Introduction to social sensing and big data computing for disaster management

Zhenlong Li [a*], Qunying Huang [b], Christopher T. Emrich [c]

[a] *Geoinformation and Big Data Research Laboratory, Department of Geography, University of South Carolina, South Carolina, USA*

[b] *Department of Geography, University of Wisconsin-Madison, Wisconsin, USA*

[c] *School of Public Administration and National Center for Integrated Coastal Research, University of Central Florida, Florida, USA*

[*]zhenlong@sc.edu

Abstract: Traditional data collection methods such as remote sensing and field surveying often fail to offer timely information during or immediately following disaster events. Social sensing enables all citizens to become part of a large sensor network, which is low cost, more comprehensive, and always broadcasting situational awareness information. However, data collected with social sensing is often massive, heterogeneous, noisy, unreliable from some aspects, comes in continuous streams, and often lacks geospatial reference information. Together, these issues represent a grand challenge toward fully leveraging social sensing for emergency management decision making under extreme duress. Meanwhile, big data computing methods and technologies such as high-performance computing, deep learning, and multi-source data fusion become critical components of using social sensing to understand the impact of and response to the disaster events in a timely fashion. This special issue captures recent advancements in leveraging social sensing and big data computing for supporting disaster management. Specifically analyzed within these papers are some of the promises and pitfalls of social sensing data for disaster relevant information extraction, impact area assessment, population mapping, occurrence patterns, geographical disparities in social media use, and inclusion in larger decision support systems.

Keywords: social media, VGI, big data, natural hazards, spatial computing

To cite: Li, Z., Huang, Q., Emrich, C., (2019) Introduction to Social Sensing and Big Data Computing for Disaster Management, *International Journal of Digital Earth*, doi: 10.1080/17538947.2019.1670951

1. Introduction

Rapid onset disasters, often difficult to prepare for and respond to, make disaster management a challenging task worldwide. Disaster and emergency management effectiveness depends heavily on making good decisions in near-real time under extreme duress. These key, often life-saving, decisions are possible only with real-time situational awareness (SA) describing “the idealized state of understanding what is happening in an event with many actors and other moving parts, especially with respect to the needs of command and control operations” (Vieweg et al., 2010). However, detailed levels of SA are mainly depended on real-time data availability and a capacity to collect, process, synthesize, and analyze these multi-sourced data in a timely fashion.

Traditional data collection practices such as in-situ sensing, field surveying and remote sensing often fail to offer timely information during or directly following a disaster event. For example, stream gauges positioned in rivers (in-situ sensing) are only useful while the stations are functioning properly and before they are overtopped by floodwaters and rendered inoperable. Field surveying is often costly, labor intensive, and suffers significant delays making it a less optimal method for obtaining information in affected areas during and immediately following disasters as lifesaving missions take place. Remote sensing renders a synoptic view in a large geographic area, significantly contributing to a holistic understanding of disaster situation. Remotely sensed imagery generally provides substantial visual evidence of impact and damage areas and has widely been utilized for monitoring pre-disaster conditions, evaluating impacts, and assessing damages following disaster. However, some inherent restrictions within remote sensing imagery hamper its utility for disaster management. For example, the coarse temporal resolution (e.g., long revisit cycle) and extreme weather conditions (e.g., cloud cover during a storm) prevent remote sensors from acquiring timely good-quality images during critical disaster response phases (Li et al., 2018).

As a new data collection approach, social sensing first appearance in the literature (2006) where it was viewed as one paradigm of the citizen-initiated sensing in personal, social, and urban settings enabled by embedded sensing technology and internet infrastructure (Srivastava et al., 2006). It was later discussed in Campbell et al. (2008) where social sensing was considered as a type of people-centric sensing along with private sensing and public sensing, “in which information is shared within social and special interest groups (p.13)”. Meanwhile, geographer Goodschild (2007) coined he terms “voluntary geographic information (VGI)” when considering human as sensors in the world of volunteered geography. VGI draws upon the concept of collaborative user-generated content through crowdsourcing, by which many users with varying levels of expertise contribute geographic data via the web (Goodchild & Glennon, 2010). A more recent study defines social sensing as “a category of spatiotemporally tagged big data that provide an observatory for human behavior, as well as the methods and applications based on such big data.” (Liu et al., 2015, p.525). In that study, VGI is differentiated from social sensing as an active data collection approach while social sensing is more passive (such as data collected from the social media platforms). We consider social sensing inclusive of both active human sensors (e.g., those participating in a crowdsourcing project) and passive human sensors (e.g., social media users), where they sense the environment and broadcast such sensed data via social media or dedicated crowdsourcing apps.

In the disaster management context, social sensing enables all citizens to become part of a large sensor network and a homegrown disaster response team by sharing information such as texts, images, and videos through social media platforms (e.g., Twitter, Facebook, Instagram) as well as dedicated crowdsourcing applications such as the USGS “Did You Feel It” project (Atkinson and

Wald, 2007). Compared to traditional physical sensors, such a citizen-sensor network is low cost, more spatially comprehensive, and always broadcasting SA information. Social sensing enables wide-scale interaction where primary data becomes "collectively resourceful, self-policing, and d generative of information that cannot otherwise be easily obtained" in a useful timeframe for disaster areas (Sutton, Palen & Shklovski, 2008, 7). For example, context-rich micro-level disaster information (e.g., site specific damage and flood water depth) can be captured in real-time through social media platforms such as Twitter, enabling rapid assessment of evolving disaster situations.

While social sensing offers a new data source in disaster management, the data itself is often massive, heterogeneous, noisy, unreliable, comes in continuous streams, and mostly lacks geospatial reference information. Millions of microblog posts from different social media platforms can be generated in a short time during or right after an impactful disaster. For example, during the 2011 East Coast earthquake impacting regions from North Carolina, USA to Quebec, Canada, over 5,000 earthquake-related tweets were being sent per second according to Twitter (DHS, 2013). Effectively synthesizing and processing such social sensing data exhibit the typical five-V challenges of big data, including volume (millions of posts), variety (different data formats such as textual, multimedia), veracity (noisy and biased), velocity (rapid data streams), and value (transformation from data to knowledge). Hence, big data computing methods and technologies such as high-performance computing, distributed geoinformation processing, deep learning, spatial statistics/modeling, and multi-source data fusion become critical components when using social sensing to understand impacts from and response to disaster events in a timely fashion. We believe that social sensing coupled with innovative big data computing methods offers enormous opportunities for disaster management by examining the physical infrastructure (e.g., physical damage), environment (e.g., disaster specific extents), and nature-human interaction (e.g., evacuation) from spatial, temporal, and social dimensions.

Along these lines, this special issue captures and highlights recent advancements in leveraging social sensing and big data computing for enabling decision making in one or more disaster phases (mitigation, preparedness, response, and recovery). The articles included in the issue makes contribution to the research of social sensing and big data computing in the context of disaster management by presenting new methodologies and techniques for the examination, extraction, modeling, mapping and/or assessment of disaster relevant information (Sit et al. 2019, Huang et al. 2018, Yu et al. 2019), impact area (Wang et al. 2019), geographic location information (Mao et al. 2019), population (Kubicek et al. 2019), occurrence patterns (Zheng et al. 2018), and geographical disparities in social media use (Zhou et al. 2019), as well as the development of the decision making system (Zhang et al. 2018).

2. Overview of the articles featured in this special issue

Extracting SA information from massive "citizen sensed" social media datasets is quite challenging as only a small portion of such datasets are disaster relevant and useful. In addition, it is difficult to further classify disaster relevant datasets into finer "actionable" categories (e.g., caution, donation, damage) after discarding noise (Huang and Xiao 2015). To address this issue, the first paper by Sit, Koylu, and Demir (2019) introduces a framework integrating deep learning and natural language processing for identifying disaster-relevant information and specific information categories. Within this framework, a deep artificial neural network architecture is first used for classifying social media messages (e.g., tweets in this case) about the disaster context. Topic modeling is leveraged to identify various categories of disaster relevant tweets such as affected individuals, donations and support, advice, caution and alerts. Finally, temporal, spatial, and spatiotemporal aspects of different types of disaster relevant information are analyzed. The results with hurricane Irma revealed the progression of the disaster including the preparedness, response and recovery efforts, and pinpoint the potential areas of damage and affected individuals.

Similarly, Huang et al. (2019) also intend to address the challenge of extracting disaster relevant tweets from a massive twitter data pool. This paper introduces a flood relevant tweet extraction approach leveraging both visual and textual information contained within a tweet. A convolutional neural network (CNN) model first classifies the visual content of flood pictures, followed a sensitivity test and duplication test to create a tweet pool for a flood event. The results indicated that coupling CNN classification results with flood sensitive words in tweets improves classification performance. The elimination of tweets containing duplicative photos greatly contributes to higher spatiotemporal relevance to the flood. Using 2017's Houston Flood as a study case, the proposed approach to tagging the flood tweets from massive social media data is demonstrated. This work enables the selection of tweets with flood-related visual information and flood-sensitive textual information. The automatically derived geotagged flood pictures could seed a wide range of flood studies for improved SA and rapid flood response.

Yu et al. (2019) further refine the utility of disaster relevant information in real time, by looking back at past disasters in preparation for future threats. Each disaster event evolves differently through time, leading to a diverse set of topics discussed throughout social media platforms. However disparate an individual disaster event appears, looking across multiple disasters will produce common themes. Once identified, a classification model trained on datasets generated from historic events should be capable of automatically categorizing the text messages for future events. Unfortunately, existing models based on machine learning techniques, such as support vector machine (SVM), are mostly event specific and difficult to handle data generated during a new disaster event. To tackle this challenge, Yu et al. (2019) examines the capability of a CNN model in cross-event Twitter topic classification based on three geotagged twitter datasets collected during Hurricanes Sandy, Harvey, and Irma. Experiment results indicate that CNN models, comparing to two traditional machine learning methods, achieved a consistently better performance for both single event and cross-event evaluation scenario. The study also reveals that the CNN model has the capability of pre-training Twitter data from past events to classify for an upcoming event for SA.

As noted earlier, situational awareness is crucial for emergency management decision making in time-critical situations. Having access to the most reliable and accurate information can save lives and protect property. The fourth paper by Wang et al. (2019) utilizes social media (geotagged tweets) to quickly estimate the impact area of a large earthquake. At its core, positive results from studies such as these can help the emergency management community gain access to better situational awareness information much more quickly than traditional impact assessment techniques. This paper combines social media and other ancillary data using a spatial logistic growth model to estimate earthquake impact areas. Results and confirmatory analytics indicate that social media data containing geospatial references representing real-time citizen sensor data, can be used to substantially enhance emergency management by rapidly mapping the impact area of a disaster.

Mao et al. (2019) take a slightly different approach to understanding disasters from twitter information recognizing that although geo-spatially enabled social networking data provides detailed locational information the fact remains that geo-tagged tweets, among other platforms, account for a very small portion of social media's big data. Because location is an important aspect of disaster impacts and damage estimation we must attempt to identify, develop, and deploy novel approaches that can utilize a broader set of social media data. The paper by Mao et al. (2019) explores a new deep learning method to extract outage locations from social media messages to map near real-time power outages rather than relying on locational information attached to such messages. Promising results point to strong positive linear correlations between socially derived outage maps and national outage indicators developed by leading national research labs. This novel approach to extracting locations directly from messages helps overcome data scarcity challenges with current low "buy in" to geotagged tweets and highlights the value add provided by free and public socially sensed data.

Kubicek et al. (2019) provide a case study of cellular phone information for situational awareness. Cellular phones, perhaps the most ubiquitous personal technology available to society, provide a unique platform from which people can communicate not only verbally, but also through social media. Although cellular phone data is not easy to access because of privacy concerns, this data contains a wealth of useful information for understanding the site and situation of users. During disaster scenarios specifically, linking locational information with hazard zone information can quickly inform decision makers about the number of at-risk people. This paper devised a novel approach to model and map population distribution at fine spatiotemporal scale based on big mobile phone data. A number of research questions drive the focus of this study on the utility of cellular data to better understand population movement and locational patterns in the context of a disaster. The general, and specific emergency management utility of cellular based location data is proven through a series of geostatistical tests. Furthermore, use of cellular data identified deficits in current census level demographics for understanding how many people are at a given place and time.

Zhou et al. (2019) move from data coverage toward examining coverage disparities during disasters. Knowing where at risk and vulnerable populations might not have adequate access to information is a key component for emergency management planning and response. Zhou et al. (2019) examined the questions related to geographical disparities in twitter use before, during, and after a major disaster (Hurricane Harvey) in Texas and Louisiana using ratio and sentiment analysis. This article explores twitter use patterns under the presumption that better socioeconomic conditions would also have higher twitter usage. A novel twitter data mining framework developed by the authors captured disaster-related tweets in a Hadoop computing environment searching on four hurricane related key words. One of two approaches provided a geolocation for each tweet, enabling geospatial analysis across the study area. Ratio and sentiment analysis, derived for multiple geographic and time scales, provided specific twitter datasets for each phase of the emergency management cycle. Summary statistics, correlations, and linear regressions revealed significant social and geographical disparities in Twitter usage.

Zheng et al (2019) build from the premise that a single hazard can cascade into multiple disaster events to develop a better understanding of how socially sensed data can help us understand these linkages. Many historical disaster events prove that disasters are typically not isolated or static. Rather, many cases have shown that secondary and derivative disasters will occur along with the primary event, resulting in a chain of cascading disasters. For example, a blizzard could trigger a series of hazards, such as blocked roads, power grids collapse, air travel delays, and flash floods. Interlocking disaster-causal factors, including natural, biological, physical, chemical, and human characteristics/interactions further contribute to often complex chains of disasters. Understanding the regular patterns of the occurrence and development of disasters in order to control and reduce their negative impacts of disasters. To achieve this goal, Zheng et al.

(2019) proposed an approach to extract the disaster chain of a specific disaster type by collecting the big scholar and social news data with disaster-related keywords, analyzing the strengths of their relationships with the co-word analysis method, and constructing a complex network of all defined disaster types. Within their work, the scholar and social news data are collected from Google Scholar, Baidu Scholar, and Sina News search engines, and the obtained disaster chins are compared with each other to demonstrate the feasibility of the proposed approach. The achieved disaster chains are also compared with the ones concluded from traditional research methods and the very reasonable result is demonstrated. The proposed method can assist in the emergency manager to better prepare in advance and minimum the losses by eliminating the key triggered factors and cutting off the chain.

With the availability of timely and critical information from various data sources including social sensing, the next step is to develop a responsive decision-making process that can integrate this multi-sourced information to support disaster management in the event of disasters. The paper by Zhang et al. (2018) proposes a cyberGIS-enabled multi-criteria spatial decision support system for informing rapid decision-making during emergencies. The system includes a high-performance computing environment, and multi-criteria decision analysis models, including Weighted Sum Model (WSM) and Technique for Order Preference by Similarity to Ideal Solution Model (TOPSIS), along with various types of social vulnerability indicators to solve decision problems with conflicting evaluation criteria during a flood disaster. Using two decision goals generated based on a historic flood event as a case study, the results indicate that WSM generates more diverse values and higher output category estimations than the TOPSIS model. This spatial decision support system enables collaborative problem solving and efficient knowledge transformation between decision makers, where different responders can formulate their decision objectives, select relevant evaluation criteria, and perform interactive weighting and sensitivity analyses.

3. Conclusion and innovation opportunities

Emergency and crisis emergency management remains a challenging task worldwide. Real-time data sources and the ability to timely collect, process, synthesize, and analyze these data is paramount for enabling better decisions in near-real time under extreme duress. The research articles featured in this issue shed light on some of the opportunities, challenges and solutions of leveraging social sensing and big data computing for supporting disaster management. As evidenced in this special issue, social sensing coupled with innovative analytical and modeling methods offers enormous possibilities for disaster management, including mapping near real-time power outages during a disaster, rapidly estimation of an earthquake impact, modeling population distribution at a fine spatiotemporal scale, and uncovering the relationships of multiple disasters. Meanwhile, innovative big data computing methods based on deep learning and high-performance computing have proven to be effective and efficient in extracting relevant and actionable information and identify useful patterns from massive unstructured social sensing data in a timely manner. Lastly, the cyberGIS-enabled spatial decision support system demonstrates the utility of integrating social sensing and big data computing in supporting flood emergency management.

Moving forward, research opportunities related to social sensing and big data computing for disaster management lie in four focus areas: 1) theory innovations: how to develop new theoretical frameworks for effectively fusing multi-sourced social sensing datasets by leveraging state-of-the-art big data computing infrastructures and models; 2) application innovations: how to innovatively use social sensing, a new data collection approach, for disaster management activities including event detection, early warning, flood mapping and damage assessment; 3) computing innovations: how to effectively and efficiently analyze big social sensing data for rapid information extraction and knowledge discovery to better understand the physical, environmental, and social dynamics of a disaster; and 4) engineering innovations: how to develop practical systems by integrating big data computing and social sensing along with other data sources for rapid decision making. From the GIScience perspective, advancements in these innovation areas demand a cross-disciplinary and collaborative approach. Application innovations require domain knowledge from disaster experts, computing innovations benefit from collaborating with computer scientists, and engineering innovation demands collaborations with the first responders.

Acknowledgement

We would like to thank the authors who contributed to this special issue and the reviewers for their helpful suggestions and criticisms that improved the papers in the special issue. We also want to extend our gratitude to Professor Changlin Wang and Dr. Linlin Guan for their editorial support and guidance in preparing the special issue.

Disclosure statement

No potential conflict of interest was reported by the authors

References

Atkinson, G. M., & Wald, D. J. (2007). "Did You Feel It?" intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, 78(3), 362-368.

Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., ... & Ahn, G. S. (2008). The rise of people-centric sensing. *IEEE Internet Computing*, 12(4), 12-21.

DHS, 2013, Innovative Uses of Social Media in Emergency Management, Available online at: https://www.dhs.gov/sites/default/files/publications/Social-Media-EM_0913-508_0.pdf, last accessed on 09/11/2019

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, 3(3), 231-241.

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal 69(4): 211-221.

Huang, Q., & Xiao, Y. (2015). Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, 4(3), 1549-1568.

Huang, X., Wang, C., Li, Z., & Ning, H. (2018). A visual–textual fused approach to automated tagging of flood-related tweets during a flood event. *International Journal of Digital Earth*, 1-17.

Kubíček, P., Konečný, M., Stachoň, Z., Shen, J., Herman, L., Řezník, T., ... & Leitgeb, Š. (2019). Population distribution modelling at fine spatio-temporal scale based on mobile phone data. *International Journal of Digital Earth*, 1-22.

Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science*, 45(2), 97-110.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., ... & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512-530.

Mao, H., Thakur, G., Sparks, K., Sanyal, J., & Bhaduri, B. (2019). Mapping near-real-time power outages from social media. *International Journal of Digital Earth*, 1-15.

Sit, M. A., Koylu, C., & Demir, I. (2019). Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *International Journal of Digital Earth*, 1-25.

Srivastava, M., Hansen, M., Burke, J., Parker, A., Reddy, S., Saurabh, G., ... & Estrin, D. (2006). Wireless urban sensing systems, available online at https://escholarship.org/uc/item/6sj003r4

Sutton, J. N., Palen, L., & Shklovski, I. (2008). Backchannels on the front lines: Emergency uses of social media in the 2007 Southern California Wildfires, in *Proceedings of the 5th International ISCRAM Conference*, Washington, DC, USA, May 2008

Vieweg, S.; Hughes, A.L.; Starbird, K.; Palen, L. (2010). Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In *Proceedings of the 2010 SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, April 2010.

Wang, Y., Ruan, S., Wang, T., & Qiao, M. (2018). Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data. *International Journal of Digital Earth*, 1-20.

Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C. (2019). Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth*, 1-18.

Zhang, Z., Hu, H., Yin, D., Kashem, S., Li, R., Cai, H., ... & Wang, S. (2018). A cyberGIS-enabled multi-criteria spatial decision support system: A case study on flood emergency management. *International Journal of Digital Earth*, 1-18.

Zheng, L., Wang, F., Zheng, X., & Liu, B. (2019). Discovering the relationship of disasters from big scholar and social media news datasets. *International Journal of Digital Earth*, 1-23.

Zou, L., Lam, N. S., Shams, S., Cai, H., Meyer, M. A., Yang, S., ... & Reams, M. A. (2019). Social and geographical disparities in Twitter use during Hurricane Harvey. *International Journal of Digital Earth*, 1-19.