Enabling Disaster Early Warning via a Configurable Data Collection Framework and Real-time Analytics

Young-Woo Kwon
Utah State University
Logan, UT, USA
young.kwon@usu.edu

Seungwon Yang
Louisiana State University
Baton Rouge, LA, USA
seungwonyang@lsu.edu

Haeyong Chung
University of Alabama in Huntsville
Huntsville, AL, USA
chungh@cs.uah.edu

ABSTRACT
The detection and prediction of natural catastrophes or man-made disasters before they occur has recently shone the light on several relatively new technologies. Due to the significant development of mobile hardware and software technologies, a smartphone has become an important device for detecting and warning about such disasters. Specifically, disaster-related data can be collected from diverse sources including smartphones' sensors and social networks, and then the collected data are further analyzed to detect disasters and alert people about them. These collective data enable a user to have access to a variety of essential information related to disaster events. Using the example of a communicable disease outbreak, such information helps to identify and detect the ground zero of a disaster, as well as make sense of the means of transmission, progress, and patterns of the disaster. In this paper, we discuss a novel approach for analyzing and interacting with collective sensor data in a visual, real-time, and scalable fashion, offering diverse perspectives and data management components.

ACM Classification Keywords
H.4.0. Information Systems Applications: General; I.2.4. Artificial Intelligence: Knowledge Representation Formalisms and Methods

Author Keywords
Crisis informatics, disaster, early warning, mobile devices, visual analytics, ontology

INTRODUCTION
Extreme environments are frequently characterized by devastation of and absence of infrastructure as well as hazardous conditions, as exemplified by a scene of a natural catastrophe (e.g., Tohoku earthquake and Hurricane Katrina) or a man-made disaster (e.g., the September 11 attacks). Although it is very difficult to predict such disasters before they occur, many technologies have recently received much attention for their potential to detect and warn about these disasters. By analyzing Twitter messages about disaster events, for example, it is possible to find warning signs of imminent earthquakes or typhoons [4]. Furthermore, due to the significant development of mobile hardware and software technologies, a smartphone has become an important tool for sharing information and warning about such disasters.

To detect a large-scale disaster event, much research has been conducted using diverse data sources in the area of crisis informatics. In particular, a massive amount of information including news articles, reports, videos and pictures generated online in a short period of time can be shared using social media such as Twitter [5]. The MyShake project [3] was initiated by UC Berkeley Seismological Lab and Deutsche Telekom Silicon Valley Innovation Center to detect an earthquake using smartphones. The MyShake project established a new type of a seismic network using smartphones to detect earthquakes. The CTRnet project [2] has been organized by Virginia Tech to collect disaster related resources as well as campus and other major shooting events. The overarching goal of the CTRnet Digital Library project was to archive online information including both Web articles and tweets, analyze and disseminate the information to various stakeholders (e.g., researchers, emergency management agents, general public).

In this study, we aim to design a system through which a number of catastrophes could be detected in time and possibly be avoided. To that end, we will explore how different disasters such as earthquakes, flooding, hurricanes, fires, disease epidemics (e.g., flu, cold, etc.) can be detected using the proposed early warning platform. Specifically, we will first create a new disaster ontology that illustrates how different disasters can be detected using diverse data sources including mobile sensors, social media, RSS feeds, etc. Then, an early warning framework will be dynamically configured and adapted to collect disaster-related data from diverse sources, thereby detecting different calamitous events without major modifications of deployed mobile software. In addition, our cloud-based crisis analytics engine will process the collected data to identify potential signs of disasters and send out warning messages. We also will present analysis results as multiple visualization views, through web-based interfaces, in order to facilitate rapid sense-making of the situation thus helping users of our system (e.g., first-responders and administrators making decisions).

HAI'15, Oct 26–Oct 1, 2015, Daegu, Republic of Korea.

DOI string from ACM form confirmation

In the balance of the paper, we show our initial ontology on different disaster types and their detection mechanisms, and then present the proposed approach and our vision for early detection of disasters using diverse data sources.

DISASTER EARLY DETECTION

In this section, we discuss how one can detect different disasters using different data sources. We first outline our approach and then present an initial disaster ontology as well as two major research thrusts to realize our vision of disaster early detection.

Approach Overview

We outline some possible solutions to the technical challenges that must be addressed in order to realize the vision of disaster early warning. The first step is to aggregate and archive crisis information generated from both mobile (e.g., sensors such as GPS, gyroscope, accelerometer, etc.) and online information (e.g., social media, news, RSS feeds, etc.) sources. The second step is to rate the reliability of information that is being aggregated in the first step. Based on the ratings of the sources, the first part—aggregation and archiving—may dynamically change allowing for a refined focus on sources of higher reliability ratings. The third step is to analyze aggregated content in real time. Text mining, natural language processing, or time-series analysis techniques are used for archived webpages and social media to identify the trending topics, location information, and named entities (e.g., organizations, human names). The trending topics found can be used as search keywords in the first step, allowing for a more focused archiving of information.

Figure 1 describes the approach overview. Since different types of disasters can be only detected using different data sources, the proposed disaster early warning framework will be dynamically adapted in accordance with a disaster detection configuration, which is generated at the cloud-based system based on our disaster ontology. For example, a system administrator determines a disaster type and area that he/she wants to watch. Then, the cloud-based system sends out a new configuration that includes which data should be collected from mobile devices. Finally, the framework automatically configure a mobile device's sensors and execution patterns using the received configuration to collect required disaster-related data.

In the following discussion, we present the technical challenges and our approach.

Reclassifying Disasters

Much research has been conducted in the area of disaster ontology, and many disaster ontologies have been proposed thus far. However, those ontologies were mainly focused on categorizing the disaster types lacking details of technologies for detecting and predicting such disasters. Thus, we have been incorporating such technologies in our ontology along with their corresponding disaster types. As an initial ontology, we classified various disasters into natural, human-instigated and compound categories described in Figure 2. The purpose of our ontology is to classify disasters in accordance with their detection technologies and relevant information sources such as mobile and online sources.

Collecting Quality and Reliable Disaster Information

In this section, we discuss how one can collect high quality and reliable disaster information from diverse sources to detect disasters.

Challenges

It is challenging to identify high-quality and reliable information and to sift out incorrect and misleading information considering that a great deal of information is generated very quickly for a large-scale disaster. This situation may become exacerbated if time-critical information has to be aggregated from multiple online sources of different types. For example, the latest news about a hurricane may be reported on websites of the local government, emergency management agencies (e.g., FEMA), and non-government organizations (e.g., Red Cross and Red Crescent). At the same time, people near the affected region may post messages and pictures on social media such as Twitter, Instagram, or Facebook.

Methodology

We will explore how diverse data can be collected from different data sources without redeployment of the mobile software. To detect different disasters, different data are required. For example, while both accelerometer and Tweeter data can be used to detect earthquakes [3, 4], social media data and local news data are required to detect hurricanes. To that end, we will develop a disaster early warning framework that can evolve to collect diverse information using different data collection mechanisms. First, a system administrator provides disaster information including a specific disaster that he/she want to detect, the region in which interested he/she is, etc. Then, in accordance with the newly created disaster ontology, our system determines required data, which will be collected by mobile devices and online sources. Using an agent-based configuration mechanism, the mobile device and crisis analytics engine automatically change their behavior to collect required data. Then, these data are continuously archived in real-time using tools such as web crawlers, plugins, and APIs provided by social media and mobile platforms.

However, considering that there exists no regulating body for monitoring the reliability of information on the Web, especially for the social media content, rating the information and sources can be crucial to have access to high-quality and reliable information. Each source and its content are rated based on our reliability criteria. A semi-automatic approach might be employed to enhance accuracy of the ratings.

Analyzing Collected Disaster Information

In this section, we discuss how one can analyze the collected disaster related information in real time and how to visualize the analyzed information.

Challenges

When a disaster occurs, key stakeholders (e.g., scientists, government officials, doctors, policy makers, etc.) need to explore and gain insight into a large amount of data in order to make appropriately informed decisions in a crisis situation.

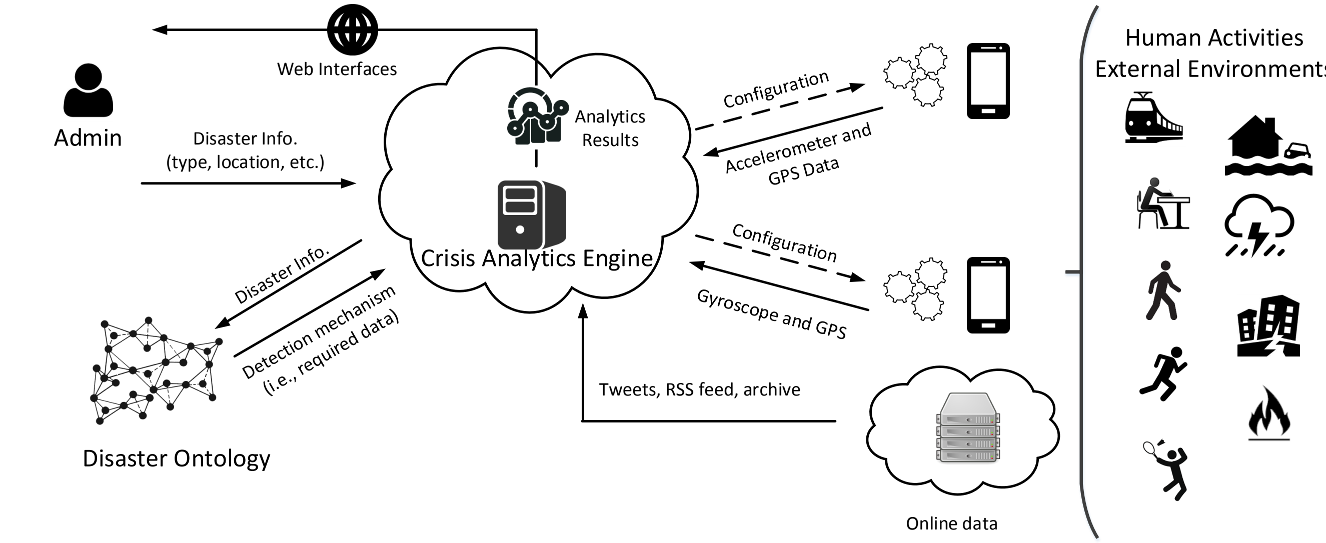

Figure 1. Approach overview.

However, analyzing and understanding collective sensor data features several inherent research challenges due to the characteristics of scale and uncertainty, as well as ubiquity of the data. To address such challenges, we present and discuss the three main challenges of identifying and understanding disasters and crisis information.

First, if a large number of personal devices can be used in a given disaster scenario, such crowd-sourced sensor data can increase information volume and area coverage in and around the outbreak location. Although these diverse data sources can be advantageous, the volume of sensor data associated with a given emergency situation may be very large, it could easily exceed conventional analysis and exploration methods since hundreds of thousands (or more) individuals could live in a disaster area or be affected by it.

Second, heterogeneous sensor data, such as data from accelerometers, magnetometers, gyroscopes, one or more GPS devices, etc., may be essential in analyzing a situation and making an informed decision. Thus, another challenge lies in gathering and understanding complex information derived from heterogeneous sources (sensors)—especially given the fact that different sensor data fed by a larger number of personal devices may also increase complexity and analytical uncertainty. Therefore, it is essential for analysts to see and compare information derived from heterogeneous types of data sources in order to solve analysis problems.

Finally, crisis and disaster events are likely to occur randomly, with little warning, and could be of indeterminate duration (e.g., the recent Ebola outbreak West Africa). The sensor data continuously obtained from thousands of smartphones could be viewed as a form of valuable streaming data. Thus, informed analysis will require the efficient handling of streaming data and visualization in real-time. Such a goal will inevitably entail investigating a new analysis infrastructure—not to mention practices for data aggregation, delivery, and storage of streaming sensor data from a large number of mobile devices.

Methodology

The above challenges are closely related to visual analytics problems, wherein analysts are "being asked to make decisions on ill-defined problems" [1]. As noted earlier, while aggregate sources of data represent potentially powerful sources of information, collective sensor data fed in by mobile devices may be uncertain or incomplete, and could very likely be extremely difficult to piece together. To address these challenges and problems, we will investigate how massive and heterogeneous sensor data can be employed to achieve deep and dynamic analysis of different disasters.

We will build a crisis analytics system that seamlessly integrates the strengths of computational methods with the visual analysis capabilities of the human analysts to aid in illuminating patterns, trends, clusters, outliers, etc. in collective sensor data. The proposed system will provide computational analytics methods for detecting, filtering, and extracting anomalies in sensor data, and will also afford corresponding visual representations and interaction facilities for analysts exploring and understanding such anomalies.

The proposed system enables crisis management personnel and general users to take advantage of the multiple perspectives offered by a large aggregation of sensor and GPS data. Furthermore, the system allows analysts to see, combine, and interpret diverse and often complex types of sensor data via either single or multiple views. For instance, relevant sensor data can be overlaid on a single large map with appropriate visual representations, tracking and showing the geographical distribution of seismic intensity, an overview of crisis status, and user behaviors concurrently in real time. Users could be represented by small shapes on a map that indicate different states and behaviors at any given time or place. Initial work for this approach has applied visual analytics to enable users

	Name	Signs	Detection mechanisms	Online sources
Natural Disasters	Earthquake	Land shake, sound	Accelerometer, load	NEIC, USGS
	Hurricane	Violent waves, rain	Humidity, temperature, barometer, speed	NHC
	Flood	Intensive rain, water level in the rivers, precipitation measure	Humidity, temperature	http://www.weather.gov/
	Snowstorm	Intensive snow, low temperature	Humidity, temperature	http://www.weather.gov/
	Mudflow	Heavy rain/snowmelt/ground water, land shake	Accelerometer, humidity	http://landslides.usgs.gov/
	Tornado	Cloud movement, rain, hail, thunder, pressure change	Humidity, temperature, barometer	http://www.spc.noaa.gov/ http://www.philanthropyjournal.org/
	Blizzard	Heavy snow, wind, temperature decrease	Barometer, temperature, humidity	http://www.weather.gov/ http://www.handsnet.org/
	Tsunami	Land shake, violent waves	Accelerometer	http://wcatwc.arh.noaa.gov/
	Land slide	Land shake, damp soil	Accelerometer	http://landslides.usgs.gov/ http://www.episcopalrelief.org
	Volcanic eruption	Heat, smoke, ash, land shake	Humidity, temperature, Accelerometer, barometer	https://volcanoes.usgs.gov/
	Cold wave	Rapid temperature decrease	Temperature	http://www.fema.gov/
	Heat wave	Rapid temperature increase	Temperature	http://www.fema.gov/
	Infectious		Temperature	

	Name	Signs	Detection mechanisms	Online sources
Man-made Disasters	Robbery	No movement, people lie on the ground/sit, people shake, stress, sweat	Accelerometer, force, gyroscope, heart rate	
	Shooting	No movement, people lie on the ground/sit sound, blood smell, gunfire smell, people shaking, stress, sweat	Accelerometer, force, gyroscope, heart rate	
	Civil unrest	Crowd, running, uniform movement of people	Accelerometer, force, gyroscope, heart rate	
	Car accident	Traffic jam, slow movement of cars, sudden braking, sudden stop of a car, unusual movement of a car	Collision detection, accelerometer, gyroscope, speed detector, force	NHTSA http://udottraffic.utah.gov/CLALertViewer.aspx?CLType=1
	Structural failure	Shaking of structure, sudden fall of car(S)/people/part(S) of structure	Accelerometer, gyroscope, force, speed	
Compound Disasters	Fire	Heat, smoke, sweat	Temperature, heart rate, barometer	http://www.srh.noaa.gov/ridge2/fire/ http://www.usfa.fema.gov/
	Avalanche	Strong wind, heavy snowfall	Force, barometer, speed	www.fsavalanche.org/
	Air pollution	Dust, harmful gases (carbon monoxide, nitrogen oxides, particulate matter, sulfur dioxide, lead and ozone)		http://www.epa.gov/oar/oaqps/ http://www.cleanairworld.org/ http://www.4cleanair.org
	Black out	Darkness, heat, cold	Light, temperature	http://energy.gov/

Figure 2. Initial disaster classification based on disaster detection mechanisms.

to control parameters of automatic natural language processing methods in order to visualize the computed results on a map [6].

In addition to the computational and visualization approaches, we will construct an interactive, easy-to-use interface for handling sensor data and empowering analytic results. This interface will facilitate an analysts ability to locate important piece of information within a large dataset using computational methods and visualization techniques.

CONCLUSION AND FUTURE WORK

In this concept paper, we presented a methodology by which one can detect a disaster by leveraging state-of-the-art computing technologies including mobile computing, social media, big data analytics, data visualization, etc. Furthermore, we discussed what research challenges must be addressed. In the near future, we plan to realize the proposed approach for the limited number of disasters (e.g., earthquake and disease) as a proof-of-concept system. Then, we will generalize our approach for other types of disasters.

REFERENCES

1. Kristin A Cook and James J Thomas. 2005. *Illuminating the path: The research and development agenda for visual analytics*. Technical Report. Pacific Northwest National Laboratory (PNNL), Richland, WA (US).
2. CTRnet 2015. The Crisis, Tragedy, and Recovery Network (CTRnet). (2015). Retrieved from http://www.eventsarchive.org/?q=node/22.
3. Qingkai Kong, Young-Woo Kwon, Louis Schreier, Steven Allen, Richard Allen, and Jennifer Strauss. 2015. Smartphone-based Networks for Earthquake Detection. In *15th International Conference on Innovations for Community Services (I4CS '15)*.
4. T. Sakaki, M. Okazaki, and Y. Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web (WWW '10)*.
5. K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*.
6. Seungwon Yang, Haeyong Chung, Xiao Lin, Sunshin Lee, Liangzhe Chen, Andrew Wood, Andrea L Kavanaugh, Steven D Sheetz, Donald J Shoemaker, and Edward A Fox. 2013. PhaseVis: What, When, Where, and Who in Visualizing the Four Phases of Emergency Management Through the Lens of Social Media. In *10th International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany*.