A Framework to Integrate Social Media and Authoritative Data for Disaster Relief Detection and Distribution Optimization

Timothy Schempp∗1, Haoran Zhang†2, Alexander Schmidt3, Minsung Hong4, and Rajendra Akerkar4

1San Diego State University
2University of Tokyo
3University of Hohenheim
4Western Norway Research Institute

April 1, 2019

Abstract

In this paper, we propose an interdisciplinary approach to (natural) disaster relief management. Our framework combines dynamic and static databases, which consist of social media and authoritative data of an afflicted region, respectively, to model rescue demand during a disaster situation. Using Global Particle Swarm Optimization and Mixed-Integer Linear Programming, we then determine the optimal amount and locations of temporal rescue centers. Furthermore, our disaster relief system identifies an efficient distribution of supplies between hospitals and rescue centers and rescue demand points. By leveraging the temporal dimension of the social media data, our framework manages to iteratively optimize the disaster relief distribution.

Keywords: disaster relief, disaster management, social media, optimization

∗Corresponding author: Center for Information Convergence and Strategy, San Diego State University, Los Angeles, USA, e-mail: tschempp@sdsu.edu
†Second Corresponding author: Center for Spatial Information Science, University of Tokyo, Tokyo, Japan, e-mail: zhang_ronan@sis.u-tokyo.ac.jp
1 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 academic research into environmental studies and sustainability has well-characterized the impacts of disasters, the real-world application of disaster management practices remains as one of the primary responsibilities of federal and local governments [1]. It is necessary to ensure the design and implementation of a nation's disaster-relief system are always maximizing efficiency and utilizing all available resources. Regarding this disaster-relief as a national-level priority attempts to mitigate the immense physical destruction of infrastructure, property, and loss of life [2]. Given the complexity of a disaster-relief system, the traditional method of unidirectional communication from established organizations to the public appears to be an antiquated and inefficient approach to disaster relief. By utilizing modern technologies, which enable multi-directional communication between institutions and citizens, advanced systems (e.g. cyber-physical systems) have emerged as a contemporary means of interfacing [3].

Developing a reliable and practical disaster-relief architecture is a complex process which involves several technical challenges [4]. The first challenge is the extraction of rescue demand requests and their location from social media data. The second challenge requires the establishment of a framework for providing decision makers with relief-distribution allocation models that address time-sensitive situational needs. With the advent of big data technologies, data-driven methods now show the potential to overcome these challenges. Despite the remarkable success of big-data applications in the manufacturing, transportation, and finance sectors, research on data-driven disaster management appears to fall behind. A few rare examples can be found in [5] and [6], which focus on both disaster-relief system design and performance evaluation, respectively.

In this paper we establish the framework for an optimal approach to disaster-relief management, which addresses both of the aforementioned challenges. Specifically, the study establishes the emerging value and effective use of near-real time data provided by social media along with various authoritative data now available from ground surveys, cadastral surveys, censuses, and satellite imagery. The acquisition of actionable data from social media is not a primary objective of this research for a variety of reasons.
The adoption rate of a particular social media platform varies greatly depending on the country. The adoption rate also changes over time, what is most popular in a one country today may fall into obscurity over time [7]. Finally, the technical approaches to acquiring the data can be volatile due to monetization of data services or changes in corporate data policy. Because of the sparseness of the actionable data we were able to retrieve, we leverage spatial patterns found within this data and the authoritative data to generate additional, simulated demand points as a pilot study to develop our framework. These actual and simulated demand points then serve as an input for our optimization method. The method we use can be considered a hybrid approach, combining Particle Swarm Optimization (PSO) and Mixed-Integer Linear Programming (MILP), which estimates optimal locations for temporal disaster-relief stations. These temporal stations or rescue centers can be perceived as distribution centers for first-aid and supplies in disaster situations. Thus, finding optimal locations close to the demand points is crucial for effective and efficient disaster management. Since we also consider the temporal dimension of the social media data, this framework holds important practical implications for government institutions in future disaster management.

We proceed as follows. Section 2 provides an overview of research related to the usage of social media and optimization algorithms for disaster management. Section 3 then describes the problem at hand and the framework that we propose for an efficient disaster relief system. The methodology used is elaborated in Section 4 and applied in Section 5, where we use data collected during a disaster that occurred in 2017, when hurricane “Harvey” hit the United States, to illustrate our framework in a practical application. Lastly, Section 6 concludes.

2 Related Work

The upsurge of social media usage, as witnessed over the last decade, holds huge potential for multi-directional communication during disaster situations. Individuals are increasingly using social media to express their needs, opinions, descriptions, and urgency of their current situation. Thus, social media data can serve as an indicator for situational awareness and rescue demand in disaster situations.

The Ushahidi Haiti Project, now infamous due to its role during the Haiti disaster, is
a prime example of how using social media data can help derive situational awareness during a disaster event [8]. This project was one of the first large-scale applications of using social media and authoritative data to populate a crisis map of the 2010 earthquake in Haiti. The data provided by Ushahidi helped to inform various US military organizations as well as private NGOs of the location of communities in need. As [9] point out, the primary limitations are related to the unknown nature and capabilities of the project amongst the humanitarian community and variances in data quality.

Applications of and research on combining social media and authoritative data can be found across a variety of domains, which include geography, public administration, computer science, engineering, environmental studies, and many others. One of the first geographic perspectives on disaster management comes from [10]. Here the authors discuss how Volunteered Geographic Information (VGI) can be provided by individuals via their device’s GPS location when using social media. This additional locational element is a significant factor for providing geographic context towards supporting disaster relief efforts. Logically, unless social media content can be contextualized with geographic location, there is little this data can contribute towards enhancing situational awareness throughout a disaster event. This is what separates VGI from crowdsourced data, as the former enables people to generate actionable geospatial intelligence [11]. One important consideration, however, is the quality of both location and the actual content of the social media data regarding both the certainty and accuracy of message contents and position [10]. Although this is an important aspect in disaster management, this study will not focus on certainty validation. While social media contextualized with location can be used to inform emergency services about individual requests and reports, there is also a great potential to aggregate this data and combine it with other authoritative data sources to characterize the afflicted areas [12].

Authoritative data by itself can benefit disaster management, as can be seen in the assessment of flood vulnerability. Flood vulnerability is a crucial tool for disaster planners and mitigation strategists, for which besides social media data authoritative data are extremely helpful. Research in this area is of interest across many domains, and because of its complexity, most studies on the topic utilize an interdisciplinary approach. According to [13], there are 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 char-
acteristics, and land cover [14, 15]. These flood vulnerability studies that incorporate land-use data into their analysis conclude that areas with high populations and which are heavily residential tend to be the most vulnerable to floods [16]. While these past studies are successful in assessing vulnerability, integrating actual demand information or attempting to optimize disaster relief efforts based on their results are possible extensions that appear to be worthwhile to explore.

Thus, social media and authoritative data constitute a satisfactory basis to visualize the disaster in geographic space. But in order to generate an optimal solution for disaster-relief distribution, those data can be used as input factors for an optimization procedure.

Such optimization of distribution infrastructure is a common topic in operational research [17]. However, only recently have the potential benefits of this approach for disaster-relief management been recognized, seeking to develop a novel solution that balances optimality and promptness [18]. Research in this area uses, among others, statistical and probabilistic models [19, 20], queuing theory [21], simulation [22], decision theory [22, 23], fuzzy methods [24, 25], and, most commonly, optimization methods [26, 27, 28] to obtain the optimal locations and capacities of relief infrastructures (e.g., temporary rescue centers, alternate care sites, and relief shelters). The main weakness of these approaches can be seen in the several hypotheses needed for the optimization. These hypotheses are only imposed after all data have already been collected rendering the optimization rather impractical for timely disaster management. While a temporally-staged integrated system would intuitively have a higher practical relevance for real-world applications, temporally-staged data analysis is also more difficult, since one does not have the benefit of hindsight-algorithms. Thus, it appears that more possibilities need to be explored by combining advanced computation algorithms to implement temporally-staged solutions for those complex problems.

In summary, while a great deal of research on vulnerability and relief optimization has been conducted over the past decade, their results tend to exist in silos. The missing integration of social media and authoritative data represents the primary gap in the research that we seek to address in the following.
3 Framework description

Since the purpose of this research involves the integration of heterogeneous data and rescue-demand predictions with practical implications for disaster-relief management, we provide an overview of the fundamental steps of the resulting disaster relief system (DRS) in Figure 1. As the figure shows, for the problem at hand there are three kinds of distributed databases in the system: a dynamic database, a static database, and a decision database. The dynamic database refers to temporally staged data that serve as one of the inputs for the DRS. During a crisis both the government, emergency services, and relief organizations but also the public and the media can contribute
towards prompt situational awareness. In our framework, social media data represent
the main information channel through which people collectively build awareness, with
the advantages of being distributed, far-reaching, and instantaneous. As a disaster
evolves, the data stemming from this channel will grow in quantity as illustrated by the
vertical timeline on the left of Figure 1. This dynamic data source serves as an input for
the DRS in order to emulate the disaster development. Considering that the nature of
the data is highly heterogeneous and from multiple sources with varying levels of quality
and accuracy, a collaboration mechanism will be established within each development
step to cleanse the information.

The static database will also play a critical role as an additional input to the DRS. Since
the dynamic database alone is not enough to perceive and contextualize the disaster,
authoritative data including urban points of interest information (e.g., locations and
capacities of each hospital, Red Cross society, and governmental relief center) will be
necessary to support the construction of a 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,
opimization methods for relief infrastructure distribution will be developed to obtain
the optimal locations and capacities of temporal relief stations (rescue centers) as well
as their connection to hospitals and responsibilities for demand areas.

Figure 2 should serve to further elaborate on the optimality conditions and capacity
requirements for the rescue centers. As mentioned above, the rescue demand points
($P_1$ to $P_5$) of the dynamic database and the locations of the hospitals ($H_1$ and $H_2$)
serve as input for the rescue system, supported by further authoritative data. With
this information, we can then determine the location of the rescue centers ($R_1$, $R_2$, and
$R_3$). These temporal stations are the hubs connecting hospitals and demand points.
Thus, they have a significant impact on the rescue efficiency of the system. The lines
in 2 represent the possible connection among hospitals, rescue centers, and demand
points. Each rescue center will connect with only one hospital and similarly, each
demand point will receive rescue by only one rescue center. Here, we use the overall
relief distance, i.e. accumulated distance between all demand points and rescue centers
and hospitals in a given period, to evaluate the system’s efficiency. The optimal rescue
Figure 2: Schematic rescue system
This figure provides an overview of the proposed rescue system of our disaster relief system. $H_1$ and $H_2$ represent two examples of hospitals, $R_1$ to $R_3$ are temporal rescue centers that provide supplies to individuals $P_1$ to $P_5$ which are affected by the disaster.

plan with the highest efficiency (shortest distance) will be computed, adhering to the above requirements.

4 Methodology

4.1 Rescue demand

In the first step of our analysis, we process a social media dataset in order to extract geographic information related to demand requests or damage reports. There are three possible ways that social media messages can be geographically contextualized. The
first is through “geo-tagging” where messages are explicitly encoded with latitude and longitude coordinates. The second and third ways have their geographic context implicitly encoded, through mentions of either 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 in four steps:

In the first step, the dataset is explored. We use the term frequency approach to examine the content of the social media messages. This is accomplished by tokenizing the messages into individual words, then parsing through all tokens to create a term frequency table. The table facilitates the identification of relevant keywords, which may vary depending on the disaster phenomenon of interest. These keywords can then be used to subset only the meaningful messages and can also inform labels for categorization.

Conceptualization is done in the second step. With the term frequency table in hand, it is possible to generate a list of keywords and phrases that may indicate messages of reported damages or requests for rescue and supplies. Some terms such as “water” are very ambiguous when observed in isolation, so phrases such as “clean water” or “standing water” are better indicators of supply requests or damage reports, respectively. The creation of the keyword/phrase list and categorization dictionary is an iterative process that is appended and updated each time. This requires starting with a conservative set of keywords and stopwords, then performing the subset, and refining the keywords and stopwords based on this result.

Step three deals with subsetting and extraction. Besides using meaningful keywords, key phrases, and stopwords for extraction, we must also capture those messages that contain the implicit geographic context. As previously mentioned, this research focuses on those messages that contain specific address mentions in addition to information regarding supply demands or damage reports. Because mailing addresses in the United States are relatively uniform in their format, regular expressions provided a viable solution for this additional filtering process. Two patterns are utilized, one that accounts for a single street name such as “123 main street”, and another that would handle double street names, “123 east main street”. Although there are potentially many other combinations for address formats, formulating even more robust regular expressions is not the focus of this research. If a particular tweet meets the keyword/stopword criteria and matches one of the regular expression patterns, it is saved in a new dataset. This dataset contains
Finally, the fourth and last step geocodes the data. In order to transform this new dataset into an actual geographic dataset that could be mapped, a geocoding procedure must be performed. Geocoding refers to the process of transforming a mailing address into geographic coordinates, i.e. its latitude and longitude. This can be accomplished through different application programming interface (API) calls, through Geographic Information System (GIS) software, or even through a spatial database. After the data is transformed into spatial data, it is then necessary to re-project the data into a format that can be reasonably processed by the optimization algorithms. As these algorithms utilize cartesian coordinates, using a State Plane or Universal Transverse Mercator (UTM) projection was appropriate. In our case, the Texas State Plane S Central 4204\footnote{All information taken from https://www.e-education.psu.edu/natureofgeoinfo/c2_p26.html and last accessed on July 27th, 2018.} was the logical decision as these projections are less susceptible to distortions when calculating distances, areas, and shapes than just using an UTM projection.

After the filtering and processing of the social media data, we were left with a sparse dataset of 69 points where rescue demand was requested. There are many possible explanations as to the sparseness of this outcome, such as the removal of address information post-disaster or the fact that we only used the address-extraction approach and not point of interest mentions. Because this paper is not focused on the data acquisition component, we generated additional data points to proceed with a simulated dataset, resulting in 150 data points. Therefore, the next task was to utilize the geocoded rescue demand points and the authoritative data to generate additional data for the simulated model. This requires the retrieval of different authoritative data from various sources and is highly specific to the application at hand, thus, we refer the reader to Section 5 for more details on the data used in our empirical application.

4.2 Optimization

As elaborated in Section 3, the aim of our relief distribution system is to optimize the infrastructure of the relief distribution. Figure 3 shows in greater detail the flow chart of our optimization design.
The dashed boxes represent the input dataset of the method, and the box in the bottom right corner represents the output. First, social media data serves as an input for the rescue demand detection model to detect the time and location information of rescue demands. Then, based on the detection result and the hospital information, Particle Swarm Optimization (PSO) is employed to generate an initial location result of rescue centers. With these results, we can establish a Mixed-Integer Linear Programming (MILP) model to work out the optimal rescue plan and the total rescue distance. We introduce the obtained total rescue distance as the fitness function value of the PSO and adjust the location of rescue centers until the result can meet with the condition of convergence.

In the following, let us provide a few more details about the optimization process. To begin with, PSO is a widely used nature-inspired optimization algorithm based on an overall population. It proves to be quite robust with a good global accuracy [29, 30]. PSO is originally intended for simulating social behavior, as a representation of the
foraging behavior in a bird flock or fish school. Each member of the population is named as particle, representing a potential feasible solution. Each particle has its own fitness value which is depended on the objective function. Relatively, the location of the food represents the global optimal solution. All particles search for the global optimal solution in the solution space.

In the search process, the best position of each particle (personal best) will be recorded and each particle could get the position of the particle closest to the optimal solution in the population (population best). In order to find the optimal solution, each particle will update its position by learning from the personal best and population best and eventually approaches the optimal solution which is reflected in the convergence of the search process [31].

There are two factors influencing the convergence speed in the initial version of PSO. One is exploration and the other is development. Exploration refers to the fact that particles leave the original search trajectory and search in a new direction, which demonstrates the ability to develop into unknown regions, just like global search. Development refers to further detailed search of particles on the original trajectory, just like local search. The strong exploration ability of a population can accelerate convergence and reduce the possibility of falling into a local optimal solution, but at the same time, the accuracy of calculation is relatively low. On the contrary, the strong development ability of the population can improve calculation accuracy by conducting detailed research around the feasible solution, but the search process converges relatively slowly.

In order to accelerate convergence and improve accuracy, numerous researchers have proposed improved PSO algorithms. By introducing the concept of inertia weights to effectively control the exploration and development process, Global PSO (GPSO) and Local PSO (LPSO) are two of these alternatives [32]. The value of inertia weights can alter the global and local search ability of PSO. When the value of an inertial weight is relatively high, the global search ability of particles is strong and the local search ability is weak. On the contrary, when the value of an inertial weight is relatively small, the local search ability of particles is strong, but the global search ability is weak. Due to the different focus of search ability, the topological structure of particles’ neighbors is different. This structure is called All topology in GPSO, whereas the topological structure of particles’ neighbors in LPSO include Ring topology, Four Cluster topology, Pyramid topology, and Square topology.
Inspired by the phenomenon of symbiosis in natural ecosystems, another PSO method, called Multi-Swarm Cooperative PSO (MCPSO), is proposed by [33]. MCPSO is based on a master-slave model, in which a population consists of one master swarm and several slave swarms. The slave swarms execute a single PSO or its variants independently to maintain the diversity of particles, while the master swarm evolves based on its own knowledge and also the knowledge of the slave swarms. In MCPSO, the evolution of each particle is not only influenced by the information of its own group, but also influenced by that of the symbiosis group. In this way, the possibility of getting into the local optimal solution caused by individual information misjudgment is reduced.

In order to consider the impact of structural heterogeneity on individual behavior, Selectively-Informed PSO (SIPSO) is proposed by [34], in which the particles choose different learning strategies based on their connections: a densely-connected hub particle gets full information from all of its neighbors while a non-hub particle with few connections can only follow a single yet best-performed neighbor. Therefore, SIPSO usually outperforms the initial PSO in success rate, solution quality, and convergence speed.

We will test all four methods in a small simulation study in order to choose the most promising one for our framework.

As the second step in our optimization model, based on the location of the possible rescue centers, a Mixed-Integer Linear Programming (MILP) model is set up by using the rescue efficiency of the current system as the objective function and taking the existing location of hospitals and externally given restrictions into consideration. In the following, we elaborate the underlying algorithm in greater detail.

The efficiency of the system reflects how much time it needs to rescue the trapped people (represented by demand points) and the less time it needs, the higher its efficiency. Then, the total time needed in the rescue system is composed of four parts as shown in Equation (1): the time needed to set up the rescue centers \( f_{1k} \), the time needed to transport goods and materials from hospital \( i \) to a possible location of rescue center \( k \) \( f_{2k,i} \), the time needed to travel from a possible location of rescue center \( k \) to demand point \( j \) \( f_{3j,k} \), and the time needed to travel from hospital \( i \) to demand point \( j \) \( f_{4j,i} \).
\[ \min f = \sum_k f_{1k} + \sum_i \sum_k f_{2k,i} + \sum_k \sum_j f_{3j,k} + \sum_i \sum_j f_{4j,i} \quad i \in I, k \in K, j \in J, \quad (1) \]

where \( i \in I = \{1, 2, \ldots, i_{\max}\} \) is the set of all hospitals in the rescue system, \( k \in K = \{1, 2, \ldots, k_{\max}\} \) is the set of all possible locations for temporal rescue centers in the rescue system, and \( j \in J = \{1, 2, \ldots, j_{\max}\} \) is the set of all demand points in the rescue system.

Setting up the temporary rescue centers takes some time. And the time is different to set up the rescue centers at different regions. By Equation (2), the time needed to set up the centers can be obtained.

\[ f_{1k} = P_{U_k} \cdot B_{W_k} \quad k \in K, \quad (2) \]

where \( P_{U_k} \) is the time needed to set up a rescue center at location \( k \), and \( B_{W_k} \) is an indicator variable, which equals 1 if a rescue center is set up at location \( k \), 0 otherwise.

Next, by Equation (3) we can determine the time needed to transport medicine and rescue equipment from hospitals to rescue centers. For each hospital, the speed used to transport goods and materials to rescue centers that are located in different regions is different.

\[ f_{2k,i} = B_{SM_{k,i}} \cdot \frac{D_{k,i}}{V_{k,i}} \quad i \in I, k \in K, \quad (3) \]

where \( B_{SM_{k,i}} \) is an indicator variable which equals 1 if goods or materials are transported from hospital \( i \) to location \( k \), 0 otherwise, and the term \( \frac{D_{k,i}}{V_{k,i}} \) is the ratio of the distance between hospital \( i \) and possible location of the rescue center at \( k \), \( D_{k,i} \), and the average speed by which material travels from \( i \) to \( k \), \( V_{k,i} \).

Similarly, the time needed to arrive at the demand points from rescue centers and hospitals can be determined by Equation (4) and (5), respectively. The degree of urgency for each demand point is expressed by \( \alpha_j \). The larger the value of \( \alpha_j \), the more
urgent it is to rescue individual $j$.

$$f_{3,j,k} = \alpha_j B_{SN,j,k} \cdot \frac{D_{j,k}}{V_{j,k}} \quad j \in J, k \in K$$

(4)

$$f_{4,j,i} = \alpha_j B_{S,j,i} \cdot \frac{D_{j,i}}{V_{j,i}} \quad j \in J, i \in I,$$

(5)

where $B_{SN,j,k}$ ($B_{S,j,i}$) is an indicator variable, which equals 1 if location $k$ (hospital $i$) provides rescue service for demand point $j$, 0 otherwise, and the term $\frac{D_{j,k}}{V_{j,k}}$ ($\frac{D_{j,i}}{V_{j,i}}$) is the ratio of the distance between the position of rescue center $k$ (hospital $i$) and a demand point $j$, $D_{j,k}$ ($D_{j,i}$), and the average speed of rescue services between the location of the rescue center $k$ (hospital $i$) and the demand point $j$, $V_{j,k}$ ($V_{j,i}$).

For a successful optimization, we need several constraints on the previously mentioned parameters, many of which are straightforward. Equations (6) and (7) constrain the flow of goods and materials between hospitals and possible locations of rescue centers: if no rescue center is set up at location $k$, no goods and materials will be deployed from any hospital. On the contrary, if a rescue center is set up at location $k$, it will receive goods and materials from only one hospital while one hospital could provide supplies for several rescue centers.

$$\sum_i B_{SM,k,i} = B_{W,k} \quad k \in K, i \in I$$

(6)

$$\sum_i B_{SM,k,i} \geq 1 + (B_{W,k} - 1) \cdot M \quad k \in K, i \in I,$$

(7)

where $M$ is large, positive integer.

Furthermore, Equation (8) introduces a rather obvious restriction: only when a rescue center is set up at location $k$ can it provide rescue services for nearby demand points. In addition, each demand point must be rescued by one center or one hospital which is constrained by Equation (9).
\[
\sum_j B_{SN_{j,k}} \leq B_{W_k} \cdot M \quad j \in J, k \in K \tag{8}
\]

\[
\sum_i B_{S_{j,i}} + B_{SN_{j,k}} = 1 \quad j \in J, k \in K \tag{9}
\]

Considering that there is only a limited amount of beds available per hospital to provide rescue service for disaster victims, the number of victims taken to one hospital should be less than the hospital’s maximum capacity, as expressed by Equation (10). The constraints given by Equation (11) and (12) refer to the indicator variable \( B_{SMN_{i,k,j}} \), which equals 1 if all victims at demand point \( j \) can be taken to hospital \( i \) via rescue center \( k \), 0 if not.

\[
\sum_k \sum_j P_{E_j} B_{SMN_{i,k,j}} + \sum_j P_{E_j} B_{S_{i,j}} \leq C_{A_i} \quad k \in K, j \in J, \tag{10}
\]

\[
B_{SMN_{i,k,j}} \leq \frac{B_{SM_{i,k}} + B_{SN_{k,j}}}{2} \quad i \in I, j \in J, k \in K, \tag{11}
\]

\[
B_{SMN_{i,k,j}} \geq \frac{B_{SM_{i,k}} + B_{SN_{k,j}} - 1}{2} \quad i \in I, j \in J, k \in K, \tag{12}
\]

where \( P_{E_j} \) is the number of victims at demand point \( j \) and \( C_{A_i} \) the number of beds at hospital \( i \).

Lastly, as the rescue proceeds, more and more demand points will emerge. This means more temporary rescue centers need to be set up. When planning the layout of new rescue centers, the centers that have been set up should be taken into consideration since they do not simply disappear but can be used for the ongoing disaster relief. This is expressed by the constraint in Equation (13). To these centers, the relationship with hospitals and demand points is different from that of the new centers. By the constraints in Equations (14) and (15), only when there are demand points taken to the rescue centers will goods and materials be sent to the centers.
\[ B_{W_{k_1}} = 1 \leq C_{A_i}, \quad k_1 \in K_1 \] (13)

\[ \sum_i B_{SM_{k,i}} \leq \sum_j B_{SN_{k,j}} \cdot M \quad i \in I, j \in J, k \in K \] (14)

\[ \sum_i B_{SM_{k,i}} \cdot M \geq \sum_j B_{SN_{k,j}} \quad i \in I, j \in J, k \in K \] (15)

where \( k_1 \in K_1 = \{1, 2, \ldots, k_{1_{max}}\} \) is the set of locations where rescue centers have already been set up.

### 4.3 Model Testing

In order to test the four different PSO methods, we conduct a small simulation study. We use the 4,600 square km area, 150 demand points, 15 hospitals, and the number of possible locations for rescue centers. Although we use the small dataset for our simulation study, it is worth to note that the demand points are extracted by the conceptualization from overall social media data (i.e., 8.5 million tweets) generated during Hurricane Harvey, since social media data usually include a lot of feeds/tweets having no relation with disasters [35, 36, 37], as mentioned in the Section 4.1 and 5.1. The coordinates of the hospitals and demand points are randomly generated and known. Matlab R2015a with Gurobi solver on a computer with Interl(R) Core(TM) i5-4200U CPU @ 1.60GHz is employed to solve the system by using the previously mentioned four types of PSO. The results are shown below. Each type of PSO is tested four times. Convergence curves of the system are shown in Figure 4. Calculation time for each test run of a PSO is shown in Table 1.
Figure 4: Convergence curves of the PSO methods
These plots show the convergence curves for the GPSO in panel (a), the LPSO in panel (b), the MCPSO in panel (c), and the SIPSO in panel (d).

Table 1: Calculation time for each test of the PSO
This table shows the performance of the four different models tested in our PSO simulation. Each model is tested four times, resulting run times are noted in columns 2 to 5. The last column shows the average run time of each model over all four test runs. Unit of measurement is seconds.

| Model | Run 1 | Run 2 | Run 3 | Run 4 | Average |
|-------|-------|-------|-------|-------|---------|
| GPSO  | 488.7 | 732.6 | 506.4 | 497.7 | 556.4   |
| LPSO  | 249.1 | 107.1 | 203.8 | 101.6 | 165.4   |
| MCPSO | 249.2 | 369.0 | 381.6 | 290.5 | 322.6   |
| SIPSO | 105.8 | 111.6 | 109.6 | 116.8 | 111.0   |
According to the simulation results, GPSO appears to be best suited for our relief distribution system, since it displays a comparatively robust convergence across all test runs.

5 Empirical application

In order to demonstrate our framework for efficient disaster relief management, we apply the proposed model to data collected in August 2017, when Hurricane Harvey hit the United States. Hurricane Harvey was a category 4 storm that hit Texas on August 26, 2017. According to the National Hurricane Center (NHC), it is the second ranked costliest storm with a total estimated economic cost of $125 billion (tied with the 2005 Hurricane Katrina) and death estimates between 68 and 89 as reported by the NHC and National Oceanic and Atmospheric Administration. Analysis by the Kinder Institute showed that almost 30% of Houston’s population was impacted by the storm.

5.1 Data

The social media dataset used for our analysis consists of over 8.5 million tweets that were extracted using the keywords “Harvey” and “hurricane Harvey”. These tweets are collected over a period of three days August 27th through the 29th, 2017. However, none of the tweets in the dataset are geo-tagged, which means that the geographic context must be extracted from the message contents. As a result, the actual data on unambiguously identified demand points is rather sparse, with only 69 points extracted. Thus, additional data are generated to increase the complexity of the problem. This process relies upon two datasets retrieved from authoritative data sources. The first is a digital elevation model (10-meter), retrieved from the US Geological Survey, which is used to derive the stream network seen in Figure 5. The second is the land use

\[\text{Dataset 1: Digital Elevation Model (10-meter)}\]

\[\text{Dataset 2: Land Use}\]

2 All information taken from https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf and last accessed on July 27th, 2018.

3 Data can be accessed via https://ricegis.maps.arcgis.com/apps/Cascade/index.html?appid=6ea5082d69484c7a922bd18705afbf85 and were last accessed on July 27, 2018.
As illustrated in Figure 5, the distribution of rescue demand points strongly corresponds to the location of the stream network and its surrounding area. Taking this into consideration, we are able to generate additional data that realistically represent rescue demand points for each of the three days. We use these three days as the temporal “steps” for the input of our optimization model.

In order to generate additional rescue demand points, we consider the temporal-distribution of the actual twitter-derived data as well as proximity of the demand points to the stream network. Finally, as mentioned in the literature on land-use patterns, specifically areas designated as residential are more likely to correspond to flood areas that contain vulnerable populations [16]. Through overlay analysis, we are thus able to generate additional points where more rescue demand requests are likely to occur. This bolsters our dataset such that performing temporal optimization could resolve a meaningful output.

The final authoritative data used in this study focuses on the hospital locations. This data is available from the US Department of Homeland Security’s open spatial data platform, the Homeland Infrastructure Foundation-Level Data (HIFLD). The data consists of a nation-wide table of hospitals as well as their location, category, number of beds, and many other attributes. From this dataset, we subset the data to only those hospitals that were located within Harris County. From this subgroup, we utilized the top 15 hospitals (according to their number of beds) for input to the optimization.

### 5.2 Results

As for our simulation in Subsection 4.3, Matlab R2015a with Gurobi solver on a computer with Interl(R) Core(TM) i5-4200U CPU @ 1.60GHz is employed to solve the system by GPSO. The results are described in the following.

On the first day, 28 demand points are recorded. Two temporal rescue centers, represented by triangularly shaped objects in the bottom panel of Figure 6 should be set up to satisfy the demand. Once the demand is satisfied, the capacities of two out of the 15
hospitals in Harris County are already exhausted. As elaborated above, these hospitals will not participate in the DRS in future periods.

On the second day, 96 demand points with victims are recorded. According to the results of the GPSO, seven new rescue centers should be set up and one old rescue center joins the network, too. Satisfying all rescue demands leads another hospital to drop out of the system, since all of its free beds are taken by victims.

**Figure 5:** This figure shows the distribution of Twitter-derived rescue demand points and the Digital Elevation Model (DEM)-derived stream network. Demand points occurring at day 1 are depicted in light gray, those occurring at day 2 in gray with a black dot in their center, and demands points of day three are shown as black dots.
Figure 6: Optimization results: Day 1
These plots show the temporally-staged optimization results for day 1: the upper panel shows the status quo at the beginning of the day, while the lower panel shows the situation after all temporal rescue stations are set up and rescue demands are dealt with. Black squares represent permanent hospitals. Hospital capacity is illustrated by the size of the respective square (small: < 360 beds; large: < 800 beds; highlighted square: 4000 beds). Hospitals with no more free capacities are represented with filled squares. Triangles represent temporary rescue centers and gray dots demand points.
Figure 7: Optimization results: Day 2
These plots show the temporally-staged optimization results for day 2: the upper panel shows the status quo at the beginning of the day, while the lower panel shows the situation after all temporal rescue stations are set up and rescue demands are dealt with. Black squares represent permanent hospitals. Hospital capacity is illustrated by the size of the respective square (small: < 360 beds; large: < 800 beds; highlighted square: 4000 beds). Hospitals with no more free capacities are represented with filled squares. Triangles represent temporary rescue centers and gray dots demand points.
Figure 8: Optimization results: Day 3
These plots show the temporally-staged optimization results for day 3: the upper panel shows the status quo at the beginning of the day, while the lower panel shows the situation after all temporal rescue stations are set up and rescue demands are dealt with. Black squares represent permanent hospitals. Hospital capacity is illustrated by the size of the respective square (small: < 360 beds; large: < 800 beds; highlighted square: 4000 beds). Hospitals with no more free capacities are represented with filled squares. Triangles represent temporary rescue centers and gray dots demand points.
On the third day, 26 more locations with victims are recorded. Two new rescue centers are needed and five old rescue centers still participate in the network. After this last day, about half of the hospitals in our data set have exhausted their capacities.

6 Conclusion

Effective disaster relief systems will play an increasingly critical role in the future for supporting decision makers. To this end, several challenges remain, such as the accurate comprehension, detection, and prediction of the disaster state, in addition to dynamic and real-time optimization for the relief planning. This research has sought to conceptualize and develop an integrated disaster relief system (DRS) by combining natural language processing techniques, VGI acquisition, PSO, and MILP. Ingesting huge volumes of heterogeneous and unstructured datasets, the system is able to detect locations of relief demand and optimize the relief distribution layout for these affected areas. This study and the produced DRS attempt to bridge domain gaps, enrich disaster management theory, and yield societal benefits.

However, as mentioned in Section 2, there are many definitions of disaster vulnerability, and therefore huge variances exist in what data is required to predict it. This translates to difficulty in determining what are the significant variables that can be used as predictors, something that is also highly dependent on the disaster phenomenon of study. Again, the sparseness of the actual demand data available on social media makes this prediction process much more difficult and uncertain. In this study, we focused on data from only one disaster for our empirical application, which may cause an inaccuracy of the relief detection. To address this issue in the future, one can expand the study areas to several collected datasets of similar disasters. Furthermore, incorporating these datasets to perform more detailed contrastive analysis and feature extraction may prove worthwhile, as [38] show.

While this research can be considered only as a starting point, our real-world application shows how these tools and systems can be an invaluable resource for decision makers. Each disaster event is unique in that it affects various countries, locations, topographies that all have different characteristics. The disaster phenomenon itself, whether it be
hurricanes, earthquakes, wildfires, etc., has distinct impacts on local populations that require a particular response from disaster relief agencies. These complexities make the creation of a universal DRS a very ambitious goal, something that requires a huge undertaking to accomplish. It requires interdisciplinary teams consisting of experts who utilize GIS, optimization, and NLP who are proficient in many different languages. These teams also rely upon domain experts for each particular disaster, which could include climatologists, meteorologists, hydrologists, and geologists. Finally, input and direction from the emergency management community ultimately drives what information will be required to make important decisions. Then, this information can be conveyed effectively by visualization experts, web developers, and web-GIS platforms. This effort depends upon the integration of many future and past research endeavors which involve disaster-specific emergency management approaches, the extraction of relevant VGI from social media data, and disaster relief optimization. With the framework that we propose in this paper we established a solid foundation on which such future efforts can be built upon.

7 Acknowledgements

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] L. Zhou, X. Wu, Z. Xu, H. Fujita, Emergency decision making for natural disasters: An overview, International journal of disaster risk reduction doi:https://doi.org/10.1016/j.ijdrr.2017.09.037.

[2] M. İvgin, The decision-making models for relief asset management and interaction with disaster mitigation, International Journal of Disaster Risk Reduction 5 (2013) 107–116. doi:https://doi.org/10.1016/j.ijdrr.2013.08.005.

[3] K. Villela, C. Nass, R. Novais, P. Simões, A. Traina, J. Rodrigues, J. M. Menendez, J. Kurano, T. Franke, A. Poxrucker, Reliable and smart decision support system for emergency management based on crowdsourcing information, in: Exploring Intelligent Decision Support Systems, Springer, 2018, pp. 177–198. doi:https://doi.org/10.1007/978-3-319-74002-7_9.

[4] R. Das, Disaster preparedness for better response: Logistics perspectives, International Journal of Disaster Risk Reduction 31 (2018) 153–159. doi:https://doi.org/10.1016/j.ijdrr.2018.05.005.

[5] G. Chatzimilioudis, C. Costa, D. Zeinalipour-Yazti, W.-C. Lee, Crowdsourcing emergency data in non-operational cellular networks, Information Systems 64 (2017) 292–302. doi:https://doi.org/10.1016/j.is.2015.11.004.

[6] Z. Xu, H. Zhang, C. Hu, Y. Liu, J. Xuan, L. Mei, 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. doi:https://doi.org/10.1504/IJAHUC.2017.083481.

[7] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, J. N. Rosenquist, Understanding the demographics of twitter users, in: Fifth international AAAI conference on weblogs and social media, 2011.

[8] J. Heinzelman, C. Waters, Crowdsourcing crisis information in disaster-affected Haiti, US Institute of Peace Washington, DC, 2010.

[9] N. Morrow, N. Mock, A. Papendieck, N. Kocmich, Independent evaluation of the
ushahidi haiti project, Development Information Systems International 8 (2011) 2011.

[10] M. F. Goodchild, J. A. Glennon, Crowdsourcing geographic information for disaster response: a research frontier, International Journal of Digital Earth 3 (3) (2010) 231–241. doi:https://doi.org/10.1080/17538941003759255.

[11] M. F. Goodchild, Citizens as sensors: the world of volunteered geography, GeoJournal 69 (4) (2007) 211–221. doi:https://doi.org/10.1007/s10708-007-9111-y.

[12] J. P. De Albuquerque, B. Herfort, A. Brenning, A. Zipf, A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, International Journal of Geographical Information Science 29 (4) (2015) 667–689. doi:https://doi.org/10.1080/13658816.2014.996567.

[13] J. Liu, Z. Shi, D. 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. doi:https://doi.org/10.1007/s11069-016-2375-0.

[14] A. H. Thieken, M. Müller, H. Kreibich, B. Merz, Flood damage and influencing factors: New insights from the august 2002 flood in germany, Water resources research 41 (12). doi:https://doi.org/10.1029/2005WR004177.

[15] B. Saghafian, H. Farazjoo, B. Bozorgy, F. Yazdandoost, Flood intensification due to changes in land use, Water resources management 22 (8) (2008) 1051–1067. doi:https://doi.org/10.1007/s11269-007-9210-z.

[16] C. Armenakis, E. X. Du, S. Natesan, R. A. Persad, Y. Zhang, Flood risk assessment in urban areas based on spatial analytics and social factors, Geosciences 7 (4) (2017) 123. doi:https://doi.org/10.3390/geosciences7040123.

[17] H. Zhang, Y. Liang, Q. Liao, X. Yan, Y. Shen, Y. Zhao, A three-stage stochastic programming method for lng supply system infrastructure development and inventory routing in demanding countries, Energy 133 (2017) 424–442. doi:https://doi.org/10.1016/j.energy.2017.05.090.

[18] T. R. Robinson, N. J. Rosser, A. L. Densmore, J. G. Williams, M. E. Kincey,
J. Benjamin, H. J. Bell, Rapid post-earthquake modelling of coseismic landsliding intensity and distribution for emergency response decision support., Natural hazards and earth system sciences. 17 (9) (2017) 1521–1540. doi:https://doi.org/10.5194/nhess-17-1521-2017.

[19] X. Xu, Y. Qi, Z. Hua, Forecasting demand of commodities after natural disasters, Expert Systems with Applications 37 (6) (2010) 4313–4317. doi:https://doi.org/10.1016/j.eswa.2009.11.069.

[20] S. Coles, L. Pericchi, Anticipating catastrophes through extreme value modelling, Journal of the Royal Statistical Society: Series C (Applied Statistics) 52 (4) (2003) 405–416. doi:https://doi.org/10.1111/1467-9876.00413.

[21] J. R. Artalejo, G-networks: A versatile approach for work removal in queueing networks, European Journal of Operational Research 126 (2) (2000) 233–249. doi:https://doi.org/10.1016/S0377-2217(99)00476-2.

[22] W. Hu, Y. Qing, Y. Ming-Hui, F. Qi, Grid-based platform for disaster response plan simulation over internet, Simulation Modelling Practice and Theory 16 (3) (2008) 379–386. doi:https://doi.org/10.1016/j.simpat.2008.01.001.

[23] H. Jia, F. Ordóñez, M. Dessouky, A modeling framework for facility location of medical services for large-scale emergencies, IIE transactions 39 (1) (2007) 41–55. doi:https://doi.org/10.1080/07408170500539113.

[24] L. Cret, F. Yamazaki, S. Nagata, T. Katayama, Earthquake damage estimation and decision analysis for emergency shut-off of city gas networks using fuzzy set theory., Structural safety 12 (1) (1993) 1–19.

[25] H. Tamura, K. Yamamoto, S. Tomiyama, I. Hatono, Modeling and analysis of decision making problem for mitigating natural disaster risks, European Journal of Operational Research 122 (2) (2000) 461–468. doi:https://doi.org/10.1016/S0377-2217(99)00247-7.

[26] S. Yan, Y.-L. Shih, Optimal scheduling of emergency roadway repair and subsequent relief distribution, Computers & Operations Research 36 (6) (2009) 2049–2065. doi:https://doi.org/10.1016/j.cor.2008.07.002.
[27] A. Bozorgi-Amiri, M. S. Jabalameli, M. Alinaghian, M. Heydari, A modified particle swarm optimization for disaster relief logistics under uncertain environment, The International Journal of Advanced Manufacturing Technology 60 (1-4) (2012) 357–371. doi:https://doi.org/10.1007/s00170-011-3596-8.

[28] H. Wang, L. Du, S. Ma, Multi-objective open location-routing model with split delivery for optimized relief distribution in post-earthquake, Transportation Research Part E: Logistics and Transportation Review 69 (2014) 160–179. doi:https://doi.org/10.1016/j.tre.2014.06.006.

[29] H. Zhang, Y. Liang, W. Zhang, N. Xu, Z. Guo, G. Wu, Improved pso-based method for leak detection and localization in liquid pipelines, IEEE Transactions on Industrial Informatics doi:https://doi.org/10.1109/TII.2018.2794987.

[30] H. Zhang, Y. Liang, J. Ma, Y. Shen, X. Yan, M. Yuan, An improved pso method for optimal design of subsea oil pipelines, Ocean Engineering 141 (2017) 154–163. doi:https://doi.org/10.1016/j.oceaneng.2017.06.023.

[31] H. Zhang, M. Yuan, Y. Liang, Q. Liao, A novel particle swarm optimization based on prey–predator relationship, Applied Soft Computing 68 (2018) 202–218. doi:https://doi.org/10.1016/j.asoc.2018.04.008.

[32] R. Mendes, J. Kennedy, J. Neves, Watch thy neighbor or how the swarm can learn from its environment, in: Swarm Intelligence Symposium, 2003. SIS’03. Proceedings of the 2003 IEEE, IEEE, 2003, pp. 88–94. doi:https://doi.org/10.1109/SIS.2003.1202252.

[33] B. Niu, Y. Zhu, X. He, H. Wu, Mcpso: A multi-swarm cooperative particle swarm optimizer, Applied mathematics and computation 185 (2) (2007) 1050–1062. doi:https://doi.org/10.1016/j.amc.2006.07.026.

[34] Y. Gao, W. Du, G. Yan, Selectively-informed particle swarm optimization, Scientific reports 5 (2015) 9295. doi:https://doi.org/10.1038/srep09295.

[35] T. Spielhofer, R. Greenlaw, D. Markham, A. Hahne, Data mining twitter during the uk floods: Investigating the potential use of social media in emergency management, in: Proceedings of the 3rd International Conference on Information and
Communication Technologies for Disaster Management (ICT-DM), IEEE, 2016, pp. 1–6.

[36] D. Pohl, A. Bouchachia, H. Hellwagner, Online indexing and clustering of social media data for emergency management, Neurocomputing 172 (2016) 168–179. doi:10.1016/j.neucom.2015.01.084.

[37] J. Stephenson, M. Vaganay, D. Coon, R. Cameron, N. Hewitt, The role of facebook and twitter as organisational communication platforms in relation to flood events in northern ireland, Journal of Flood Risk Management 11 (3) (2018) 339–350. doi:10.1111/jfr3.12329.

[38] S. Han, H. Huang, Z. Luo, C. Foropon, Harnessing the power of crowdsourcing and internet of things in disaster response, Annals of Operations Research (2018) 1–16 doi:10.1007/s10479-018-2884-1.