Rapid Assessment of Disaster Impacts on Highways Using Social Media

Yudi Chen, S.M.ASCE[1]; Qi Wang, A.M.ASCE[2]; and Wenying Ji, A.M.ASCE[3]

Abstract: A timely and reliable assessment of disaster impacts on highways is desired for executing evacuations, providing emergency services, and planning relief and recovery activities in disaster management. Although social media—a near-real-time and human-centered information source—has been increasingly investigated to assess disaster impacts on communities, the employment of social media for assessing disaster impacts on highways is still less explored due to the lack of a systematic and reliable approach. To fill this gap, this research presents a social media–based approach to assess disaster impacts on highways. Two social media–based indicators, namely, the number of impact-related tweets (NIT) and the geolocation of impact-related tweets (GIT), are designed to indicate the severity and location of disaster impacts on highways, respectively. Disaster impacts on highways during Hurricane Harvey in Houston were studied to demonstrate the developed approach. The feasibility and applicability of the research outcomes are validated through the highway high-water incidents data collected by the Texas Department of Transportation (TxDOT). **DOI: 10.1061/(ASCE)ME.1943-5479.0000836.**

Author keywords: Disaster management; Highway infrastructure; Social media mining.

Introduction

Highways are of vital importance in various aspects of disaster management, such as evacuation planning (Chang 2003), emergency service provision (Litman 2006), and relief and recovery resource distribution (Faturechi and Miller-Hooks 2015), which assist in mitigating and reducing the disastrous impacts on communities. However, as per the 2017 ASCE infrastructure report card, highways are graded as D, representing poor to fair conditions (ASCE 2017), which makes highways vulnerable to massive disruptions caused by natural disasters (e.g., hurricanes, floods, and earthquakes) (Ha et al. 2017). Due to the importance and vulnerability of highways, a timely and reliable assessment of disaster impacts on highways (indicated by interrupted traffic flow) is needed for effective route planning and enhanced highway recovery. In practice, two perspectives of disaster impacts on highways, severity and location, are assessed through damage surveys and/or site inspections operated by the FHWA and state-level DOTs (Kirk 2012). However, during the strikes of large-scale natural disasters, these assessment processes are time consuming, labor intensive, and unsafe, which leads to delayed information collection [e.g., delayed site inspections due to blocked roads (Rau et al. 2007)], inefficient resource allocation [e.g., lack of inspection professionals (FHWA 2013)], and unsafe working environments [e.g., potential casualty (OSHA 2018)], respectively.

With the advances of information technologies, social media platforms (e.g., Twitter, Facebook, and Instagram) have been broadly utilized for crowdsourcing near-real-time emergency information and public opinions during disasters (Imran et al. 2018). Researchers have claimed the feasibility of measuring disaster impacts through social media due to its capability of reflecting the digital traces of a disaster. The potential of using social media for rapid flood mapping has been demonstrated through the 2015 South Carolina floods (Li et al. 2018b). Disaster-related Twitter activities have been shown to be highly correlated with physical disaster damage (in monetary value) reported by FEMA for a variety of disasters (e.g., hurricane, earthquake, and tornado) (Kryvasheyeu et al. 2016). Similarly, the geographical distributions of damage-related Twitter activities during Hurricane Matthew are consistent with those derived from the postdisaster insurance claims (Yuan and Liu 2018, 2020). Such presented research outcomes, together with the near-real-time and human-centered nature of social media, make it promising for achieving a timely and reliable assessment of disaster impacts on highways. In spite of this intriguing idea, the use of social media for assessing disaster impacts on highways has yet been systematically examined.

This study aims to develop a systematic approach to timely and reliably assess disaster impacts on highways using social media data. Disaster impacts from Hurricane Harvey on the highways in Houston were studied for this purpose. The developed approach is capable of (1) reliably extracting disaster impact–related data from social media by developing a systematic data adaption module; (2) quantitatively assessing the severity of disaster impact (SDI) via the number of impact-related tweets (NIT); and (3) geographically indicating the location of disaster impacts (LDI) through the geolocation of impact-related tweets (GIT). The content of this study is organized as follows. First, previous research studies focusing on social media data processing in disaster

[1]Ph.D. Student, Dept. of Civil, Environmental and Infrastructure Engineering, Univ. of George Mason, 4603 Nguyen Engineering Bldg., 4400 University Dr., MS 6C1, Fairfax, VA 22030. Email: ychen55@gmu.edu

[2]Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Northeastern, 433 SN, 360 Huntington Ave., Boston, MA 02115. Email: q.wang@northeastern.edu

[3]Assistant Professor, Dept. of Civil, Environmental and Infrastructure Engineering, Univ. of George Mason, 1411 Nguyen Engineering Bldg., 4400 University Dr., MS 6C1, Fairfax, VA 22030 (corresponding author). ORCID: https://orcid.org/0000-0002-1222-2191. Email: wji2@gmu.edu

Note. This manuscript was submitted on November 26, 2019; approved on May 1, 2020; published online on June 30, 2020. Discussion period open until November 30, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Management in Engineering***, © ASCE, ISSN 0742-597X.

management are introduced. Second, the methodology is composed of three modules, namely, data collection, data adaption, and impact assessment. Third, the feasibility of the proposed approach for assessing disaster impacts on highways is validated using the highway high-water incidents data collected by the Texas Department of Transportation (TxDOT). Fourth, the advantages and potential applications of the proposed approach are discussed. Finally, research contributions, limitations, and future work are concluded.

Previous Research

Social media has become increasingly popular in obtaining enhanced awareness of disaster situations (Imran et al. 2018). During a disaster, individuals and organizations are highly active in social media for distributing emergency information and sharing their opinions and sentiments (Chen et al. 2019; Kim et al. 2018; Kim and Hastak 2018), which has inspired researchers to examine disaster damage through analyzing various information forms of social media [e.g., text (Fan et al. 2020), image (Ham and Kim 2020), and video (Tian et al. 2019)]. However, the potential of assessing disaster impacts on highways through social media is still less investigated, and therefore requires further research efforts.

Obtaining a specific assessment of disaster impacts using social media is challenging mainly due to the difficulty of identifying specific information of interest from massive social media data (Imran et al. 2018). To address this challenge, supervised classification models [e.g., support vector machine (Ragini and Anand 2017), random forest (Li et al. 2018a), and convolutional neural network (Yu et al. 2019)] have been applied to filter out irrelevant information and to categorize relevant information. However, high-quality reference samples are required to train the classification models, which limits the implementation of this type of technique in disaster management (Imran et al. 2018). Therefore, a systematic data adaption module capable of avoiding labeling issues is desired to extract relevant social media data.

Methodology

In this research, a systematic approach was designed to assess disaster impacts on highways using social media, as shown in Fig. 1. The developed approach comprises three modules: data collection, data adaption, and impact assessment. In the data collection module, social media data are collected based on disaster information, and a highway lexicon is defined as per highway names (i.e., official designations and nicknames). In the data adaption module, data related to disaster impacts on highways are adapted from social media through three processes, i.e., data cleaning, data mapping, and data filtering. The data cleaning process aims to transform each tweet into a set of representative terms. Then, the data mapping process maps the cleaned tweets to their corresponding highways by designing a set of domain knowledge–based rules. Finally, the data filtering process identifies tweets specific to disaster impacts on highways using a set of impact keywords determined by manual iterations. The purpose of having the data mapping process prior to the data filtering process is to reduce the manual efforts involved in determining impact keywords in the data filtering process. In the impact assessment module, the adapted data are employed to assess disaster impacts on highways from perspectives of severity and location, both of which are crucial for practitioners to plan disaster relief and recovery activities.

Disaster impacts on highways from Hurricane Harvey were studied to illustrate the developed approach. Hurricane Harvey inflicted $125 billion in damage, primarily from catastrophic rainfall-triggered flooding in the Houston metropolitan area (hereafter referred to as Houston for simplicity). Most areas in Houston observed more than 50.8 cm (20 in.) of precipitation, which prompted high-water incidents for almost all major highways (Blake and Zelinsky 2018). These widespread and devastating floods caused unsafe living conditions and impeded or even blocked traffic. As a consequence, the efficiency of distributing disaster relief and recovery resources was significantly reduced due to the malfunction of the highway system, which in turn led to more disastrous impacts on communities (e.g., increased casualties and extended infrastructure restoration) (Davies 2017). Through the developed approach, rapid highway situation awareness is developed to facilitate the selection of functional routes, which is beneficial for delivering valuable resources (e.g., life-saving supplies and utility professionals) to disaster-affected communities in a timely manner.

Data Collection

Social media data and highway lexicon are prepared in the data collection module. More specifically, social media data provide raw information that is further processed for assessing disaster impacts on highways, and the highway lexicon is used in the process of identifying highway-related social media data.

Social Media Data

With the prevalence of mobile communication devices, an increasing number of individuals and organizations are using social media to share emergency information and opinions during a disaster (Kirilenko and Stepchenkova 2014). This boosting trend makes social media an ideal candidate to collect social responses that can be further employed to evaluate disaster impacts. In this research, Twitter was used to collect social media data due to its popularity in disaster management for sharing emergency information as well as its effortless accessibility for collecting large-scale data sets

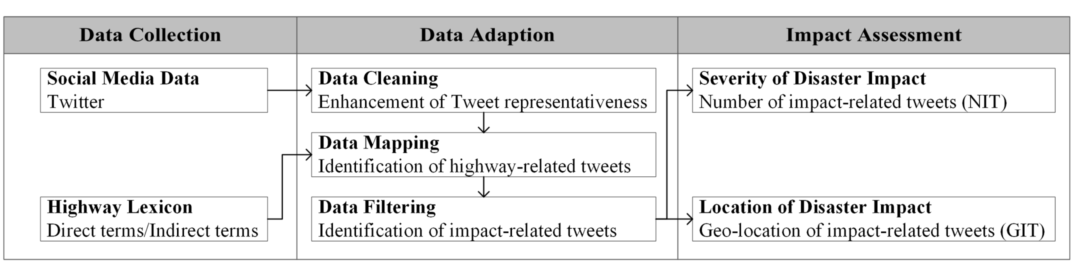

Fig. 1. Research methodology.

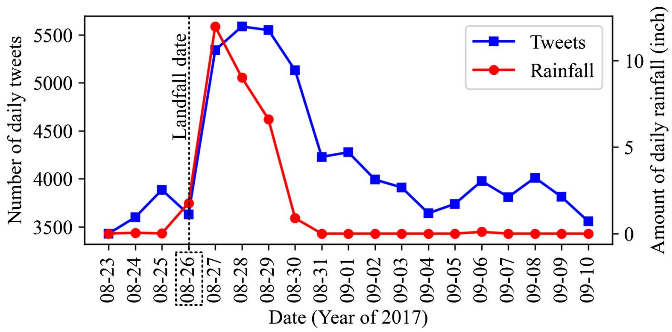

Fig. 2. Temporal variation of Twitter activities and rainfall.

(Kryvasheyeu et al. 2016). Tweepy, a Python package for implementing the Twitter streaming application programming interface (API), was used to collect general geotagged tweets through the previously developed module (Wang and Taylor 2016). Two types of constraints, time span and location bounding box, were employed as filters to ensure that tweets used in this study were posted during the period of Hurricane Harvey in Houston. The time span was from August 23, 2017, to September 10, 2017, which covers 3 days before the start of the unprecedented precipitation plus 2 weeks after the peak of precipitation. The location bounding box was comprised of latitudes and longitudes of the boundaries for the area of interest. In this case, most areas of Houston, with corresponding latitudes [29.2°N, 30.2°N] and longitudes [96.0°W, 94.8°W], were covered. In this research, retweet activities were removed due to their duplicated information. Eventually, 87,365 tweets posted by 12,865 unique users were collected.

Fig. 2 depicts the temporal variations of the collected Twitter activities (represented by the number of daily tweets) and the amount of daily rainfall during Hurricane Harvey. The number of daily tweets increases substantially from August 26 (the landfall date of Hurricane Harvey) to August 27, 2017, due to the catastrophic amount of rainfall brought by Harvey. Compared with the sharp decrease in the amount of daily rainfall, the number of daily tweets maintains a high level until August 30, 2017, due to the continuous flooding impact. To ensure a representative data size for assessing disaster impacts on highways, Twitter activities from the peak period (i.e., from August 27 to 30, 2017) were examined.

Highway Lexicon

In social media, a highway is typically described by a set of specific terms to avoid ambiguity and for simplicity. For instance, in the tweet "This is the flooding I-10 Katy Fwy #traffic," Interstate Highway 10 is indicated by the term "I-10." Exploiting this specification, a comprehensive highway lexicon was defined to provide various search terms to map the tweets to each highway in the data mapping process. The highway lexicon was built based on the official designations of the highways as well as their nicknames and abbreviations to cater to the informal expressions in social media. Ten highways were selected based on the annual average daily traffic (AADT) to ensure they are potentially with high levels of social media activities as depicted in Fig. 3. Their AADTs were extracted from the transportation planning map of Texas in the year of 2017 (TxDOT 2017) as shown in Table 1.

To construct the highway lexicon, two types of search terms—direct terms and indirect terms—were defined, as listed in Table 2. The direct terms are the official designations of highways and their variants or nicknames. For example, "IH-45" is the designation of the Interstate Highway 45, while "i45" and "i-45" are the variants

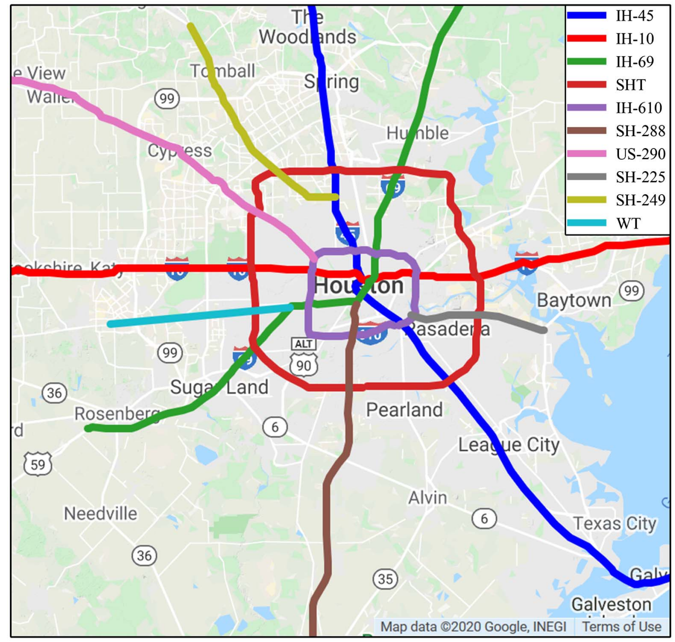

Fig. 3. Highways of interest in Houston. (Map data © 2020 Google, INEGI.)

Table 1. AADTs of the selected highways

Highway	AADT
IH-45	2.80×10^5
IH-10	2.60×10^5
IH-69	2.60×10^5
IH-610	2.90×10^5
SHT	2.00×10^5
SH-288	1.60×10^5
US-290	2.40×10^5
SH-225	1.10×10^5
SH-249	1.50×10^5
WT	1.40×10^5

Note: IH = Interstate Highway; SHT = Sam Houston Tollway; US = US Highway; and WT = Westpark Tollway.

Table 2. Search terms for the selected highways

Highway designation	Direct term	Indirect term
IH-45	IH-45, I-45, I45	45, gulf, north
IH-10	IH-10, I-10, I10	10, katy, baytown-east
IH-69	IH-69, I-69, i69, US59, US-59	69, 59, eastex, southwest
IH-610	IH-610, I-610, I610, the loop	610
SHT	beltway8, belt8, beltway 8	Sam Houston
SH-288	SH-288, SH288	288, south
US-290	US-290, US290	290, northwest
SH-225	SH-225, SH225	225, la Porte
SH-249	SH-249, SH249	249, Tomball
WT	Westpark	Westpark

Note: IH = Interstate Highway; SHT = Sam Houston Tollway; US = US Highway; and WT = Westpark Tollway.

of the designation. Also, "beltway8" is the nickname for the Sam Houston Tollway (SHT). Regarding the indirect terms, they are simply composed of a route number (e.g., "45," "10," and "610") or a part of the highway names (e.g., "Sam Houston," "Tomball," and "Westpark"). These search terms were utilized in the data mapping process to extract highway-related tweets from the originally collected tweets.

Data Adaption

In practice, transforming collected raw data into interpretable data is necessary for accurately mining knowledge because the raw data are usually too noisy and overwhelming to be utilized (Bramer 2007). A data adaption module, which is capable of adapting raw data into a tidy format, is particularly important for utilizing large-scale social media data (Immonen et al. 2015). In the present study, the collected raw tweets were adapted to each highway to assess disaster impacts through three processes, namely, data cleaning, data mapping, and data filtering.

Data Cleaning

Data cleaning, a process of detecting and removing errors and inconsistencies, is needed for mining social media data that usually contain excessive amounts of informal expressions (e.g., abbreviations and slang) and noisy information (e.g., erroneous spells and invalid symbols) (Cai and Zhu 2015). In this study, the decisive information of each tweet is the text content related to disaster impacts on highways. Five cleaning steps were employed to enhance the representativeness of tweet text content. Details are illustrated in a step-by-step pseudocode form as follows:

Input: Raw tweets
1. Remove all URL links and retain the text content from a tweet
2. Tokenize a tweet into separate terms based on a set of delimiters
3. Lemmatize each term into its root, such as "flooded" to "flood"
4. Remove invalid symbols through regular expressions
5. Remove stopping terms (e.g., is, of, and often)

Output: Cleaned tweets

With the cleaning process, each tweet is represented by a set of enhanced terms to ensure the quality of operations detailed in the data mapping and data filtering processes.

Data Mapping

To describe the disaster impacts on highways using social media activities, cleaned tweets needed to be mapped to each highway. In this process, a set of rules was designed to map these cleaned tweets to their corresponding highways. To be specific, a tweet containing one of the direct search terms for a specific highway was considered to be related to this specific highway, while a tweet containing one of the indirect search terms only indicated there was a high possibility of relatedness and required further investigation. For example, although the tweet "It's like 45 songs that isn't R&B RT @Jdxthompson: 2000's R&B was the best" contains "45," which is one of the indirect search terms for IH-45, it is apparently not related to IH-45. To solve this issue, a set of highway functional classification terms, which is usually used together with the indirect terms to represent a highway, was employed to facilitate further investigations. These terms are "highway," "hwy," "freeway," "fwy," "tollway," "tlwy," "parkway," "pwy," "loop," and "lp." If only one of the defined functional classification terms was detected in the neighbors of indirect terms, the tweet was deemed to be related to this highway. Here, the detailed data mapping process is illustrated in a step-by-step pseudocode form as follows:

Input: Cleaned tweets, highway lexicon, and highway functional classification terms (HFCT)
1. **Repeat**
2. Select a tweet from the cleaned tweets
3. **Repeat**
4. Select a highway from the highways of interest in Houston
5. Load corresponding direct terms and indirect terms from the highway lexicon
6. **If** there is an intersection between the tweet and the direct terms
7. The tweet is related to the highway
8. **Else if** there is an intersection between the tweet and the indirect terms
9. Name the intersection as the indirect intersection
10. Find the neighbors of the indirect intersection from the tweet
11. **If** there is an intersection between neighbors and HFCT
12. The tweet is related to the highway
13. **End If**
14. **End If**
15. **Until** all highways of interest are selected
16. **Until** all cleaned tweets are mapped

Output: Highway-related tweets

Data Filtering

From the previous two steps, the mapped highway-related tweets comprise information related to disaster impacts and other types of information (e.g., normal traffic) as well. To reliably assess disaster impacts using only impact-related information, tweets capable of indicating disaster impacts on highways need to be further sorted out from the mapped highway-related tweets. Here, keyword search was employed to identify impact-related tweets due to its effectiveness and simplicity (Boyd et al. 2010). The impact keywords were initialized using "hurricane," "Harvey," and "flood," which are straightforward and intuitive for indicating disaster impacts. Tweets containing one or more of the impact keywords were categorized as impact-related tweets. The filtering process is iteratively conducted in a step-by-step pseudocode form as follows:

Input: Highway-related tweets and impact keywords
1. **Repeat**
2. Categorize the tweets into impact- or nonimpact-related based on impact keywords
3. Obtain the top terms in nonimpact-related tweets
4. **If** there are top terms obviously related to disaster impacts
5. Update the impact keywords with the newly identified top terms
6. **End If**
7. **Until** none of the top terms is obviously related to disaster impacts

Output: Impact-related tweets

In this study, the relevance of a top term to disaster impacts was manually determined. The number of analyzed top terms in each iteration was empirically set as 20 to balance the efficiency and filtering accuracy of this determination process. Finally, "rain," "water," and "river" were determined to supplement the original impact keywords. Notably, the identified hazard-specific words used in the filtering process were general words (e.g., "water," "rain," and "river"), which can be directly applied to identify impact-related tweets for a future hurricane with a similar disaster impact (i.e., flooding). Through this iteration process, most impact-related tweets were captured. These impact-related

Table 3. Numbers of highway- and impact-related tweets for the selected highways

Highway	N_H (through the data mapping process)	N_I (through the data filtering process)
IH-45	467	143
IH-10	408	199
IH-69	402	107
IH-610	478	143
SHT	589	217
SH-288	240	82
US-290	171	22
SH-225	89	44
SH-249	82	15
WT	98	44
Total	3,024	1,016

Note: N_H and N_I represent the numbers of highway-related and impact-related tweets, respectively; IH = Interstate Highway; SHT = Sam Houston Tollway; US = US Highway; and WT = Westpark Tollway.

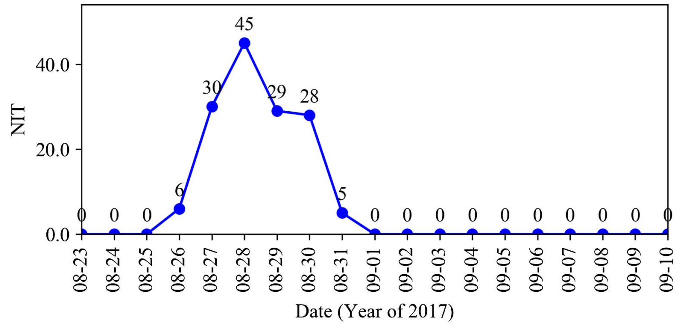

Fig. 4. Temporal variations of NIT for IH-45.

tweets represent true information related to disaster impacts on highways.

The numbers of highway-related and impact-related tweets are listed in Table 3. Through the specialized data mapping process, tweet size reduced significantly from the original 87,365 to 3,024, which indicates that only a small portion of the original tweets are related to the selected 10 highways. Still, the highways were frequently mentioned on social media during Hurricane Harvey as there were large numbers of highway-related tweets for each highway. Among the 3,024 highway-related tweets, 1,016 tweets describe disaster impacts on highways.

The performance of this systematic data adaption module is demonstrated through the following validation steps. First, 1,000 tweets were selected from the originally collected tweets in a random manner to avoid bias. Second, an English native speaker was invited to manually identify impact-related tweets that were taken as ground truth. Third, the manually identified impact-related tweets and the tweets extracted using the developed data adaption module were compared through two measures, namely, precision and recall. In detail, precision represents the fraction of correctly extracted impact-related tweets among all the extracted impact-related tweets using the data adaption module, while recall represents the fraction of the correctly extracted impact-related tweets over the manually identified impact-related tweets. In this validation, the precision is 80% and the recall is 92%. The high values of these two measures demonstrate the capability of the data adaption module for extracting tweets related to disaster impacts on highways, which in turn demonstrate the uniqueness and completeness of the search terms and keywords employed in the data mapping and filtering processes.

Impact Assessment

In this section, disaster impacts on highways are assessed from the perspectives of severity and location to generate actionable insights for practitioners to perform effective planning in disaster management. Details of these two perspectives are discussed as follows.

Severity of Disaster Impact

During Hurricane Harvey, highways in Houston were severely impacted by floods, wherein high-water incidents on highways impeded or even blocked the traffic. Measuring the SDI can enhance relief resource allocation, and therefore reduces disaster impacts on communities (Orabi et al. 2010). For example, FHWA would benefit from such an index to distribute an appropriate amount of funding and experts to disaster-impacted roads (FHWA 2013).

SDI is defined to reveal the interrupted traffic flow through incorporating the factors of blockage duration and road usage. The blockage duration is represented by the duration of a highway high-water incident recorded by TxDOT, and the road usage is represented by the AADT. To quantify the interrupted public traffic, the SDI is formulated through the multiplication of high-water incident duration $D_{\mathrm{high-water}}$ and AADT as Eq. (1)

$$\mathrm{SDI} = \frac{D_{\mathrm{high-water}} \times \mathrm{AADT}}{1{,}440} \tag{1}$$

The denominator, 1,440, transforms the time resolution of AADT from day into minutes to match the time resolution of high-water incident duration. Ultimately, the severity of disaster impact on a highway is computed through aggregating disaster impacts on all affected segments that are associated with the specific highway. The SDI is utilized to validate the derived social media–based metric in the validation section.

To achieve a timely measurement of SDI, we derived a social media-based metric, NIT. For illustration purposes, the temporal variations of NIT for IH-45, which connects Houston, Dallas, and the Gulf of Mexico for delivering essential resources, are depicted in Fig. 4. The larger the NIT, the more severe the disaster impact, and vice versa. Through this derived social media–based metric (i.e., NIT), the severity of disaster impact is measured in a daily manner to assist practitioners to plan relief and recovery efforts in the dynamic environment of a disaster.

Location of Disaster Impact

A timely and reliable indication of the LDI is needed to distribute valuable yet limited relief and recovery resources to proper highway segments, as well as to avoid impacted segments for safe and effective planning (Kim et al. 2017). The continuous trajectory of a highway is segmented via a set of blocks to represent the highway discrete locations. Fig. 5 shows the segmentation of IH-45 as an example. Each block spans 0.05° of latitude (i.e., approximately 5.56 km) and 0.05° of longitude (i.e., approximately 4.81 km) for properly balancing the number and the geo-granularity of highway locations. The exact coordinates of each location (marked by a dot) are calculated by averaging the highway trajectory (marked by a line) that is within the location-associated block.

With this discrete representation of a highway, the distance d between the geolocation of each impact-related tweet GIT and the discrete highway location HL is calculated using Eq. (2)

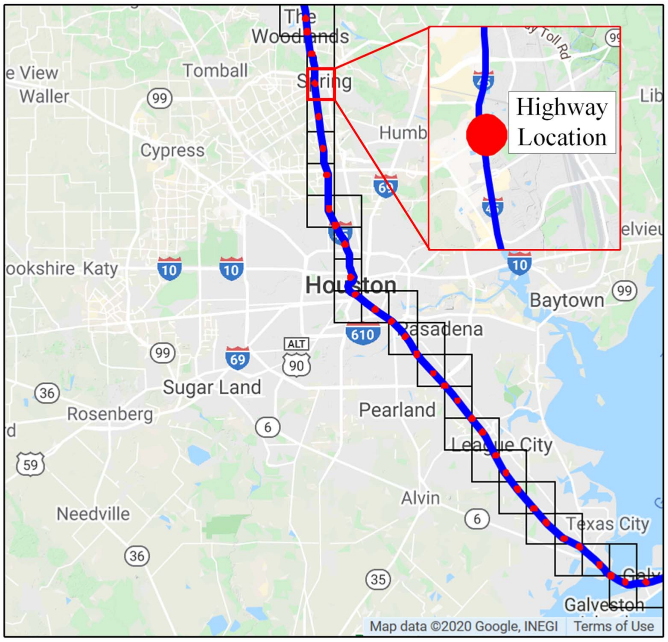

Fig. 5. Discrete representation of IH-45 (rendered by the gmplot package in Python). (Map data © 2020 Google, INEGI.)

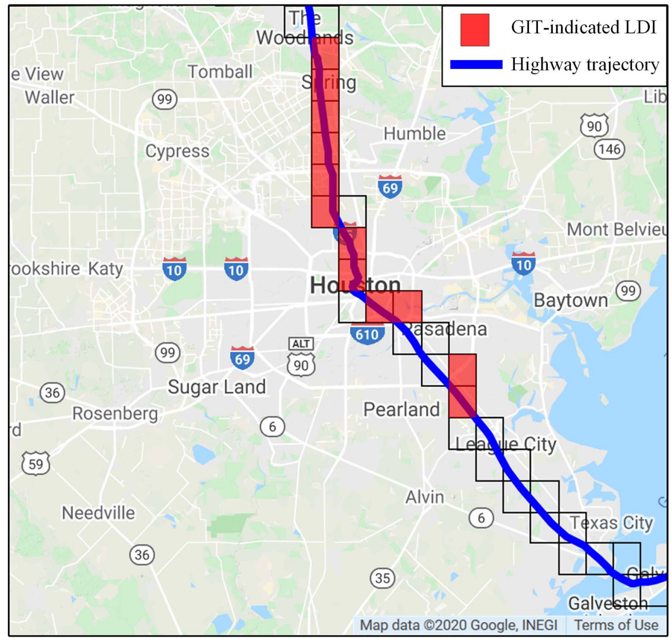

Fig. 6. Geographic distribution of GIT-indicated LDIs for IH-45 on August 29, 2017 (rendered by the gmplot package in Python). (Map data © 2020 Google, INEGI.)

$$d = \sqrt[2]{(\mathrm{GIT_{lat}} - \mathrm{HL_{lat}})^2 + (\mathrm{GIT_{lng}} - \mathrm{HL_{lng}})^2} \tag{2}$$

The subscripts lat and lng represent latitude and longitude, respectively. A detailed process for indicating impacted locations is illustrated in a step-by-step pseudocode form as follows:

```
Input: Impact-related tweets, highway trajectories
  Repeat
    Select a highway and load its corresponding trajectory
    Segment the selected highway into discrete blocks
    Load impact-related tweets associated with this highway
    Repeat
      Select a tweet and obtain its geolocation
      Calculate a distance between the GIT and the highway
      location using Eq. (2)
      Search the nearest highway location and indicate it as an
      LDI
    Until all loaded impact-related tweets are selected
  Until all highways of interest are selected
Output: GIT-indicated LDIs
```

The geographic distribution of GIT-indicated LDIs (shaded areas) on August 29, 2017, for IH-45 is shown in Fig. 6. Notably, the central and northern locations of IH-45 were widely impacted.

Validation

To demonstrate the feasibility and applicability of the proposed approach, the capability of NIT for measuring SDI and the capability of GIT for indicating LDI were systematically validated from the perspectives of timeliness and reliability. Here, the timeliness indicates that the two derived metrics are capable of measuring disaster impacts in a daily manner that is much faster than conventional approaches (i.e., site visits and surveys), and the reliability implies that the assessed disaster impacts are consistent with the actual disaster impacts derived from highway high-water incident data recorded by TxDOT during Hurricane Harvey. Sample data of the high-water incidents are listed in Table 4, which presents the impacted highways and their corresponding time and location information. Time information includes detection time, cleared time, and duration of a high-water incident; and location information comprises a pair of latitude and longitude. In detail, time information was used to group high-water incident impacts into daily time intervals, and location information was processed to depict the actual impacted geographical areas. Comprehensive validations for both NIT and GIT are discussed in the following sections.

Validation of NIT

To validate the timeliness and reliability of NIT for measuring SDI, cross-correlation and Pearson correlation analyses were performed. In this section, the timeliness is validated through analyzing temporal variations of two time series, daily NIT and SDI. The relationships between daily NIT and SDI for all selected highways are illustrated in Fig. 7. Consistent patterns—a sharp initial increase, followed by a peak, and then a gradual decrease—are observed in both NIT and SDI sequences for all selected highways. Although the general patterns are visually consistent, certain variations still exist for these highways. For instance, for US-290, US-225, IH-610, and Westpark Tollway (WT), there are 1-day time lags between their NITs and SDIs in the initial increasing period, while for other highways, their NITs and SDIs increased simultaneously.

Given these temporal variances, cross-correlation analysis was performed to indicate the similarity of these two time series as a function of the time lag of NIT to SDI. Fig. 8 shows the cross-correlation results, which reveal the similarity between daily NIT and SDI with various time lags. The level of the similarity is indicated by the value of the correlation coefficient, which has a value between −1 and 1, where −1 is a total negative correlation and 1 is a total positive correlation. Considering the fact that the employed

Table 4. Sample data of high-water incidents for highways

Incident	Highway	Latitude	Longitude	Detection time	Cleared time	Duration (min)
483401	IH-10	29.7790	−95.9514	August 26, 2017 at 2:44	August 29, 2017 at 19:21	5,314
483446	IH-45	30.2073	−95.4560	August 26, 2017 at 9:15	August 26, 2017 at 13:55	274
483447	IH-45	30.3252	−95.4752	August 26, 2017 at 9:21	August 26, 2017 at 15:39	376
483388	IH-69	29.4943	−95.9116	August 25, 2017 at 22:19	August 30, 2017 at 18:50	6,988

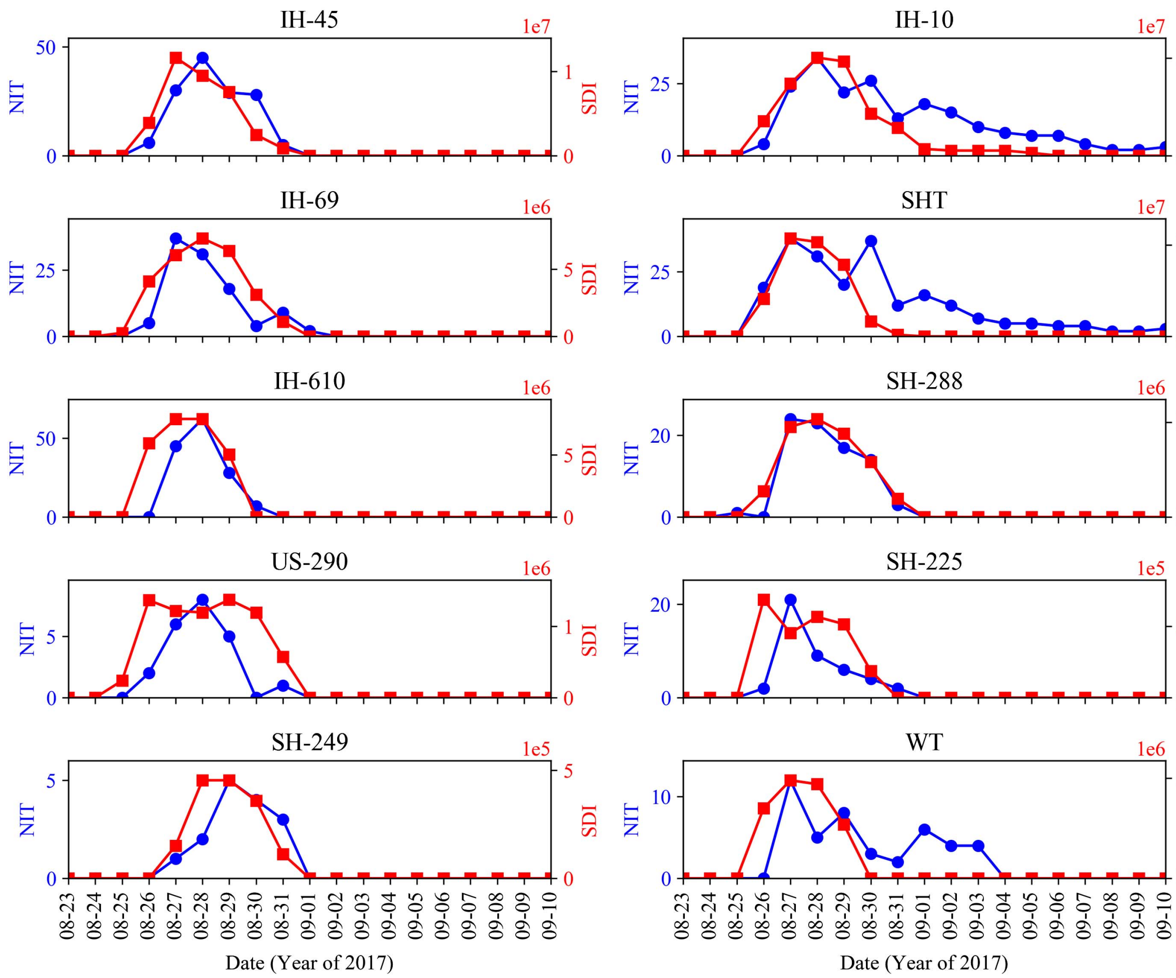

Fig. 7. Temporal variations of NIT and SDI.

social media activities are induced by disasters, the NIT is always shifted with a positive time lag (i.e., leftward shift) in each analysis to find the best positive correlation. The best correlations between daily NIT and SDI for all selected highways are significant and located at time lag zero or one. This implies that NIT is capable of reliably explaining the temporal trends of SDI in a daily (time lag is 0) or near-daily (time lag is 1) manner.

Following the validation of timeliness, the accuracy of NIT for measuring the absolute value of SDI was proven. Pearson correlation analysis between NIT and SDI was performed on the dates of the predefined peak period (i.e., from August 27 to 30, 2017). For each date, the NIT and SDI for a highway were taken as a sample. In total, 40 samples (10 highways with the four dates of the peak period) were extracted, as shown in Fig. 9. A linear regression model was fit, showing the correlations, in which the Pearson correlation coefficient R is 0.76 at a significance level of 0.001. This strong correlation, together with the validated timeliness, demonstrates the capability of NIT for measuring SDI in a timely and reliable manner.

Validation of GIT

In this section, the capability of GIT for indicating LDI is validated through comparing the GIT-indicated LDIs with the actual LDIs. The actual LDIs—locations with one or more high-water incidents—were derived from high-water incidents recorded by TxDOT. For comparison purposes, the GIT-indicated LDIs and the actual LDIs were characterized by three types of LDI, namely, correctly indicated LDI, incorrectly indicated LDI, and missed LDI. A correctly indicated LDI is a location that is indicated by GIT and was recorded by TxDOT. An incorrectly indicated LDI is a location that is indicated by GIT yet not recorded by TxDOT. A missed LDI is a location that is not indicated by GIT yet was recorded by TxDOT. Therefore, each GIT-indicated LDI was categorized as a correctly or an incorrectly indicated LDI, while each actual LDI was categorized as a correctly indicated LDI or missed LDI. For illustration purposes, the geographic distributions of three LDI types for IH-45 on August 29, 2017, are shown in Fig. 10. While GIT is capable of indicating disaster impacts on the central

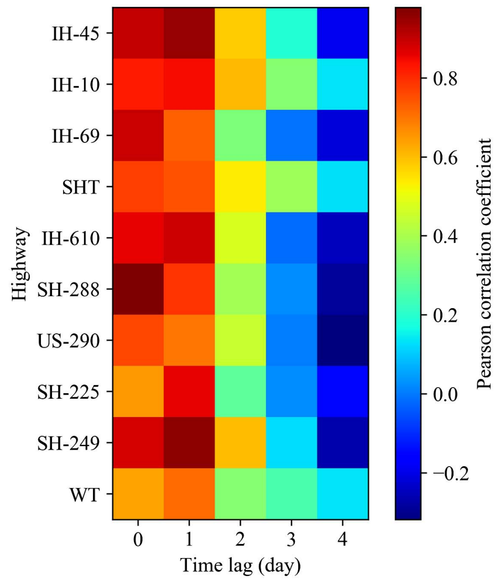

Fig. 8. Cross-correlation between daily NIT and SDI for various time lags.

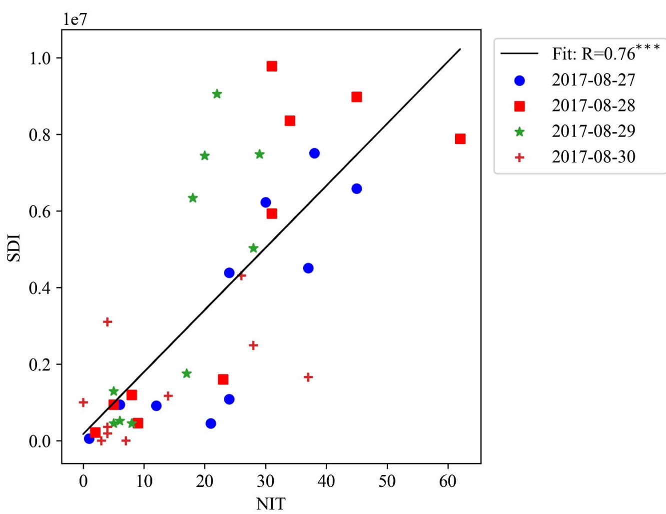

Fig. 9. Linear regression plot between NIT and SDI ($R = 0.76$, $p < 0.001$).

and northern segments, it is less capable of indicating disaster impacts on the southeastern segments. The potential reason for this incapability is that the residents in southeastern area were evacuated during Hurricane Harvey prior to the flooding (McCausland and Chuck 2017).

To quantify the capability of GIT for indicating LDI, two evaluation measures, correct indication rate and total indication rate, are formulated as Eqs. (3) and (4)

$$\text{Correct indication rate} = \frac{N_c}{N_c + N_i} \tag{3}$$

$$\text{Total indication rate} = \frac{N_c}{N_c + N_m} \tag{4}$$

where N_c, N_i, and N_m = number of correctly indicated, incorrectly indicated, and missed LDIs, respectively. The correct indication

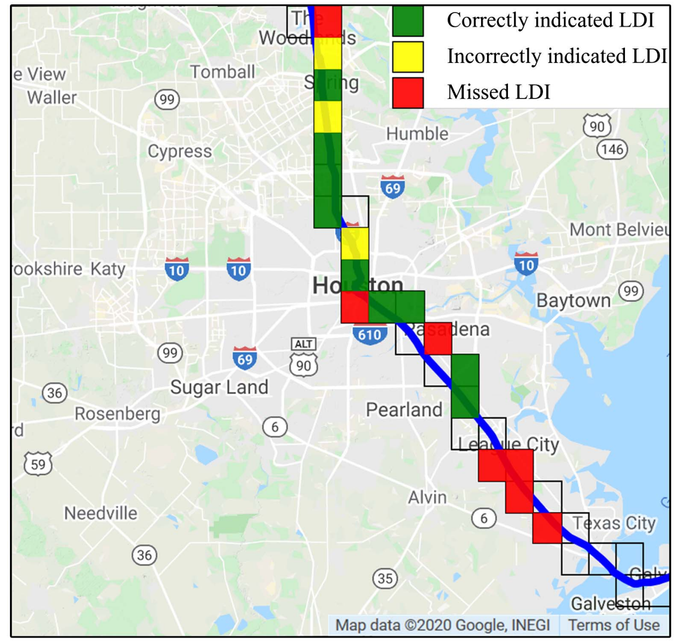

Fig. 10. Geographic distributions of three types of LDI for IH-45 on August 29, 2017 (rendered by the gmplot package in Python). (Map data © 2020 Google, INEGI.)

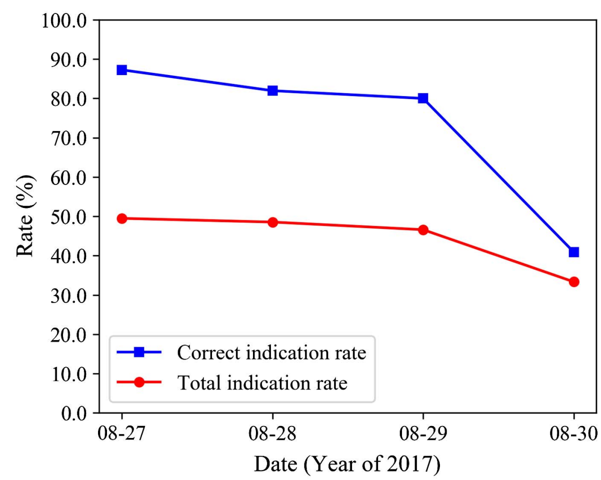

Fig. 11. Temporal variations of correct capture rate and total capture rate.

rate, i.e., the percentage of GIT-indicated LDIs that are actual LDIs, represents the confidence of taking a GIT-indicated LDI as an actual LDI. A higher correct indication rate indicates stronger confidence. The total indication rate is the percentage of actual LDIs that are indicated by GIT. The numerator is the number of correctly indicated LDIs and the denominator is the number of actual LDIs that are composed by correctly indicated LDIs and missed LDIs. Total indication rate depicts the probability that an actual LDI is detected by GIT, and a higher total indication rate reveals a higher detection probability.

The temporal patterns of correct and total indication rates for all the selected highways are shown in Fig. 11. In the first three days of the peak period (i.e., August 27–29, 2017), both measures have constant values, wherein the correct indication rate is around 80% and the total indication rate is around 50%.

But on August 30, 2017, there is a sharp decrease for both measures. Through manually checking the impact-related tweets posted on August 30, 2017, this decrease was found to be the result of most tweets not describing disaster impacts on highway main lanes but disaster impacts on highway exits and frontages, which were not recorded as highway high-water incidents by TxDOT. Tweet examples include "Exit blocked due to flooding on I-69" and "Closed due to flooding. in #Houston on I-45 Frontage Rd." This is consistent with the fact that Hurricane Harvey had left Houston on August 30, 2017, and the main lanes of highways recovered from the flooding (Murphy 2018). In summary, the overall correct indication rate and total indication rate are 80% and 46% during the peak period, which indicates that a GIT-indicated LDI is taken as an actual LDI with 80% confidence and an actual LDI is identified by GIT with 46% probability. These consistencies demonstrate the GIT's capability of providing a preliminary assessment of LDI.

Discussion

While most social media–based studies have focused on general disaster situations (e.g., donation, caution and advice, and causalities and damage), the potential of assessing disaster impacts on a specific type of infrastructure using social media is less investigated. This research demonstrates the feasibility and applicability of obtaining a rapid assessment of disaster impacts on highways through social media, which contributes to decision making in disaster response. The developed approach has the potential to be generalized to obtain rapid insights of disaster impacts on other critical infrastructure systems (e.g., power, public health, and housing).

In contrast to previous research efforts on assessing general disaster damage, this study explores the potential of a more specific, applicable, and reliable approach in evaluating disaster impacts on highways using social media. The reliability of the proposed approach was validated through a strong correlation ($R = 0.76$, $p < 0.001$) between NIT and SDI, and acceptable geographical consistencies (correct indication rate $= 0.74$, total indication rate $= 0.46$) between GIT-indicated LDIs and actual LDIs. Superior to conventional assessment approaches (i.e., site visits), the developed approach is capable of providing timely insights on disaster impacts, which can be utilized as a preliminary assessment tool to support decision-making (e.g., timely allocation of relief funding and assessment professionals). Meanwhile, the developed approach is easy and safe to implement to assess disaster impacts on highways during the chaotic time of a disaster because it solely requires crowdsourced data that are accessible through social media platforms. The rapidly assessed disaster impacts on highways are expected to be employed to enhance the planning effectiveness with the overall aim of mitigating disastrous impacts and maintaining community resilience. Two potential applications are identified from the perspectives of route planning support and highway recovery.

Route Planning Support

During and after a disaster, time-sensitive relief and recovery efforts are required to ensure community resilience. Mostly, these time-sensitive efforts are distributed through land transportation that heavily relies on functional highways (Hamedi et al. 2012). Compared with other information sources (e.g., cameras, sensors, and mapping applications) for route planning, the proposed approach is capable of providing timely insights of disaster impacts for enhanced awareness of highway conditions. Take Hurricane Harvey as an example: Trucking fleets were continuing to enter Houston through these affected highways to deliver relief supplies (e.g., food, water, and medical supports) to the overwhelmed communities, which led to inefficient responses (Kilcarr 2017). Through the proposed approach, disaster impacts on highways can be timely assessed, which is helpful for trucking fleets to avoid flooded highways, thereby ensuring delivery of supplies and mitigating safety risks.

Highway Recovery

Planning the recovery efforts for damaged highways in the aftermath of a disaster is a challenging task due to limited recovery resources (Stumpf et al. 2009). Optimal distribution of these limited resources has significant impacts on the societal needs of minimizing the recovery cost and the performance loss of highways (Eid and El-adaway 2017). Therefore, these limited resources need to be deployed and utilized in a timely and effective manner. Through the developed approach, the severity and location of disaster impacts on highways are derived to assist the allocation of recovery resources, thereby prompting the efficiency of highway recovery.

Conclusion

This research developed a systematic social media data adaption module and inventively designed social media–based decision metrics to achieve a timely and reliable assessment of disaster impacts on highways. Through the social media data adaption module, data related to disaster impacts on highways are extracted in an efficient and reliable manner. Utilizing the extracted data, disaster impacts on highways are assessed from the perspectives of severity and location through two derived social media–based metrics, namely, NIT and GIT. A systematic and comprehensive validation is performed to demonstrate the timeliness and reliability of the proposed approach for assessing disaster impacts on highways using social media.

Academically, this approach achieves the theoretical and methodological advancements of assessing disaster impacts on highways through social media. From the theoretical side, disaster impacts on highways are assessed using the related social media activities, wherein NIT is introduced for measuring SDI and GIT is employed for indicating LDI. In terms of the methodological advancement, a systematic data adaption module is developed for reliably extracting social media data related to disaster impacts on highways.

To demonstrate the feasibility and applicability of the presented approach, disaster impacts of Hurricane Harvey on highways in Houston were studied. Highway high-water incident data, collected by TxDOT, were used to validate the timeliness and reliability of the two derived indicators, NIT and GIT. Results demonstrate that the presented social media–based approach is capable of assessing disaster impacts on highways in a timely and reliable manner. Practitioners can use the assessed disaster impacts to enhance the effectiveness of route planning support and highway recovery. The research also sheds light on further implementation of social media for assessing disaster impacts on various infrastructure systems (e.g., power and water systems).

Although this research proposed a systematic approach to assessing disaster impacts on highways using social media, its capability might be limited when the related social media activities (highway-related tweets) are relatively insufficient. In the future, information from alternative highway-related data sources (e.g., mapping applications and traffic sensors) could be incorporated with social media information to achieve an enhanced assessment of disaster impacts on highways.

Data Availability Statement

Highway high-water incidents data were provided by a third party. Requests for data should be directed to the provider indicated in the "Acknowledgments." Social media data are proprietary and may only be provided with restrictions.

Acknowledgments

This study is funded by the Thomas F. and Kate Miller Jeffress Memorial Trust (Grant No. 223476) and the National Science Foundation (Grant No. 1761950). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the Thomas F. and Kate Miller Jeffress Memorial Trust and the National Science Foundation. The authors would like to acknowledge Valerie K. Taylor at TxDOT for sharing the valuable validation data.

References

ASCE. 2017. *Infrastructure report card*. Reston, VA: ASCE.

Blake, E. S., and D. A. Zelinsky. 2018. *Tropical cyclone report: Hurricane Harvey.* Silver Spring, MD: National Oceanic and Atmospheric Administration.

Boyd, D., S. Golder, and G. Lotan. 2010. "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter." In *Proc., Annual Hawaii Int. Conf. System Science*, 1–10. Piscataway, NJ: IEEE.

Bramer, M. 2007. *Principles of data mining*. London: Springer.

Cai, L., and Y. Zhu. 2015. "The challenges of data quality and data quality assessment in the big data era." *Data Sci. J.* 14 (May): 2–29. https://doi.org/10.5334/dsj-2015-002.

Chang, S. E. 2003. "Transportation planning for disasters: An accessibility approach." *Environ. Plan. A* 35 (6): 1051–1072. https://doi.org/10.1068/a35195.

Chen, Y., W. Ji, and Q. Wang. 2019. "A Bayesian-based approach for public sentiment modeling." In *Proc., Winter Simulation Conf.*, 3053–3063. Piscataway, NJ: IEEE.

Davies, A. 2017. "Houston's trucking lifeline rumbles back in Harvey's aftermath." Accessed April 22, 2019. https://www.wired.com/story/hurricane-harvey-houston-trucks-resupply.

Eid, M. S., and I. H. El-adaway. 2017. "Sustainable disaster recovery decision-making support tool: Integrating economic vulnerability into the objective functions of the associated stakeholders." *J. Manage. Eng.* 33 (2): 04016041. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000487.

Fan, C., Y. Jiang, and A. Mostafavi. 2020. "Social sensing in disaster city digital twin: Integrated textual–visual–geo framework for situational awareness during built environment disruptions." *J. Manage. Eng.* 36 (3): 04020002. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000745.

Faturechi, R., and E. Miller-Hooks. 2015. "Measuring the performance of transportation infrastructure systems in disasters: A comprehensive review." *J. Infrastruct. Syst.* 21 (1): 04014025. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000212.

FHWA. 2013. *Disaster assessment and damage survey summary.* Washington, DC: FHWA.

Ha, S., K. Kim, K. Kim, H. Jeong, and H. Kim. 2017. "Reliability approach in economic assessment of adapting infrastructure to climate change." *J. Manage. Eng.* 33 (5): 04017022. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000530.

Ham, Y., and J. Kim. 2020. "Participatory sensing and digital twin city: Updating virtual city models for enhanced risk-informed decision-making." *J. Manage. Eng.* 36 (3): 04020005. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000748.

Hamedi, M., A. Haghani, and S. Yang. 2012. "Reliable transportation of humanitarian supplies in disaster response: Model and heuristic." *Procedia Soc. Behav. Sci.* 54 (Oct): 1205–1219. https://doi.org/10.1016/j.sbspro.2012.09.835.

Immonen, A., P. Pääkkönen, and E. Ovaska. 2015. "Evaluating the quality of social media data in big data architecture." *IEEE Access* 3 (Oct): 2028–2043. https://doi.org/10.1109/ACCESS.2015.2490723.

Imran, M., C. Castillo, F. Diaz, and S. Vieweg. 2018. "Processing social media messages in mass emergency: A survey." In *Proc., Int. World Wide Web Conf. Communication*, 507–511. New York: Association for Computing Machinery.

Kilcarr, S. 2017. "Trucks roll for Harvey recovery effort." Accessed November 12, 2019. https://www.trucker.com/equipment/trucks-roll-harvey-recovery-effort.

Kim, D., T. Hong, M. Lee, C. Koo, J. Kim, and K. Jeong. 2017. "Mapping the rescue equipment mobilization potential: Decision support tool for emergency management." *J. Manage. Eng.* 33 (6): 04017037. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000550.

Kim, J., J. Bae, and M. Hastak. 2018. "Emergency information diffusion on online social media during storm Cindy in U.S." *Int. J. Inf. Manage.* 40 (Jun): 153–165. https://doi.org/10.1016/j.ijinfomgt.2018.02.003.

Kim, J., and M. Hastak. 2018. "Social network analysis: Characteristics of online social networks after a disaster." *Int. J. Inf. Manage.* 38 (1): 86–96. https://doi.org/10.1016/j.ijinfomgt.2017.08.003.

Kirilenko, A. P., and S. O. Stepchenkova. 2014. "Public microblogging on climate change: One year of Twitter worldwide." *Global Environ. Change* 26 (1): 171–182. https://doi.org/10.1016/j.gloenvcha.2014.02.008.

Kirk, R. S. 2012. "Emergency relief program: Federal-aid highway assistance for disaster-damaged roads and bridges." Accessed November 12, 2019. https://fas.org/sgp/crs/homesec/R42804.pdf.

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

Li, H., D. Caragea, C. Caragea, and N. Herndon. 2018a. "Disaster response aided by tweet classification with a domain adaptation approach." *J. Contingencies Crisis Manage.* 26 (1): 16–27. https://doi.org/10.1111/1468-5973.12194.

Li, Z., C. Wang, C. T. Emrich, and D. Guo. 2018b. "A novel approach to leveraging social media for rapid flood mapping: A case study of the 2015 South Carolina floods." *Cartography Geogr. Inf. Sci.* 45 (2): 97–110. https://doi.org/10.1080/15230406.2016.1271356.

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

McCausland, P., and E. Chuck. 2017. "*Hurricane Harvey evacuations: Residents warned to leave, stay away.*" Accessed August 25, 2017. https://www.nbcnews.com/storyline/hurricane-harvey/hurricane-harvey-evacuations-residents-college-students-warned-stay-away-n795891.

Murphy, J. D. 2018. *Service assessment: August to September 2017 Hurricane Harvey.* Silver Spring, MD: National Oceanic and Atmospheric Administration.

Orabi, W., A. B. Senouci, K. El-Rayes, and H. Al-Derham. 2010. "Optimizing resource utilization during the recovery of civil infrastructure systems." *J. Manage. Eng.* 26 (4): 237–246. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000024.

OSHA (Occupational Safety and Health Administration). 2018. *Hazard exposure and risk assessment matrix for hurricane response and recovery work: Search and rescue*. Washington DC: OSHA.

Ragini, J. R., and P. M. R. Anand. 2017. "An empirical analysis and classification of crisis related tweets." In *Proc., Int. Conf. Computational Intelligence Computing Research*, 1–4. Piscataway, NJ: IEEE.

Rau, J., L. Chen, J. Liu, and T. Wu. 2007. "Dynamics monitoring and disaster assessment for watershed management using time-series satellite images." *IEEE Trans. Geosci. Remote Sens.* 45 (6): 1641–1649. https://doi.org/10.1109/TGRS.2007.894928.

Stumpf, A., H. Kim, and E. Jenicek. 2009. *Planning post-disaster reconstruction efforts of damaged transportation networks.* Washington, DC: Congressional Research Service.

Tian, H., Y. Tao, S. Pouyanfar, S. C. Chen, and M. L. Shyu. 2019. "Multimodal deep representation learning for video classification." *World Wide Web* 22 (3): 1325–1341. https://doi.org/10.1007/s11280-018-0548-3.

TxDOT (Texas Department of Transportation). 2017. "District traffic and urban saturation web maps of 2017." Accessed November 12, 2019. http://txdot.maps.arcgis.com/apps/webappviewer/index.html?id=06fea0307dda42c1976194bf5a98b3a1.

Wang, Q., and J. E. Taylor. 2016. "Process map for urban-human mobility and civil infrastructure data collection using geosocial networking platforms." *J. Comput. Civ. Eng.* 30 (2): 04015004. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000469.

Yu, M., Q. Huang, H. Qin, C. Scheele, and C. Yang. 2019. "Deep learning for real-time social media text classification for situation awareness—Using Hurricanes Sandy, Harvey, and Irma as case studies." *Int. J. Digit. Earth* 12 (11): 1230–1247. https://doi.org/10.1080/17538947.2019.1574316.

Yuan, F., and R. Liu. 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 (Jun): 758–767. https://doi.org/10.1016/j.ijdrr.2018.02.003.

Yuan, F., and R. Liu. 2020. "Mining social media data for rapid damage assessment during Hurricane Matthew: Feasibility study." *J. Comput. Civ. Eng.* 34 (3): 05020001. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000877.