) A literature survey on algorithms for multi-label learning. Oregon State Univ Corvallis 18:1–25

223. Ramos J, others (2003) Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning. pp 133–142

224. Tsoumakas G, Vlahavas I (2007) Random k-labelsets: An ensemble method for multilabel classification. In: European conference on machine learning. pp 406–417

225. Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."

226. Khan A, Baharudin B, Lee LH, Khan K (2010) A review of machine learning algorithms for text-documents classification. J Adv Inf Technol 1:4–20

227. Binkhonain M, Zhao L (2019) A review of machine learning algorithms for identification and classification of non-functional requirements. Expert Syst Appl

228. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830

229. Hutto CJ, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media

230. NOAA (2019) National Hurricane Center Data Archive. In: Natl. Hurric. Cent. Cent. Pacific Hurric. Cent. https://www.nhc.noaa.gov/data/

231. Read J, Bifet A, Holmes G, Pfahringer B (2012) Scalable and efficient multi-label classification for evolving data streams. Mach Learn 88:243–272

232. Nguyen TT, Dang MT, Luong AV, et al (2019) Multi-Label Classification via Incremental Clustering on Evolving Data Stream. Pattern Recognit

233. NOAA National Centers for Environmental Information (NCEI) U.S. (2018) Billion-Dollar Weather and Climate Disasters. https://www.ncdc.noaa.gov/billions/. Accessed 15 Jan 2018

234. Ahmadlou M, Adeli H (2010) Enhanced probabilistic neural network with local decision circles: A robust classifier. Integr Comput Aided Eng 17:197–210

235. Rafiei MH, Adeli H (2017) A new neural dynamic classification algorithm. IEEE Trans neural networks Learn Syst 28:3074–3083

236. Cerri R, Basgalupp MP, Barros RC, de Carvalho AC (2019) Inducing Hierarchical Multi-label Classification rules with Genetic Algorithms. Appl Soft Comput 77:584–604

237. Wehrmann J, Cerri R, Barros R (2018) Hierarchical multi-label classification networks. In: International Conference on Machine Learning. pp 5225–5234

238. Bostrom A, Morss R, Lazo JK, et al (2018) Eyeing the storm: How residents of coastal Florida see hurricane forecasts and warnings. Int J disaster risk Reduct 30:105–119

239. Carlson JM, Alderson DL, Stromberg SP, et al (2014) Measuring and modeling behavioral decision dynamics in collective evacuation. PLoS One 9:

240. Lindell MK, Lu J-C, Prater CS (2005) Household decision making and evacuation in response to Hurricane Lili. Nat hazards Rev 6:171–179

241. Grimm V, Railsback SF (2005) Individual-based modeling and ecology. Princeton university press

242. DeYoung SE, Wachtendorf T, Farmer AK, Penta SC (2016) NOAA radios and

neighbourhood networks: demographic factors for channel preference for hurricane evacuation information. J contingencies Cris Manag 24:275–285

243. Xu D, Zhou W, Deng X, et al (2020) Information credibility, disaster risk perception and evacuation willingness of rural households in China. Nat Hazards. doi: 10.1007/s11069-020-04106-5

244. Hasan S, Ukkusuri S V (2011) A threshold model of social contagion process for evacuation decision making. Transp Res Part B Methodol 45:1590–1605 . doi: https://doi.org/10.1016/j.trb.2011.07.008

245. Widener MJ, Horner MW, Metcalf SS (2013) Simulating the effects of social networks on a population's hurricane evacuation participation. J Geogr Syst 15:193–209 . doi: 10.1007/s10109-012-0170-3

246. Miami-Dade County Open data hub. https://gis-mdc.opendata.arcgis.com/

247. United States Census Bureau. https://data.census.gov/cedsci/

248. Zhu Y, Xie K, Ozbay K, Yang H (2018) Hurricane evacuation modeling using behavior models and scenario-driven agent-based simulations. Procedia Comput Sci 130:836–843

249. Halim N, Kuhlman CJ, Marathe A, et al (2020) Two-Mode Threshold Graph Dynamical Systems for Modeling Evacuation Decision-Making During Disaster Events. In: Cherifi H, Gaito S, Mendes JF, et al (eds) Complex Networks and Their Applications VIII. Springer International Publishing, Cham, pp 519–531

250. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world'networks. Nature 393:440–442

251. Lorenz J (2005) A stabilization theorem for dynamics of continuous opinions. Phys A Stat Mech its Appl 355:217–223

252. Watts DJ (2002) A simple model of global cascades on random networks. Proc Natl Acad Sci 99:5766–5771

253. Widrow B, Hoff ME (1988) Adaptive switching circuits, Neurocomputing: foundations of research

254. Ghaderi J, Srikant R (2014) Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. Automatica 50:3209–3215

255. Centola D (2010) The spread of behavior in an online social network experiment. Science (80-) 329:1194–1197

256. Haque E, Abdul-Aziz OI (2020) Large-scale modeling and assessment of climate and land cover change impacts on the stormwater runoff in complex coastal-urban environments. J Hydrol under revi:

257. Rossman LA, others (2010) Storm water management model user's manual, version 5.0. National Risk Management Research Laboratory, Office of Research and~…