D-record: Disaster Response and Relief Coordination Pipeline

Shruti Kar[1*], Hussein S. Al-Olimat[1*], Krishnaprasad Thirunarayan[1], Valerie L. Shalin[1],
Amit Sheth[1], Srinivasan Parthasarathy[2]

[1]Kno.e.sis, Wright State University, Dayton, Ohio
[2]The Ohio State University, Columbus, Ohio

{shruti,hussein,tkprasad,valerie,amit}@knoesis.org;   srini@cse.ohio-state.edu

ABSTRACT

We employ multi-modal data (i.e., unstructured text, gazetteers, and imagery) for location-centric demand/request matching in the context of disaster relief. After classifying the *Need* expressed in a tweet (the WHAT), we leverage OpenStreetMap to geolocate that *Need* on a computationally accessible map of the local terrain (the WHERE) populated with location features such as hospitals and housing. Further, our novel use of flood mapping based on satellite images of the affected area supports the elimination of candidate resources that are not accessible by road transportation. The resulting map-based visualization combines disaster-related tweets, imagery and pre-existing knowledge-base resources (gazetteers) to reduce decision-making latency and enhance resiliency by assisting individual decision-makers and first responders for relief effort coordination.

CCS CONCEPTS

• **Information systems** → *Decision support systems*;

KEYWORDS

disaster relief, location-centric processing, flood mapping, need matching

1 INTRODUCTION

Recent catastrophic events, e.g., resulting from floods and hurricanes, combined with increased urbanization and interdependent infrastructure reveal an increasing vulnerability to natural and human-made hazards. A resilience framework must accommodate both precise information for action and for various levels of analysis, posing substantial challenges [4]. These include characterizing hazards with an integrated ontology across voluminous, multi-modal data sources, facilitating rescue/response actions, and assessing the effects of pre- and post-disaster actions on individual needs.

The dire real-time need for integrated, coherent multi-modal data from different sources has motivated substantial research [8]. Social media is a significant source of such data. Twitter users, for example, post short texts as updates. These messages (a.k.a. tweets) reveal *Needs* and requires matching with possible available help. These messages can provide critical information to first responders and individual decision-makers to contribute effectively to relief efforts [18]. Tweets authored about crisis events may contain crucial but unstructured location-centric information (such as incident reports) even when explicit geo-tagging is absent. The ability to extract specific location information from this unstructured terminology and the other attached information is critical for timely assistance by the first responders.

Crowdsourced data (such as OpenStreetMap) that covers the geographical area of a disaster event can assist in the extraction of location content from tweets. To achieve this, we developed D-record: **D**isaster **re**sponse and **re**lief **coord**ination, a pipeline that utilizes multi-modal data (i.e., unstructured text, gazetteers, and satellite imagery). D-record allows first responders to visualize and match location-centric *Needs* and options for available help while taking into account transportation constraints and available options in the local area of a disaster.[1]

2 RELATED WORK

The following related work informs the D-record pipeline and clarifies its contributions.

Disaster-Centric Ontologies. Lightweight lexicons or vocabularies like the Management of Crisis Vocabulary (MOAC)[2], CrisisNLP[3], and CrisisLex [21] provide the different concepts for disaster management but lack the granularity required for describing response and miss crucial spatial and thematic details. Other efforts like UNOCHA's Humanitarian eXchange Language (HXL)[4] improve information sharing during disasters but lacks specific crisis-related concepts. The Social Media and Emergency Management (SMEM) ontology [13] combines emergency domain knowledge with social media information for Incident Classification. However, SMEM focuses on major coarse-grained categories of disasters such as meteorology or biology. As stated above, state-of-the-art techniques in event representation for disaster response is critically limited with respect to the usability and dissemination of recommended action and response (intentional intervention) required for relief

---

---

[1]The source code of the pipeline is available at https://github.com/shrutikar/d-record.
[2]http://observedchange.com/moac/ns/
[3]http://crisisnlp.qcri.org/
[4]http://hxlstandard.org/

efforts coordination. Given the above limitations, D-record uses a location-centric ontology to assist in matching relevant needs with available help to support a disaster response use-case as discussed in Section 3.

Need **Classification.** For extracting the type of *Needs* (e.g, food, shelter, or medical) from text data, state-of-the-art *Need* classification techniques such as [16] focus on the identification of information type. However, these techniques do not provide a comprehensive solution for matching classification output. [5] merely summarizes disaster information at a certain geographical level without providing any actionable decisions for response and relief. Instead, they provide summaries of topics to decision makers. Our previous classification technique [3] uses an ontology and location detection to enhance situational awareness through recognizing the interdependencies between disaster events, which helps in the early identification of *Needs*. Inspired by these techniques, D-record employs a coarser *Need* classifier to make its classifications compatible with the affordances of locations (e.g., a shelter can support a variety of *Needs* such as food and water whereas a hospital can provide medical and rescue support *Needs*). In other words, we resort to coarser needs (as opposed to fine-grained needs) for effective matching because available help centers satisfy multiple needs.

Matching. Our previous technique on Seekers and Suppliers matching [18] uses hand-crafted rules to match the *Need* requested with a possible provider, e.g., matching someone offering shelter with someone requesting shelter. Other approaches to *Need* matching such as [6] uses structured input from users and utilizes moderators in the loop to match those *Needs* with organizations presently available to help. The *Need*-capacity matching technique by [17] quantifies the degree of impact of a disaster and prioritizes the matching accordingly. Finally, [14] provides both online and offline solutions for requesting, providing and coordinating resources.

Limitations of the above work include: (a) the inadequate assumption that all routes are available for matching by assuming the full accessibility of roads during a disaster [19], and (b) the matching problem being solved for a different level of granularity than for the required individual level of *Needs*.

Flood Mapping. Some logistical matching methods assume the full accessibility of routes. Hu et al. [7], for example, provides an algorithm for optimal path selection in a logistics supply-chain management system. In contrast, [15] provides a portal for UN staff to upload spatial data including satellite images and crowdsourced data that can identify blocked routes. We use our customized state-of-the-art human-guided flood mapping technique [11] to prune the flooded routes to reflect the situation on the ground relevant for better-informed transportation.

Data Visualization. The majority of the visualization techniques in the literature place geo-tagged tweets on a map [12] or display users' structured input such as in [20]. However, these techniques fail for social media-based systems where geo-tagged tweets are relatively infrequent, and the information is buried inside unstructured texts. [10] showed the effectiveness of layered visualization systems, where each layer can contain a different source of information. In D-Record, each layer can include a different *Need* class or OpenStreetMap location features. To visualize the kinds of
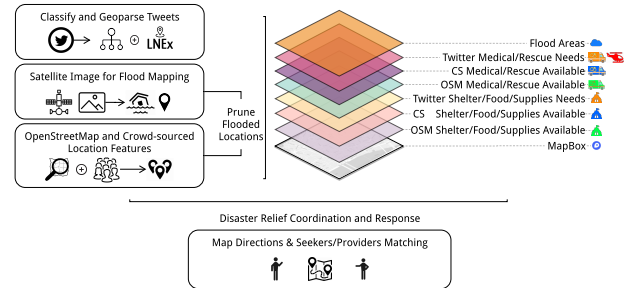

Figure 1: D-record map layers and pipeline functions

Needs effectively they are bound to their spatial footprints extracted using our tool LNEx [2].

3 METHOD

The following section describes the different data sources, the functionality of the pipeline, and the map layers we employ that support the visualization of information. Figure 1 contains a summary of map layers and their purpose.

Data Sources. D-Record utilizes three data sources: Twitter, OpenStreetMap, and Satellite Images. Twitter data consists of two groups. The first group contains a set of two targeted streams from two disaster events, namely, the Chennai Flood in 2015 and Houston Flood in 2016 [2]. The second group of crisis tweets is from CrisisNLP [9] and CrisisLexT26[16] datasets labeled for their information type (e.g., affected individuals, donations, and volunteering). The first group drives the pipeline. The second is used to train a binary classifier to recognize the expression of *Need* in tweets and expand the concepts in the underlying ontology (see Section 3). Crowd-sourced data from volunteers (contributed as Excel sheets) contain information about shelters and help requests/offers (e.g., Chennai Flood - List of Corporation and Relief Centres[5]).

The OpenStreetMap gazetteer provides location names and metadata (such as geo-coordinates) that produce the affordances of each location with respect to the needed help (e.g., a hospital can provide rescue and medical support) allowing for matching a help offer with a *Need* request from tweets. The Output of our flood mapping tool [11], which uses satellite images, determines the flooded geo-points and overlays this information on top of the map.

Location-Centric Ontological Modeling. Determining spatially and temporally specific instances for coordinating resources *Needs* benefit from a domain-specific location-centric event ontology for social media data. Event location is key to both aggregation and analysis of related event instances. Therefore, we created an event ontology that uses information from past disaster and risk reports[6] to enrich existing disaster event vocabularies.

The D-Record ontology[7] represents the temporal distinction between the response and relief phases of a disaster. The class "Concepts" includes two kinds of concepts: *Needs* and Availability. The class OSMFeatures covers the affordances of various location types (i.e., OSM map features[8]) in relation to the *Need* classes.

---

[5]https://rebrand.ly/reliefspreadsheet
[6]such as by https://www.acaps.org/
[7]The ontology can be found at https://github.com/shrutikar/d-record.
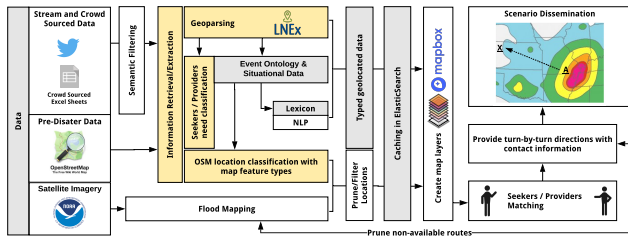[8]http://wiki.openstreetmap.org/wiki/Map_Features

Figure 2: D-record system architecture

Need **Class**	**CrisisNLP and CrisisLexT26 Classes**
Shelter/Food/ Supplies *Need*	donation_needs_or_offers_or_volunteering_services, displaced_people_and_evacuation
Medical/Rescue Help *Need*	missing_trapped_or_found_people, deaths_reports, injured_or_dead_people, affected_people

Table 1: Our mapping of CrisisNLP and CrisisLexT26 classes

The set of keywords for a "*Needs*" concept was expanded using topic modeling learned from the labeled data for the two classes: Shelter/Food/Supplies and Medical/Rescue Help. The relevance probability of each word to the Need class/topic functioned as a feature while training our *Need* classifier as shown in Section 3. Finally, the sub-classes of the *Needs* concept are mapped to the OSMFeatures sub-classes representing the options for available help at each location type (i.e., the affordances). For example, hospital maps to the Medical/Rescue Help class and pharmacy maps to the Shelter/Food/Supplies *Need* class.

Text Classification. To leverage the existing labeled data from CrisisNLP and CrisisLexT26, we combined the overlapping classes to form the two *Need* classes (see Table 1). We used tweets from these datasets to train an SVM-based classifier which allows D-record to categorize any given crisis related text into one of the two classes: Shelter/Food/Supplies *Need* or Medical/Rescue Help *Need*. The *Need* class Shelter/Food/Supplies includes donations, volunteer services, and *Needs* for food, water, shelter, or clothes. The *Need* class Medical/Rescue Help comprises affected people, death reports, injured or dead people, missing or trapped people, and other medical and rescue-related information. The classified streaming text appears on the map for matching.

The classifier had to capture tweet semantics. A simple Bag-of-Words model lacks context. Instead, feature engineering [22] vectorizes the text sentences to capture their semantics adequately. Before featurizing, the text was preprocessed by stemming, case folding and removing "noisy" lexical elements (such as URLs, non-ASCII characters, mentions, punctuations, dataset-specific stopwords, and hashtags). Finally, we designed an SVM classifier with lexicon-based features, TF-IDF vectors, and gensim's word2vec embeddings[9].

Using the relevance probabilities from topic modeling described earlier, we form lexicon-based two-tuple feature vectors (i.e., Shelter/Food/Supplies and Medical/Rescue Help). Each element in the vector represents the word frequency multiplied by its relevance score. This vector was concatenated with the other features from TF-IDF and gensim's word2vec embeddings, which captures the semantics of a word by looking at the context where it was mentioned. To address the class imbalance, SMOTE oversampled the minority class synthetically.

System Architecture. D-record uses three major forms of data (see Figure 2): Streaming and Crowdsourced, pre-disaster, and satellite imagery data. The knowledge extracted from text streams (filtered using hashtags of the disaster event) and crowdsourced excel sheets provides situational awareness and the various kinds of help available at each location. The pre-disaster information available from OpenStreetMap (sliced using a bounding box of the disaster event) represented the available help confirmed or pruned using satellite images. The output from our flood mapping method [11] helped prune out routes that were unavailable during the matching process for the location seeking help.

Our tool LNEx extracts and attaches each location to their lat-long information which locates tweet information on the map, following *Need* classification. Since OpenStreetMap provides the metadata for locations (including names, types, and geo-coordinates), the flood map information is used to prune out the locations that are flooded and not accessible for help. D-record caches all of this information in Elasticsearch. The *Need* information and location of a tweet indicate where the help is needed while OSM location features identify available help. We create the map layers on top of MapBox[10]. The system then matches a *Need* in the ontology with the locations that can fulfill that *Need*. D-record ultimately prunes all routes containing a flooded section according to the flood map.

4 RESULTS

SVM with the SMOTE (to partially overcome class imbalance) and the Gradient Boosting Algorithm performed the best and achieved a 0.8 F-Score in the *Need* classification task. We tested it using the leave-one-out technique (i.e., testing on one dataset and training using the rest). D-record extracted locations from tweets to support plotting these tweets on the map. The experiments used our Chennai and Houston datasets, containing 169,838 and 415,057 tweets, respectively. From these tweets, LNEx extracted 85,564 locations from the Chennai dataset and 241,684 locations from Houston dataset. As for the number of OpenStreetMap location features, 1,103 and 2,826 locations were retrieved for the affected areas of Chennai and Houston, respectively. Other crowdsource data received from the excel sheets were relatively few, around 41 locations.

5 DEMONSTRATION

Figure 3 shows D-record where users can choose the time range to filter the data using ⓐ, the dataset using ⓑ, and select the map layers using ⓒ. To match a *Need*, the user clicks on the orange icons as in ① which shows a textbox with the tweet text and the extracted location. Users can click on a "Match Need" button to get the closest and non-flooded nearby location which can provide help for the given *Need*, as in ②. Users obtain the matched location information by hovering the mouse over the green icons as in ③. The tool will provide the full information of the matched location and the contact number if available with the turn-by-turn directions obtained using MapBox directions API[11] as shown in ④.

---

[9]https://radimrehurek.com/gensim/

[10]https://www.mapbox.com/

[11]https://www.mapbox.com/help/define-directions-api/