

# An Integrated Crowdsourced Framework for Disaster Relief Distribution

Timothy Schempp  
Department of Geography  
San Diego State University  
San Diego, USA  
tschempp@sdsu.edu

Minsung Hong  
Western Norway Research Institute  
Sogndal, Norway  
msh@vestforsk.no

Haoran Zhang  
Center for Spatial Information Science  
The University of Tokyo  
Tokyo, Japan  
zhang\_ronan@csis.u-tokyo.ac.jp

Rajendra Akerkar  
Western Norway Research Institute  
Sogndal, Norway  
rak@vestforsk.no

Alexander Schmidt  
Department of Econometrics and  
Statistics  
University of Hohenheim  
Stuttgart, Germany  
schmidt.alexander@uni-hohenheim.de

**Abstract**—This paper outlines our ongoing efforts at building an effective disaster relief system utilizing real-time, crowdsourced responses to disaster events. This system integrates three components: social media analysis, rescue demand prediction, and relief distribution optimization. We introduce the conceptual disaster relief system, database structure, and its integrated components. Finally, a summarization of the current research challenges we expect to address in the future is provided.

**Keywords**—disaster relief, disaster management, social media, crowdsourcing, optimization

## I. INTRODUCTION

Countries around the globe have experienced a significant increase in frequency, intensity, and impact of disasters both natural and otherwise over the past decades. While research in the fields of environmental protection and related technology research have already produced many fruitful results to address the impacts of disasters, effective disaster management must remain one of the main responsibilities of governments, especially in areas determined to be vulnerable to disasters in the foreseeable future [1]. Such a disaster-relief system as a national-level priority attempts to mitigate the immense physical destruction of infrastructure, property, and loss of life [2].

However, reliable and practical disaster-relief architecture is a more complex process, which still involves two further technological bottlenecks [3]. The first is the detection, comprehension, and prediction capabilities required to understand a disaster event in real-time. The second is a timely, targeted, and optimal approach to making relief decisions regarding the infrastructure distribution. The traditional methods of unidirectional communications from established organizations to the public seem powerless in dealing with those fuzzy and systematic issues.

The main aim of our research is to support affected regions and responsible institutions by suggesting a novel, social media (SM)-based approach to disaster relief. In order to do so we (i) extract relevant information from Twitter data about rescue demands collected during the disaster situation, (ii) combine these data with authoritative data to predict high-demand areas for rescue operations, and

(iii) optimize disaster relief distribution based on these predictions.

In the following, we will briefly present related work on disaster management in Section II. Since our research is still work-in-progress, Section III describes the ongoing work being undertaken and Section IV an outlook on the next steps in our research. Further challenges are highlighted in Section V. Lastly, Section VI concludes and provides an outlook on future work.

## II. RELATED WORK

With the advent of big data technologies, novel data-driven methods demonstrate a potential breakthrough in real-time disaster relief systems. Although some remarkable successes have benefited the manufacturing, transportation, and finance sectors, the research on data-driven disaster management falls much behind those utilitarian domains. Currently, there are few studies focusing on disaster relief system design [4] and performance evaluation [5], rather more are discussing each sub-issue of disaster management.

Assessing flood vulnerability is a crucial tool for disaster planners and mitigation strategists. Research in this area has been of interest to many domains, and because of its complexity, most studies on the topic utilize an interdisciplinary approach. According to Liu et al. [6], there is a wide variety of factors that must be considered in order to assess vulnerability. These include community factors, such as demographics and socio-economics, in addition to physical factors, like elevation, soil types, drainage characteristics, land cover, and land use [7]. Our research is not focused on advancing these vulnerability prediction models, however, it does attempt to integrate volunteer geographic information (VGI) with these authoritative data to provide a means of validation of our proposed prediction model. While these past studies have been successful in assessing vulnerability, they do not integrate actual demand information or attempt to optimize disaster relief efforts based on their results.

A second crucial tool for disaster planners and mitigation strategists is relief distribution optimization. The importance of this topic is highlighted by an increased interest in it by operations and management researchers [8]. Research addresses this challenge using statistical and

probabilistic models, queuing theory, simulation, decision theory, fuzzy methods, and, most commonly, optimization methods to obtain the optimal locations and capacities of relief infrastructures (e.g., temporary rescue station, alternate care site, and relief shelter, etc.).

Although a great deal of research on vulnerability and relief optimization has been conducted over the past decade, their results tend to exist in silos. Few studies have considered the fusion of these two domains to design a disaster relief system [9]. This integration of domains represents the primary gap in the research that we seek to address.

Research exploring the dynamic relationship between public observations and situational awareness for disaster management typically relies solely on SM data [10, 11, 12]. Latonero and Shklovski [13], for example, highlight the relevance of Twitter data to raise situational awareness. This research focuses on the integration of heterogeneous datasets and seeks to uncover the interrelation between them. Previous algorithms of infrastructure distribution optimization are usually retrospective and oriented to utilitarian domains. In our research, we aim to address the special characteristics of disasters by developing a stochastic algorithm to handle real-time data.

### III. ONGOING RESEARCH

In the first step of our analysis, we process a SM dataset consisting of Tweets collected during a disaster situation in order to extract geographic information related to demand requests or damage reports. There are three possible ways that Tweets can be geographically contextualized. The first is through ‘geo-tagging’ where Tweets are explicitly encoded with latitude and longitude coordinates. The second and third ways have their geographic context *implicitly* encoded, through mentions of physical addresses or points-of-interest (POIs). Due to the lack of a robust POI database, this study focuses on the former implicit encoding. Relevant messages and their physical addresses are extracted using a two-stage filtering process. The first stage utilizes relevant keywords related to demand or damage, which also serve as labels for categorization. The second stage uses regular expressions to filter the previous results based on address patterns. After the relevant messages containing addresses are extracted, they are geocoded in order to transform their implicit geographic context into latitude and longitude coordinates.

After the filtering and processing of the SM data, the next task is to integrate the SM output with authoritative data to predict the locations and quantities of potential relief demands. This requires the retrieval of different authoritative data from various sources. Because the case study disaster phenomenon is related to flooding, a digital elevation model retrieved from the US Geological Survey<sup>1</sup> is used as one input. Additionally, the socio-demographic data for the study area from the US Census Bureau<sup>2</sup> and data on land use is incorporated as the final input for the demand

prediction modeling process. The prediction modeling represents one of our primary bottlenecks in our current research. Because of the sparseness of demand points extracted from SM messages, there is inherent difficulty in predicting the locations that will have the highest demand. Preliminary attempts at this process include the use of linear models and clustering algorithms. Overcoming this bottleneck may require revisiting the first step and implementing all three of the geographic encodings to bolster the dataset.

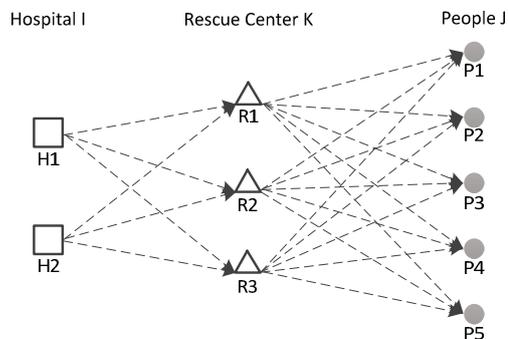


Fig.1 Schematic diagram of the relief system

The schematic in Fig.1 illustrates the conceptual rescue system. After the rescue demand model prediction, we can predict the location and degree of urgency of at-risk people (P1-P5). The locations of the hospitals (H1 and H2) are known in the rescue system, informed by authoritative data. With this information, we can then determine the location of rescue centers (R1, R2, and R3). These locations have a significant impact on the rescue efficiency of the system because these rescue centers are the hubs connecting hospitals and endangered people. Based on known and predicted information, both the management scope and the rescue scope can be optimized, which directly impacts the rescue efficiency of the system.

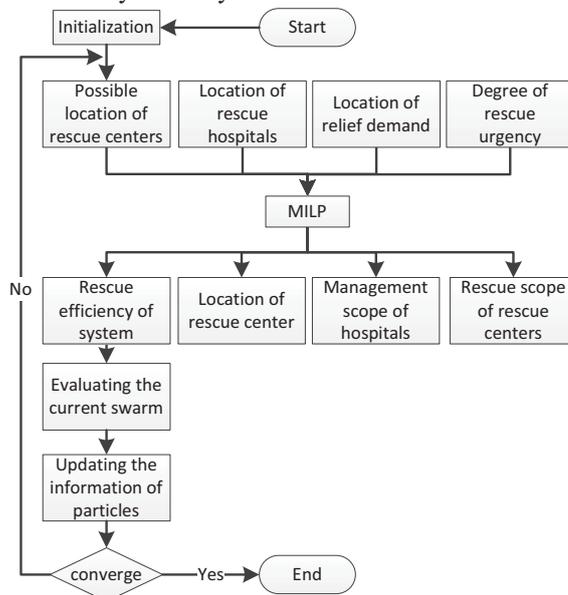


Fig.2 Flowchart of the design of the rescue plan

The dotted lines in Fig.1 represent the possible connection among hospitals, rescue centers, and at-risk people. Each rescue center will connect with only one hospital and similarly, each at-risk individual will be rescued by only one rescue center.

<sup>1</sup> U.S. Geological Survey (2018). National Elevation Dataset (NED): U.S. Geological Survey database available online at <https://viewer.nationalmap.gov/basic/?basemap=b1&category=ned,nedsrct&title=3DEP%20View/>.

<sup>2</sup> U.S. Census Bureau (2016). Total population, 2012-2016 American Community Survey 5-year estimates. Retrieved from [https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS\\_16\\_5YR\\_B01003&prodType=table](https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_B01003&prodType=table).

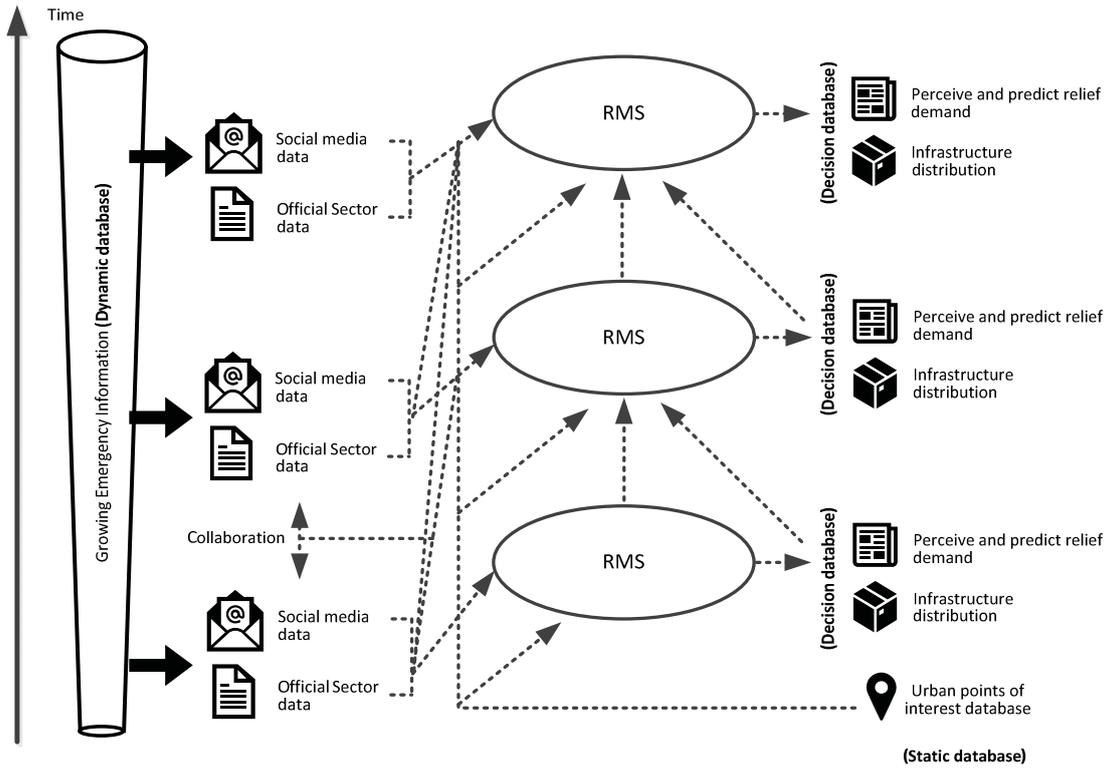


Fig.3 The conceptual disaster relief system, database structure, and integrated components

Finally, the optimal rescue plan with the highest efficiency will be computed, adhering to the above requirement.

The design of the relief distribution optimization scheme is illustrated in Fig.2. Based on a widely-used evolutionary optimization algorithm, particle swarm optimization (PSO) [14], and a mathematical programming model, mixed-integer linear programming model (MILP) [15], the procedure is divided into two steps. First, the possible locations of rescue centers can be determined by PSO, which can generate the location based on the prediction model. Second, an MILP will be established, informed by the PSO output of rescue center locations. Then model parameters will be adjusted by determining the rescue efficiency of the system as the objective function under consideration of existing locations of hospitals and other constraints. Then, the particles in PSO will be updated according to the objective function of the MILP model which is set as the fitness of PSO in each iteration. When the convergence requirements of PSO are met, the optimal rescue plan, including location and rescue scope of rescue centers and management scope of hospitals will be obtained. The solution will be the configuration which has the highest rescue efficiency.

#### IV. OUTLOOK

The further work of this research involves exploring effective methodologies and sustainable solutions to screen the critical and heterogeneous information, perceive and predict relief demand, help managers to make efficient decisions for the relief infrastructure distribution and finally integrate them to construct a practical disaster relief system (DRS). As shown in Fig.3, there are three kinds of distributed databases in the system: a dynamic database, a static database, and a decision database.

The dynamic database will be an important input for the DRS. During a crisis, everybody involved - the public, the media, the government, emergency services, relief organizations, and others - can contribute towards prompt situational awareness. SM data and authoritative data represent the main information channels through which people collectively build awareness, with the advantages of being distributed, far-reaching, and instantaneous. As the disaster evolves, the data quantity and quality will also grow as illustrated by the timeline. Hence, we will integrate those two dynamic data sources as input information of the DRS in order to emulate the disaster development. Considering the nature of the data is highly heterogeneous and from multiple sources with varying levels of quality and correctness, a collaboration mechanism will be established within each development step to cleanse the chaotic information. For this, we will first use the term frequency approach to examine the content of the SM messages. Next follows conceptualization, in which we will iteratively create a keyword/phrase list and categorization dictionary and update it each time. In the third step, we will deal with subsetting and extraction, after which, as a last step, we will geocode the extracted data (addresses) of relevant Tweets into latitudinal and longitudinal information.

The static database will also play a critical role in the DRS. Since the dynamic database alone is not enough to perceive and contextualize the chaotic disaster's status, the mass urban points of interest information (e.g., locations and capacities of each hospital, Red Cross society, and governmental relief center, etc.) will be necessary to support the construction of an DRS. After cleansing and processing the inputs of dynamic and static databases, the next task of the DRS will be to perceive and predict the locations and quantities of potential relief demands which are the primary output to the decision dataset. Finally, based on the

distribution of the demands, visualized with a geographic information system, optimization methods for relief infrastructure distribution will be developed to obtain the optimal locations and capacities of relief infrastructures.

## V. CHALLENGES

One challenge for our research is related to the lack of explicit geographic context (geo-tagging) of the SM messages in the dataset. While there is some research related to the extraction of postal addresses from unstructured text data [16] there is no perfect pattern-based approach because of the non-uniformity found in some addresses. Conversely, the use of natural language processing for identification and categorization of message contents is prevalent in disaster relief research [17, 18]. While it is not uncommon to identify and categorize SM messages for disaster relief, it is practical to additionally capture the implicit geographic context of these messages.

Another challenge that must be overcome relates to the difficulty of rescue demand prediction. As mentioned in the related works, not only are there many perspectives on the definition of vulnerability, there are also huge variances in what data is required to predict it. This translates to difficulty in determining what the significant variables that can be used as predictors are, something that is also highly dependent on the disaster phenomenon studied. Again, the sparseness of the actual demand data makes this prediction process much more difficult and uncertain. Revisiting the demand extraction may be the only way to alleviate this issue.

Finally, applying this research towards processing real-time data is the last challenge that must be addressed if we are to produce a practical, usable result. While in its current stage, we are indeed focused on retrospective data, our methods and approaches are implemented with a real-time component in mind. In terms of optimization, we are currently taking a temporally-staged approach. This way the data we base the optimizations on are not all available at once but rather in intervals.

Despite these challenges, first results, in which we utilize data collected in 2017 during the Hurricane Harvey disaster in the USA, show that our system is able to detect locations of relief demand and optimize the relief distribution layout for these affected areas. Space restrictions do not allow to visualize the optimization outcome here, however, our system optimally places temporary rescue centers within the disaster region. Furthermore, our system takes the temporal aspect of the data into account and reiterates the optimization for each period under consideration of previous optimization periods' outcome and remaining vacancies of rescue stations.

## VI. CONCLUSION

This paper has outlined our ongoing efforts at developing an effective disaster relief system to assist with real-time relief response to a disaster event through the integration of data from SM and authoritative sources. While in the literature, research towards vulnerability assessment and relief distribution optimization has been conducted, the outcomes of these studies exist in silos. The aim of our research is to integrate these two approaches and

develop an operational system that can be used by authorities and first responders. Finally, the paper summarizes our initial findings, our ongoing efforts, as well as a set of challenges that we expect to tackle in the near future.

## ACKNOWLEDGEMENT

The work is funded from the Research Council of Norway (RCN) and the Norwegian Centre for International Cooperation in Education (SiU) grant through INTPART program.

## REFERENCES

- [1] Zhou, Lei, et al. "Emergency decision making for natural disasters: An overview." *International journal of disaster risk reduction* (2017).
- [2] M. Ivgin, The decision-making models for relief asset management and interaction with disaster mitigation, *International Journal of Disaster Risk Reduction* 5 (2013) 107-116.
- [3] Das, Rubel. "Disaster preparedness for better response: Logistics perspectives." *International Journal of Disaster Risk Reduction* 31 (2018): 153-159.
- [4] Chatzimilioudis, Georgios, et al. "Crowdsourcing emergency data in non-operational cellular networks." *Information Systems* 64 (2017): 292-302.
- [5] Xu, Zheng, et al. "Crowdsourcing-based timeline description of urban emergency events using social media." *International Journal of Ad Hoc and Ubiquitous Computing* 25.1-2 (2017): 41-51.
- [6] Liu, Jie, Zhenwu Shi, and Dan Wang. "Measuring and mapping the flood vulnerability based on land-use patterns: a case study of Beijing, China." *Natural Hazards* 83.3 (2016): 1545-1565.
- [7] Saghaian, Bahram, et al. "Flood intensification due to changes in land use." *Water resources management* 22.8 (2008): 1051-1067.
- [8] Anaya-Arenas, Ana Maria, Jacques Renaud, and Angel Ruiz. "Relief distribution networks: a systematic review." *Annals of Operations Research* 223.1 (2014): 53-79.
- [9] Kuo, Yong-Hong, et al. "A Real-Time Decision Support Tool for Disaster Response: A Mathematical Programming Approach." *Big Data (BigData Congress), 2015 IEEE International Congress on. IEEE*, 2015.
- [10] Huang, Qunying, and Yu Xiao. "Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery." *ISPRS International Journal of Geo-Information* 4.3 (2015): 1549-1568.
- [11] Seppänen, Hannes, and Kirsi Virrantaus. Shared situational awareness and information quality in disaster management. *Safety Science* 77 (2015): 112-122.
- [12] Vieweg, Sarah, Carlos Castillo, and Muhammad Imran. "Integrating social media communications into the rapid assessment of sudden onset disasters." *International Conference on Social Informatics*. Springer, Cham, 2014.
- [13] Latonero, Mark, and Irina Shklovski. "Emergency management, Twitter, and social media evangelism." *Using Social and Information Technologies for Disaster and Crisis Management*. IGI Global, 2013. 196-212.
- [14] Zhang, Haoran, et al. "Improved PSO-based Method for Leak Detection and Localization in Liquid Pipelines." *IEEE Transactions on Industrial Informatics* (2018).
- [15] Zhang, Haoran, et al. "A three-stage stochastic programming method for LNG supply system infrastructure development and inventory routing in demanding countries." *Energy* 133 (2017): 424-442.
- [16] Efremova, Julia, et al. "A Geo-Tagging Framework for Address Extraction from Web Pages." *Industrial Conference on Data Mining*. Springer, Cham, 2018.
- [17] Stowe, Kevin, et al. "Identifying and categorizing disaster-related tweets." *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*. 2016.
- [18] Kryvasheyev, Yury, et al. "Rapid assessment of disaster damage using social media activity." *Science advances* 2.3 (2016): e150077