

Data-Driven Techniques in Disaster Information Management

TAO LI, School of Computing and Information Sciences, Florida International University
and School of Computer Science, Nanjing University of Posts and Telecommunications

NING XIE, CHUNQIU ZENG, WUBAI ZHOU, LI ZHENG, YEXI JIANG, YIMIN YANG,
HSIN-YU HA, and WEI XUE, School of Computing and Information Sciences,
Florida International University

YUE HUANG, School of Computer Science, Nanjing University of Posts and Telecommunications

SHU-CHING CHEN, JAINENDRA NAVLAKHA, and S. S. IYENGAR, School of Computing and
Information Sciences, Florida International University

Improving disaster management and recovery techniques is one of national priorities given the huge toll caused by man-made and nature calamities. Data-driven disaster management aims at applying advanced data collection and analysis technologies to achieve more effective and responsive disaster management, and has undergone considerable progress in the last decade. However, to the best of our knowledge, there is currently no work that both summarizes recent progress and suggests future directions for this emerging research area. To remedy this situation, we provide a systematic treatment of the recent developments in data-driven disaster management. Specifically, we first present a general overview of the requirements and system architectures of disaster management systems and then summarize state-of-the-art data-driven techniques that have been applied on improving situation awareness as well as in addressing users' information needs in disaster management. We also discuss and categorize general data-mining and machine-learning techniques in disaster management. Finally, we recommend several research directions for further investigations.

CCS Concepts: • **Information systems** → **Data mining**; *Data cleaning*; *Summarization*; Geographic information systems; Social networks; Personalization

Additional Key Words and Phrases: Disaster information management, data management, data mining, application

We would like to thank the following former members of KDRG (Knowledge Discovery Research Group) at Florida International University (FIU): Dr. Lei Li, Dr. Liang Tang, Dr. Chao Shen, and Dr. Jingxuan Li, for their early help on this article. The work was supported in part by the National Science Foundation under Grant Nos. HRD-0833093, IIS-1213026, CNS-1126619, and CNS-1461926; the U.S. Department of Homeland Security under grant Award Number 2010-ST-062000039; the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001; Chinese National Natural Science Foundation under grant 91646116; Nanjing University of Posts and Telecommunications under grant NY214135; Scientific and Technological Support Project (Society) of Jiangsu Province No. BE2016776, Ministry of Education/China Mobile joint research grant under Project No. 5-10, and an FIU Dissertation Fellowship.

Authors' addresses: T. Li (contact author), School of Computing and Information Sciences, Florida International University, 11200 SW 8th Street, ECS 354, Miami, FL, 33199 and School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, P.R. China, 210023; email: taoli@cs.fiu.edu; N. Xie, C. Zeng, W. Zhou, L. Zheng, Y. Jiang, Y. Yang, H.-Y. Ha, W. Xue, S.-C. Chen, J. Navlakha, and S. S. Iyengar, School of Computing and Information Sciences, Florida International University, 11200 SW 8th Street, ECS 354, Miami, FL, 33199; Y. Huang, School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, P.R. China, 210023.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2017 ACM 0360-0300/2017/03-ART1 \$15.00

DOI: <http://dx.doi.org/10.1145/3017678>

ACM Reference Format:

Tao Li, Ning Xie, Chunqiu Zeng, Wubai Zhou, Li Zheng, Yexi Jiang, Yimin Yang, Hsin-Yu Ha, Wei Xue, Yue Huang, Shu-Ching Chen, Jainendra Navlakha, and S. S. Iyengar. 2017. Data-driven techniques in disaster information management. *ACM Comput. Surv.* 50, 1, Article 1 (March 2017), 45 pages.
DOI: <http://dx.doi.org/10.1145/3017678>

1. INTRODUCTION**1.1. Background**

Disasters are severe, large-scale, non-routine events that disrupt the normal functioning of a community or a society, causing widespread and overwhelming losses and impacts. From ancient times, natural disasters (such as earthquakes, hurricanes, tsunamis, floods, tornadoes, and volcano eruptions) and man-made disasters (such as environmental disasters, terrorist attacks, and wars) have caused innumerable loss of lives, devastating destruction, and immeasurable economy losses. A disaster of moderate scale can easily result in a loss of hundreds of millions of dollars [Zheng et al. 2013]. Merely counting what happened in the last decade, disasters have caused more than \$800 billion in losses. Table I lists the disasters that have happened since 2005 that caused over \$10 billion of loss¹. If the losses of all the less severe disasters are also taken into account, the total amount would exceed 1 trillion dollars.

In fact, high cost is not the only devastating consequence of disasters: heavy casualties, mass destruction of infrastructure, and other aftermath of disasters all cause long-lasting adverse consequences in the affected regions. Take the *Tohoku earthquake*, for example. This 9.0 magnitude earthquake led to 15,889 deaths, 6,152 injuries, and 2,601 people missing. As for the destruction of infrastructure, it caused 127,290 building collapses, 272,788 buildings half collapsing, and another 747,989 buildings partially damaged [Tohoku 2011]. In the week right after the earthquake, the associated tsunami further triggered nuclear accidents that caused the evacuation of hundreds of thousands of residents who lived within 20km radius of the Fukushima Daiichi Nuclear Power Plant [CNN Wire Staff 2011; ABC News 2011]. As of this writing, the infrastructures of the inflicted areas have not fully recovered yet. The facilities – such as water, electricity, and gas supplies – were suspended for a long time in the affected areas. Compared with the rebuilding progress of infrastructure and facilities, the recovery from mental suffering for thousands of orphaned children will take much longer [Kyoto News 2012].

In addition to these direct losses, disasters are also fatal to the businesses in the affected areas. According to a study by *Contingency Planning and Management*, over 40% of the companies that were forced to shut down for at least three days due to a disaster would be unable to survive for more than 36 months. A key reason for businesses failing after a disaster is the lack of information on the availability of power and supplies as well as the lack of an effective and regularly tested disaster responseplan [Conference Board 2006]. The economic deterioration caused by business failures further impedes the recovery of the affected areas.

1.2. Disaster Management

Disaster management is the process of planning and taking actions to minimize the social and physical impact of disasters and reduce the community's vulnerability to the consequences of disasters. Given the devastating losses caused by disasters every year, effective disaster management has become a pressing issue for today's world, especially for disaster-prone countries such as China, Japan, and the United States [CERCS 2013].

Since disaster management is a multifaceted process, it is imperative to deploy proper management that optimizes planning and responses. Due to the limitation of resources

¹Source from Wikipedia http://en.wikipedia.org/wiki/List_of_disasters_by_cost.

Table I. Major Disasters in the Past Decade and Their Loss Estimates

Year	Name	Loss (In US Dollars)
2014	Iquique earthquake	\$30,000,000,000
2011	Tohoku earthquake and tsunami	Over \$300,000,000,000
2011	Thailand Floods	\$45,700,000,000
2011	Christchurch earthquake	\$40,000,000,000
2010	Deepwater Horizon oil spill	\$60,000,000,000 to \$100,000,000,000
2010	Haiti earthquake	\$14,000,000,000
2008	Sichuan earthquake	\$148,000,000,000
2008	Hurricane Ike	\$29,600,000,000
2005	Hurricane Katrina	\$108,000,000,000

during and after the disaster periods, collaborative efforts at the governmental, private, and community levels are indispensable. Collaboration at each level requires a prompt, coordinated, and organized effort to prepare for, respond to, mitigate against, and recover from emergencies and their consequences.

Disaster management is usually a cyclic process that consists of the following four main phases: (1) *preparation* – planing for various disasters that could strike within the area; (2) *response* – issuing warnings prior to a disaster and conducting initial actions during the disaster; (3) *recovery* – conducting actions to get the affected area and community back into normalcy; and (4) *mitigation* – taking measures that minimize the aftermath of a disaster and prevent the same damages from occurring again. Usually, the end of one phase is also the beginning of the next, but sometimes phases can overlap and several may take place simultaneously. Timely decision-making during each phase results in better preparedness, earlier warnings, and reduced vulnerability. The full disaster management cycle also includes the creation of effective plans and public policies for addressing the causes of disasters and/or mitigating their impacts on properties, people, and infrastructures.

1.3. Data-driven Disaster Management

Information exchange during and after the disaster periods can greatly reduce the losses caused by the disaster. This is because it allows people to make better use of the available resources, such as relief materials and medical supplies, and provides a channel through which reports on casualties and losses in each affected area can be delivered expeditiously. *Data-driven disaster management* refers to applying advanced data collection and analysis technologies to achieve more effective and responsive disaster management [Schmitt et al. 2007].

1.3.1. Data Sources for Data-Driven Disaster Management. Data useful for disaster management are highly diverse in nature, reflecting the diverse information and communication requirements at different times and locations with respect to disaster management. The data sources for data-driven disaster management can be classified roughly into two different types according to their temporal characteristics: *dynamic data sources* and *static data sources*. Common dynamic data sources include (a) periodically updated situation reports from the Office of Emergency Management (OEM) in each state and the participating companies that summarize the current status of threats, ongoing disaster relief operations, and the objectives of preparation and recovery efforts; (b) damage analysis reports and images from EOC and participating companies; (c) open/closed status of roadways/highways/bridges, schools, hospitals and businesses, as well as on/off status of other infrastructures (e.g., power, fuel, police stations, fire stations, and other emergency service); (d) news reports from various agencies; (e) mailing lists, direct emails, press releases, web pages, news feeds, and conference calls; (f) logs of 311 calls that report damages or status of various facilities;

and (g) data from social network sites like blogs, Facebook, and Twitter. Common static data sources include (a) historical data from the OEM at the federal and state levels; (b) incident action plans that illustrate the impending threat along with guidelines and plan of actions for preparing against the threat and recovery operations after the threat has elapsed; (c) properties and business locations; (d) public geospatial datasets; (e) public map services such as Google Maps and Microsoft Maps; and (f) road networks.

In general, there are four types of resource providers: local jurisdiction, media, non-governmental organizations (NGOs), and private sectors. Within each resource type, data are further divided into several levels of administrative division or different categories based on the type of business that generated the data.

1.3.2. The Scope of This Article. Data-driven disaster management is an emerging research area that has undergone considerable progress during the past decade. Its main advantage over traditional disaster management is that it interconnects different partners and entities in the system, allowing users to find valuable information that makes them aware of the current disaster situation and recovery status. Moreover, the community participants (the disaster management officials, utility agents, industry representatives, and so on) can collaborate to exchange critical information, evaluate the damage, and make practical recovery plans. Governments are adopting data-driven approaches to organize their disaster management assets and human infrastructures [FEMA 2009; Howitt et al. 2010; Saleem et al. 2008; Song et al. 2013]. For example, the *Division of Emergency Management at the State of Florida* has created a web-based platform to facilitate logistic and relief missions in the disaster-affected areas [Zheng et al. 2010]. Based on the user motion data of 1.6 million people during the *Great East Japan Earthquake* and the *Fukushima Daiichi nuclear accident*, a research team at the University of Tokyo has been collaborating with the government and industry partners [Song et al. 2013, 2014a]. They developed intelligent systems that automatically discover, analyze, and simulate population evacuation, and model human emergency mobility.

However, to the best of our knowledge, there is no work on data-driven disaster management that both summarizes recent progress and suggests future research directions. Our previous article [Hristidis et al. 2010], which summarizes existing research progress on data-driven disaster management from the perspective of data management, probably comes the closest to meeting these requirements. Unfortunately, some important aspects of disaster management research, such as *data processing* and *predictive analysis*, are left out in that work. To remedy this situation, the current article aims to provide a more comprehensive overview of the recent developments in data-driven disaster management and highlight the most important research challenges in the field. In particular, we will focus on the following key problems in designing a disaster management plan, which are crucial in achieving better information sharing and delivery:

Problem 1: *How to effectively capture the status information and improve situational awareness from diverse information sources?* During and after disasters, affected entities need to communicate/submit/share status reports using many sources. It is often desirable to deliver such status information as soon as it becomes available. In addition, some reports on a particular vicinity may be redundant. It is thus important to present a situation overview to users who desire concise and simple reports that are compiled from diverse information sources.

Problem 2: *How to effectively capture users' interests and deliver the relevant information to them?* Historical data and real-time data during disasters are collected from diverse sources, and they may belong to various categories. During disaster preparation and recovery stages, users usually cannot afford the time to go through the entire system to find needed information.

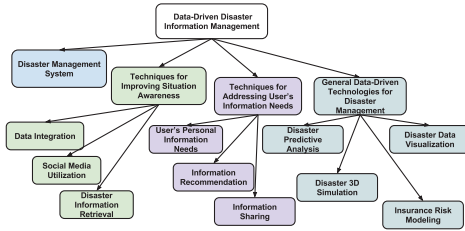


Fig. 1. Overview of this survey.

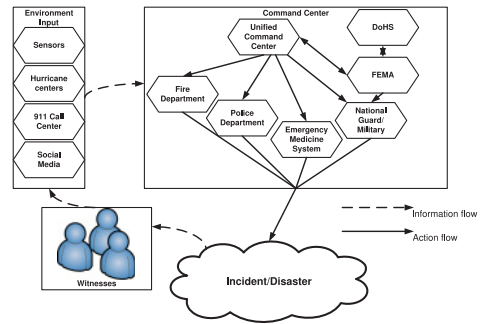


Fig. 2. Incident/disaster information and action flow.

Problem 3: *How to leverage the community information from disaster recovery?* Participating entities generally play different roles within various communities: they can locate in the same physical community or be members of the same retail chain. Identifying the interaction patterns among various communities during a disaster is quite important, since they often reveal information that is valuable in a recovery scenario.

In this article, we discuss several important aspects of the recent research on data-driven disaster management. We categorize and discuss preliminary studies that address these challenges. For *Problem 1*, we introduce the effective methods that help users understand the situation reports and improve situation awareness; for *Problem 2*, we present the intelligent information delivery techniques for effectively supporting users' information needs; and for *Problem 3*, we discuss how the techniques developed in data mining and machine learning are leveraged to improve the effectiveness of the disaster management systems.

1.4. Roadmap

The rest of this article is organized as follows (see also Figure 1). Section 2 contains an overview of existing research efforts on data-driven disaster management systems. In Section 3, we describe the techniques used to improve situation awareness. Section 4 introduces the techniques that address users' information needs in disaster management situations, and Section 5 provides a detailed discussion on the general data-mining and machine-learning techniques that have been applied in disaster management. Finally, we conclude with future research directions in Section 6.

2. DISASTER MANAGEMENT SYSTEMS

2.1. General Requirements

To best serve the purpose of saving lives and properties, a disaster management system must ensure timely interaction and coordination between disaster response mechanism and recovery efforts. Figure 2 illustrates a typical information and action flow triggered by a disaster. First, some witnesses immediately make 911 calls to inform the command centers of the emergency situation; some witnesses may post the emergency situation to social media, such as Twitter and Facebook. The emergency situation can also be detected automatically by on-site monitoring devices, such as fire alarm sensors. Then, as soon as the situation information is forwarded to the United Command Center (UCC), the rescue plan and decision are made immediately. Based on the severity of the incident and the resource capacity of each department – such as the police department, military, and fire department – the UCC gives administrative directions to

the corresponding departments, who finally proceed to perform their tasks to mitigate the situation.

The main functionality of a disaster management system is to manage and analyze disaster information in order to provide evidence for decision making by the UCC as well as to facilitate the coordination of the resource needed by all departments involved. In Kim et al. [2007], the following general requirements for an efficient and effective Critical Incident Management System are identified.

- Communication capacity.** Usually, a flood of phone calls are made to the first-responders and emergency agencies when an emergency occurs. At the same time, a surge of messages related to the emergency situation may be posted on many social media websites. These phone calls and messages are extremely important for mitigating the incident and rescuing the people in emergencies. Consequently, communication channels of a disaster system should provide sufficient load capacity to transmit information between the public and emergency response systems.
- Early detection and immediate response.** Immediate response to an emergency is critical in reducing casualty and property damage. Early detection is crucial in identification and characterization of the disaster, evaluation, and dissemination of possible damages caused by the coming incident. Quick response to the emergency situation reduces the cost of damages by providing instant evidence for the command center to make an accurate decision in a timely fashion.
- Effective information management.** As the main objective of disaster management is to form a solid basis for decision-making during emergency situations, it should competently collect, integrate, organize, analyze, and share information throughout all phases of a disaster.
- Resource allocation.** Under emergency situations, reasonable resource allocation helps to minimize the physical and social impacts caused by the incident given limited resources for rescue and mitigation.
- Security and privacy.** To carry out disaster prevention and recovery plan effectively, it is important to guarantee the security of disaster data, especially in a man-made disaster, such as a terrorist attack. In addition, the privacy issue has gained more attention recently. A disaster management system should install proper authorization access policies to keep private information from leaking to the public domain for misuse.

2.1.1. Information Processing Requirements. With respect to information processing, a disaster management system must deal with the following three challenging tasks: (1) how to efficiently collect and integrate disaster data from heterogeneous data sources, (2) how to search and identify disaster-related information to meet the users' needs, and (3) how to extract interesting patterns and trends from historical data. To a large extent, timely, accurate, and effective decision-making based on available disaster information depends on addressing these issues. Some well-studied information technologies, such as Information Extraction (IE), Information Retrieval (IR), Information Filtering (IF), and Data Mining (DM) have been applied in the disaster domain in Hristidis et al. [2010]. These information technologies form the essential building blocks in a disaster management system, as illustrated in Figure 3.

- Information Extraction.** The disaster information is extracted from heterogeneous sources such as social media and monitoring devices. Typically, the disaster information from different sources varies greatly in structure or format. To support further analysis and processing, a common format is specified for disaster information integration. Then, the integrated disaster information in this format is organized and stored for further processing.

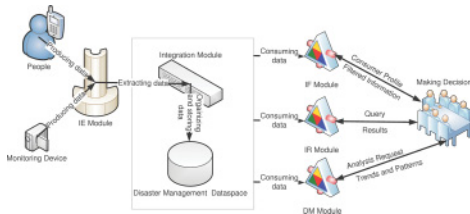


Fig. 3. Information technologies are applied in the disaster domain.

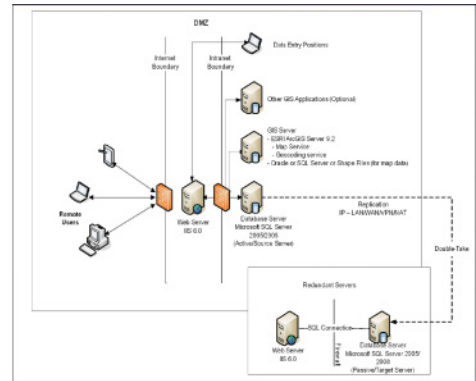


Fig. 4. The system architecture of WebEOC [Intermedix 2013].

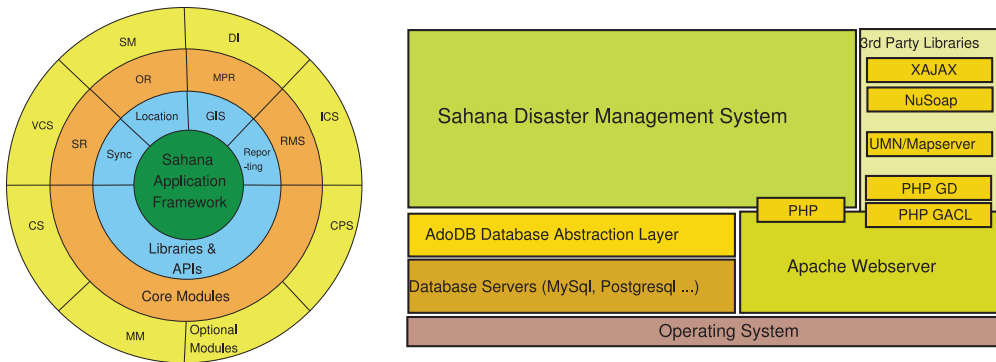
- Information Retrieval.** Disaster information relevant to the users' needs should be efficiently retrievable through queries to the disaster management system.
- Information Filtering.** If all the disaster information were presented to the users, it would cause an overwhelming workload. Instead, the disaster information should be filtered based on the specific purposes of the users.
- Data Mining.** It is very important to apply data-mining techniques to analyze the current and historical data in search of interesting patterns and trends, which form an indispensable basis for decision making.

2.1.2. System Evaluation. The criteria for evaluating a disaster information management system mainly concern the abilities of the system to (1) correctly differentiate the response from the agent-generated demands; (2) adequately carry out generic functions; (3) efficiently mobilize resources and personnel; (4) make an appropriate delegation of tasks and a division of labor; (5) process disaster-related information; (6) exercise decision making; and (7) provide instant reports for the news media [Quarantelli 1997]. Usually, a case study by users is used to measure these criteria. Different kinds of disasters and emergencies are simulated, and the user experiences with the disaster information system are recorded. The response from the users will be used to evaluate whether the system adequately satisfies these criteria.

2.2. General Overview on Disaster System Architecture

Large-scale disaster management involves a diversity of organizations and produces a great amount of heterogeneous data that are then consumed by these organizations. An effective Information and Communication Technology (ICT) solution plays a key role in effectively coordinating among these organizations, as it tracks the evolving situation and directs ongoing operations based on an overall awareness of developments. Next, we illustrate the architecture of these systems by taking a close look at several examples.

Sahana. Sahana [Careem et al. 2006], a free and open-source application, provides a comprehensive solution for information management in relief operation recovery and rehabilitation. A layered architecture of Sahana is depicted in Figure 5(a). The architecture of Sahana consists of four layers. The Sahana Application Framework lies at the core, surrounded by a set of libraries and Application Programming Interfaces (APIs) such as Location, GIS, and Reporting APIs. The core modules are then built on top of the Sahana Application Framework and libraries/APIs, with the optional



(a) The layered architecture of Sahana (b) The component-based architecture of a typical system based on Sahana

Fig. 5. The system overview of Sahana [Careem et al. 2006].

peripheral modules lying at the outermost layer, and are installed upon users' requests. In general, outer layer modules are able to use the functionality of the inner ones.

The following features are provided in the Sahana framework: a flexible and modular architecture that ensures that the systemwide tasks and events can be easily managed and synchronized; security at the modular and framework level, internationalization and localization of content; an easy way to set up and configure the Sahana and its database; support for automatic detection and dynamic plug-in of new modules due to its flexible architecture; capability to work with a common database to access its schema and disaster data; and administration functionality for other modules in the system.

The component-based architecture of a typical system built on the Sahana framework is illustrated in Figure 5(b). Such a system is composed of an Operating System, above which a Database and Web Server can be deployed to support the Sahana framework. Sahana can take advantage of other third-party libraries, such as XAJAX, NuSoap, and MapServer, to facilitate its operations and tackle more challenging tasks in disaster management.

WebEOC. A crisis-disaster management system called *WebEOC* was developed to manage large-scale events and disasters, support public safety information sharing, and provide real-time situational awareness. The system is maintained by the Division of Homeland Security and Emergency Management along with the State of New Hampshire and Department of Safety. One primary objective of *WebEOC* is to provide incident commanders, community leaders, and command-level personnel with one common operating picture of public safety operations, sensitive information, and infrastructure disruptions, upon which informed and effective decisions can be made on recovery and mitigation efforts. *WebEOC* is also used as a gateway to share information between the State Emergency Operations Centers (SEOC) and federal, state, and local public safety entities and critical infrastructure patterns. Additionally, *WebEOC* is one primary means of communications and incident management for the SEOC.

The system architecture of *WebEOC* is shown in Figure 4. The *WebEOC* system can be divided into three tiers by the Internet boundaries. The leftmost tier is the user layer, where users can access *WebEOC* by browser or service API. All business logic are implemented on the web server in the middle tier. In the rightmost tier, both the disaster information and GIS data are stored in the database. To avoid data loss due to disaster, both the web and database server are backed up with replication servers.

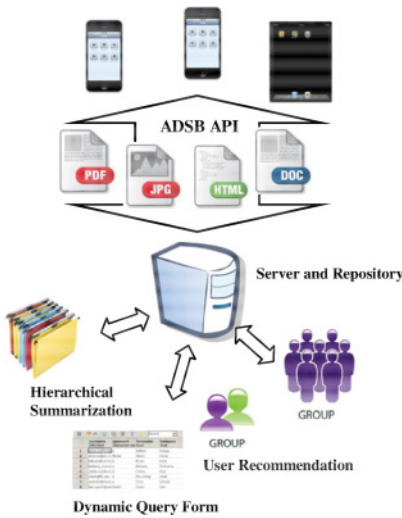


Fig. 6. The system Architecture of ADSB.



Fig. 7. Business Continuity Information Network (BCIN).

E-Team. E-Team, software offered by NC4 (National Center for Crisis and Continuity Coordination), focuses on the capability of situational awareness [NC4 2011]. To help emergency operation centers respond quickly and effectively to emergencies and incidents, E-Team – combined with existing software from the SAP – provides a powerful intuitive solution for web-based incident management and reporting. E-Team helps to leverage timely, trusted, high-quality data across the enterprise to enhance situational understanding and provides a common operating picture of the environment. It enables cross-jurisdictional data sharing and provides user information access to all operating units, regardless of the user's location. E-Team dashboards allow the user to tailor information displays to meet the specific user's role in disaster management.

ADSB. In the past decade, mobile computing has reached the commercial industry and mainstream consumers via smartphones and personal digital assistants (PDAs). Mobile devices are particularly useful in disaster situations, as they suffer almost no restrictions on time, location, or power supply. An All-Hazard Disaster Situation Browser (ADSB) system, running on mobile devices, is proposed in Zheng et al. [2011]. The system architecture of the ADSB is illustrated in Figure 6.

In order to share disaster information effectively, the ADSB supports the following functionalities: (1) provides users with a personalized list of reports that are the most important ones with respect to the users; (2) presents a comprehensive view of each individual report, in which all related information is displayed; (3) allows users to obtain summarized information by keyword searches; (4) supports the Dynamic Query Form, which allows a dynamic and heterogeneous query with desired attributes and their annotation in different reports; and (5) provides instant disaster information sharing through community management, which offers a channel for users to exchange disaster-related information and monitor the evolving situation of a specific event.

BCIN. Researchers at Florida International University have been working in collaboration with government and industry partners (such as the Miami-Dade and Palm Beach Offices of Emergency Management, Wal-Mart, Home Depot, Verizon Wireless, and many others) to develop a web-based service (i.e., a Business Continuity

Information Network (BCIN)) available year-round through which emergency management offices, local businesses, and organizations can share critical information and support efforts during all cycles of disaster management [Saleem et al. 2008; Hristidis et al. 2010; Zheng et al. 2012, 2013]. The BCIN is a business-to-business community network that allows members to track their key resources and locate required disaster recovery supplies, equipment, and services. In particular, the system facilitates professional organizations such as Chambers of Commerce to aid their members on disaster relief and expedites government agencies to assess damage in order to prioritize rescue efforts. The project website is at <http://www.bizrecovery.org>. Figure 7 is a snapshot of the current prototype, which displays information collected from various sources in an intuitive and user-friendly manner. The top panel is the situation dashboard, in which every row represents a jurisdiction and every column represents a facility. A colored symbol represents the status of a facility in a jurisdiction. For example, a green (resp., red) cycle indicates that the facility is open (resp., closed). The bottom-left panels provide users with richer views (such as pictures, images, videos) of the current threats and impacts, while the bottom-right panel shows a list of recent reports submitted by authoritative and trusted users. Multiple filtering options are provided that allow users to quickly extract valuable information from the system.

Other Systems. In addition to the disaster management systems outlined earlier, there are many other notable systems designed for different purposes and satisfying the needs of different type of users, for example, the RESCUE Disaster Portal [UCI 2015] and Puerto Rico Disaster Decision Support Tool [DDST 2015]. The National Emergency Management Network (NEMN) [Kapucu et al. 2010] allows local government to share disaster-related resources and information. These products are designed for county emergency management and participating agencies to provide effective reports sharing software systems. More recent systems include Ushahidi [2012] – which provides a platform for crowdsourcing news stories and crisis information to generate visualization and interactive maps, GeoVISTA [2010] – which monitors tweets to generate situation alerts using a map-based interface by making use of the geo-location associated with the tweets, and DI-DAP [Li et al. 2016] – which supports effective information delivery and analysis in disaster management.

Dealing with data unavailability issue. To handle data in the presence of unavailable sources and thus achieve high robustness and reliability, many techniques are adopted in disaster information management. The Distributed Data System [Özsu and Valduriez 2011] (DDS) is commonly adopted to prevent data loss and help data recovery. It allows portions of the database to be stored on multiple computers located in multiple physical locations within a network. Currently, with the rapid emergence of cloud storage, the data from the primary site is now backed up in the cloud and/or in multiple geographically separated data centers to improve fault tolerance and availability [Chidambaram et al. 2008]. Several cloud infrastructure and storage vendors, such as Amazon S3, Glacier, and Rackspace, provide storage for backup services. Data versioning arises as a another challenge for distributed collaboration systems or a system requiring frequent backup, such as a disaster management system [Clark et al. 2013]. Data versioning evolves from user-level source code/document control and enterprise backup tools to a single-assignment file system and storage system [Zhu 2003]. The Rapid Open Geospatial User-Driven Enterprise (ROGUE) Joint Capability Technology Demonstration (JCTD) [Clark et al. 2013] is committed to providing an innovative approach to geospatial data development, management, and sharing that leverages institutional interactions and enables rich collaboration, especially for disaster management institutions such as The Pacific Disaster Center and the

Department of State-Humanitarian Information Unit. They developed a versioning system, GeoGit, with the capability of allowing for distributed collaboration and versioning of geographic data and not relying on network connectivity to be fully functional.

Table II summarizes the disaster information management systems discussed; information on more such systems can be found in EMSystems [2005]. Our discussion so far indicates that disaster management systems have gone through several historical stages in the past decade. During the first stage, a system acts only as a traditional information management system that stores disaster information in a specific schema to support efficient query processing. In the second stage, a growing amount of disaster information from heterogeneous sources calls for more advanced techniques in information collection, management, analysis, and discovery. These advanced techniques often stem from state-of-the-art research results in areas such as information extraction, information filtering, information retrieval, data mining, and distributed computing platforms [Zeng et al. 2013]. In the third stage, with the surge of social media and the prevalence of mobile devices, people are able to share disaster information with little spatial or temporal constraint. As a result, disaster information management systems are demanded to provide a comprehensive solution that integrates advanced information techniques, social media, and mobile devices.

3. TECHNIQUES FOR IMPROVING SITUATION AWARENESS

Situation awareness (SA) in disaster management refers to conscious awareness of the dynamic environment with respect to time or space and understanding of the elements in the situation that are critical to disaster management decision-making. One major challenge for a disaster management system is to build a dynamic situation awareness system that can integrate all collected data, extract useful information from them, analyze and predict based on extracted information, and present results into the disaster coordinators' mental model of what is happening and what is likely to happen so that the latter can take effective and efficient actions. For instance, a vision of how, in the near future, the public will provide information about disasters through social media and how the information is processed, integrated, and utilized in emergency response is depicted in Palen et al. [2010]. In this scenario, all kinds of possible social media platforms receive crowdsourcing contents and disaster-related information is gathered and fed into repositories of different purposes via social media plug-ins or APIs.

In this section, we provide an overview of techniques that improve the situation awareness of a disaster management system. In Section 3.1, the challenge of integrating information from different data sources is addressed. In Section 3.2, the achievement of information extraction from unstructured textual data is delivered. Currently the most prevalent information-sharing platform, social media not only provide us the most diverse and substantial information but more important they provide the most up-to-date information, which can be life-saving in disaster management. In Section 3.3, we present the motivations for employing the state-of-the-art techniques that utilize social media in disaster management. Section 3.4 and Section 3.5 introduce information retrieval and advanced storyline generation techniques, respectively, both of which aim to help the user rapidly acknowledge core information from massive data.

3.1. Integrating Information from Different Data Sources

A disaster management system receives and broadcasts information many different ways (e.g., mailing lists, direct emails, press releases, web pages, and conference calls). It is desirable to obtain the status of situations as soon as the relevant reports are available. However, users may upload status information with documents of various formats (e.g., plain text, PDF, Microsoft Word, images, and videos). It is thus necessary to gather useful information from documents with different formats. In larger organizations,

Table II. Summary of Disaster Management Systems

System	Group	Description	Technology Summary	Cost	Intelligent Analysis: prost(+)/cons(-)	Phrases: support(+)/not support(-)
Sahana Careem et al. 2006	Sahana Community	The Sahana project aims to provide a set of modular, web-based disaster management applications.	(1) Location hierarchy selection and storage; (2) GIS mapping interface and spatial capability; (3) Automated and custom reporting capability; (4) Database synchronization between Sahana instances.	Open Source	(-) Document summarization; (-) Information recommendation	(-) Prevention, (+) Preparedness, (+) Response, (-) Recovery, (-) Mitigation
WebEOC [Intermedix Corp. 2013]	The State of New Hampshire, Department of Safety, Division of Homeland Security and Emergency Management	WebEOC is built to manage large-scale events and disasters and support public safety information sharing.	(1) Event reporting; (2) Situational awareness; (3) Resource management; (4) After action reports.	Commercial Solution	(-) Document summarization; (-) Information recommendation	(-) Prevention, (+) Preparedness, (+) Response, (-) Recovery, (+) Mitigation
E-Team [NC4 2011]	National Center for Crisis and Continuity Coordination	E-Team provides a complete set of web-based incident management tools that are easy to use and bring efficiency to managing an emergency.	(1) Incident and emergency event reporting; (2) Resources, critical assets and vendor management; (3) Critical infrastructure tracking, and so on.	Commercial Solution	(-) Document summarization; (-) Information recommendation	(+) Prevention, (+) Preparedness, (+) Response, (+) Recovery, (-) Mitigation
ADSB (All-Hazard Disaster Situation Browser system) [Zheng et al. 2011]	Disaster Information Management Group at Florida International University	ADSB is able to share the disaster information with others by mobile devices with the benefit of no restriction of time and location.	(1) Personalized reports; (2) Disaster information summarization; (3) Dynamic query form; (4) Instant disaster information sharing.	Free	(+) Document summarization; (+) Information recommendation	(-) Prevention, (+) Preparedness, (+) Response, (+) Recovery, (-) Mitigation
Business Continuity Information Network [BCIN 2006] (see Figure 7)	Disaster Information Management Group at Florida International University	(1) A Web-based service for information sharing and exchange to operate during and after a hurricane recovery period; (2) All-hazard situation browser on mobile devices.	(1) Information extraction from text data; (2) Document clustering and summarization for building a summarized view; (3) Spatial clustering for dynamic geo spatial community generation; (4) Information and user recommendation for community-based recovery.	Free	(+) Document summarization; (+) Information recommendation	(-) Prevention, (+) Preparedness, (+) Response, (+) Recovery, (+) Mitigation

or in the case that there is a host of members in a neighborhood (e.g., a corporate park), multiple reports about a particular area might be redundant, which calls for a concise, accurate, and consistent summarized view derived from diverse information sources. Data integration and data fusion have been extensively studied in the database community [Haas 2006], and have been applied to analyze multimedia data to further enhance the performance of multimedia content-based retrieval [Bagheri et al. 2013], event detection [Corey 2012], event summarization [Jiang et al. 2011, 2014; Chakrabarti and Punera 2011], and semantic concept detection [Atrey et al. 2010].

To improve the situation awareness of disaster management systems, much of the research effort in data integration is devoted to the following tasks: (1) to convert contents of different formats into a standard format; (2) to verify the credibility of various crowdsourcing data sources and attempt to leverage it to produce useful information for disaster decision-making; (3) to map images or texts with their corresponding geo-locations to better capture the current situation; and (4) to process and analyze the data from different sources.

3.1.1. Standardization. One of the major challenges of integrating data in different formats is that there is no standard way to represent all kinds of data. Peng et al. [2011] propose an incident information management system with three major components: data integration, data mining, and multi-criteria decision-making. The first component (i.e., data integration) is composed of three layers: first, the Distributed Heterogeneous Data Interface receives and converts heterogeneous data via wired or wireless channels; second, the XML-based Integrated Processor, by applying data fusion algorithms, processes various types of data sources by modeling, integration, and fusion; and last, the Unified Data Interface (UDI) provides an easy-to-use and standardized programming interface to the data-mining and decision-making components. Following a similar principle, Hu et al. [2011] propose storing heterogeneous data in the XML format in local files in order to speed up data transition. To visualize the disaster report in real time via corresponding geo-location and to loosen up the process of integrating data so that more observed data from volunteers can be contributed to existing disaster systems, Ortmann et al. [2011] suggest “outsourcing” a substantial part of the data integration to the crowd by asking people to process unstructured observations into the form of (subject, predicate, object) RDF-triples (Resource Description Framework), following the Linked Open Data principles [Bizer et al. 2009].

3.1.2. Verification. The challenge of verifying crowdsourcing data, which is obtained via either social media (e.g., Facebook, Twitter, and blogs) or short message service (SMS) sent through mobile phones, has been highlighted in work such as Heinzelman and Waters [2010], Gao et al. [2011], and Roche et al. [2013]. Crowdsourcing is an efficient method for quickly generating a large amount of near real-time data for a given event. This is particularly valuable for disaster management as the data is mostly from people who are actually at the disaster scene or nearby, while other channels of obtaining data are usually at least partially blocked or compromised. As a result, the accuracy and validity of crowdsourced information is crucial for further identifying actionable information and making timely decisions. Both Gao et al. [2011] and Zook et al. [2010] suggest using collective feedback to “self-adjust” errors in crowdsourcing data. However, this solution in general does not resolve the tension between experts and amateurs when data quality is of high priority. Moreover, it also fails in situations in which there are not enough people to serve as “verifiers.” Further investigations are needed to address these issues.

3.1.3. Mapping and Linking. To utilize pinging outsourcing information (e.g., text description, videos, images, links to useful websites) in disaster management, Kumar

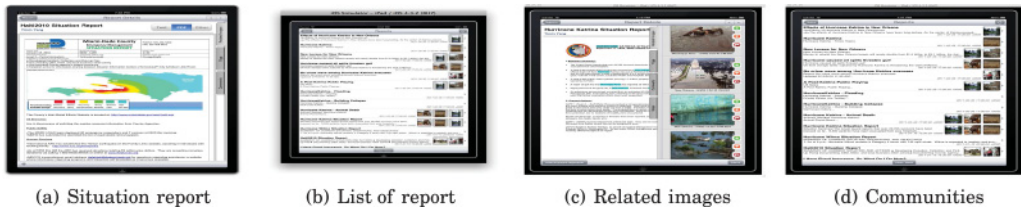


Fig. 8. MADIS major components.

et al. [2011] develop an application called “TweetTracker” to demonstrate how instantly pinging the content of tweets with disaster hashtags can be useful in visualizing situational awareness and monitoring tweet streams from both the spatial and temporal perspective. White [2010] also proposes to collaborate with social media and link all real-time multimedia information to the crisis map as a feasible solution to improve SA in disaster management. Sakuraba et al. [2013] not only recognize the need to collect disaster information in a GIS-based manner but also go one step further to display the information on a tiled display platform; thus, more disaster-related information can be analyzed on an ultra-high definition display environment.

In addition to integrating information by exploiting links among related data of a single type (e.g., microblog texts), there are also works focusing on analyzing the multimedia contents of different types (e.g., text report, images, and videos), and aiming to discover the temporal, geospatial, or topical correlations. Yang et al. [2012] and Zheng et al. [2011] introduce the Multimedia Aided Disaster information Integration System (MADIS) and the ADSB system, respectively. Both systems are designed and developed to first process each single data type independently and then associate them based on the analysis results. Specifically, MADIS fuses image and text information so that images can be accurately classified to the target disaster subject and associated with the related disaster report. Figure 8 illustrates the interface with which users can browse over the report list and all the related images.

3.1.4. Linked Open Data. The idea behind Linked Open Data (LOD) is to publish machine-readable data in the RDF format on the Web [Yu 2011]. Acted as the most exhaustive semantic knowledge base, it allows broad applications on three categories: Linked Data browsers [Becker and Bizer 2008; Hastrup et al. 2008], Linked Data search engines [Cheng and Qu 2009; Hogan et al. 2007], and domain-specific Linked Data applications [Heath and Motta 2008; Becker and Bizer 2008]. Though LOD allows public access to information originating from numerous information providers, the quality of provided information varies due to different levels of knowledge and different intentions of information providers. In Bizer and Cyganiak [2009] and Heath [2008], an overview of different content-, context-, and rating-based techniques that can be used to heuristically assess the relevance, quality, and trustworthiness of data is given. Ortmann et al. [2011] suggested crowdsourcing LOD as the next step toward a full exploitation of crowdsourced information in disaster management. Silva et al. [2013] proposed an effective way, by means of the LOD technology, to bring a large amount of data from different diversified data sources into a standardized and exchangeable common data format, enabling their interoperability and integration in disaster mitigation and preparation phases of disaster management.

3.2. Textual Information Extraction

IE studies the problem of how to automatically identify instances of facts – including entities, relations, and events – from unstructured or semistructured texts, and convert them into structured representations. IE can be used to build a database with

information on a given relation or event from text data (such as reports, news, web pages, and social posts) and is widely used as an important component in other intelligent information systems, such as question answering, summarization, machine translation, and search engines. Research on IE dates back to 1979 [DeJong 1982] and many IE systems have been built during the past two decades [Paik et al. 2000; Xu et al. 2010a; Jayram et al. 2006]; see the survey articles Cowie and Lehnert [1996], Chang et al. [2006], and Sarawagi [2008]. Early IE methods and systems are mostly knowledge-based or rule-based (MUC-I – MUC-7). With the development of machine-learning techniques, learning-based methods dominate IE over the last decade (ACE, CoNLL). Recent IE research focuses on how to extract information from larger-scaled and more complex texts, such as multilingual documents [Peters et al. 2012] and on how to make use of knowledge base (KBP) [Hoffmann et al. 2011].

3.2.1. Information Extraction Core Modules. Currently, most IE systems consist of the following three core modules:

Name Tagging. This module recognizes and classifies phrases from text as named entities. The most widely used categories of named entities are Person (PER) – named person or family; Organization (ORG) – named organizational entity (corporate, governmental, or other), and Geopolitical entity (GPE) – name of politically or geographically defined location (such as cities, states, countries, continents, international regions, bodies of water, and mountains). Other categories include location, artifact, facility, vehicle, weapon, product, time, and numerical quantities. Extended name hierarchy is introduced, which includes over 150 domain-dependent types [Sekine and Nobata 2004].

Relation Extraction. This module decides the relationship between two given named entities or finds all pairs of named entities of a given relationship.

Event Extraction. This module recognizes in texts both triggers and their arguments for predefined categories of events. An *event* is a specific occurrence that implies a change of states. Description of an event includes an *event trigger* and several *event arguments*. For example, in the text “the attack on Gaza killed 13,” “killed” is a trigger of the event and both “Gaza” and “13” are arguments of the event.

In addition to these core modules, an IE system often contains a pipeline of fundamental natural language processing (NLP) tools, such as tokenization, part-of-speech tagging, dependency parsing, and consistent parsing.

3.2.2. Challenges and Solutions for Information Extraction in Disaster Management. In the disaster management domain, unstructured text (e.g., Adobe PDF and Microsoft Word) is still one of the most widely used forms of information for storage and exchange. For example, organizations (such as the Miami-Dade Emergency Office, the Fire Department, and the Coast Guard) publish status reports frequently in unstructured text following disaster events. Note that the same organization often follows a similar format for all its reports. This fact makes it possible to extract information using a fixed set of rules if the sources (e.g., organizations) are known in advance. However, it is much more difficult to incorporate information from new organizations whose report formats are unknown, from sites whose news stories are in ad-hoc formats, or from individuals on social media. New techniques are required to achieve high-quality IE for such situations.

Adapting IE systems to new scenarios and tasks is quite difficult since most of the current technologies often require interventions from IE experts. The adaptation involves the following main tasks:

- Adapting to different domain-specific features.* Reimplement and rebuild system resources, such as lexica and knowledge bases;
- Adapting to different sublanguages features.* Modify grammar and lexicon to deal with specific linguistic constructions that are typical for the application;
- Adapting to different text genres.* Specific text genres (e.g., medical abstracts, scientific papers, police reports) may have their own lexicons, grammar, and discourse structure; and
- Adapting to different document types.* For example, web-based documents can radically differ from newspaper-like texts.

As an important scenario example for open-domain IE, disaster-related benchmark corpus data has been built. For instance, terrorist attacks are investigated under MUC-3 and MUC-4 [MUC DATA SET 2001], and a corpus data of disease outbreak is built and studied in Grishman et al. [2002]. Although not built to target the disaster domain, some systems are evaluated on disaster-related data or based on user studies of disaster management [White et al. 2001].

For a domain-specific IE system, it is vital to incorporate domain knowledge. For standard IE modules and methods, domain knowledge is coded in categories of named entities, relationships, event definitions, and labeled data. In Döhling and Leser [2011], a 5-ary relationship is defined for information on earthquake damages. A dictionary+rule-based method is used for named entity recognition and classification; a two-step approach, which consists of discovering entity pairs and identifying maximal cliques in entity graphs, is then used for relationship extraction. Instead of labeling relationship data, ontology – a widely used form of representing domain knowledge – is used in Samiha et al. [2011] for IE in disaster management. In their work, entities are first extracted from text using WordNet (a lexical database for English) and GATE (an open-source text-processing software) and then mapped onto an existing disaster ontology. To decide whether two entities have the relationship of their corresponding concepts in the ontology, the lexical chain between these two entities and ontology-defined relationship are compared.

Social media provide disaster management with more real-time data together with new challenges. To cope with the problem that data from social media is large and noisy, Imran et al. [2013] propose practical methods to extract information from social media during disasters. They first classify each tweet into one of the following categories: “Caution and Advice,” “Information Sources,” “Casualties and Damage,” “Donations,” “People,” and “Other.” Then for each tweet, using a sequential labeling algorithm, class-specific information is extracted for further analysis. For example, “Who is missing” or “Who has been found” will be identified for tweets in the “People” category.

3.3. Utilizing Social Media

Social media provide services of online interactions among people who create, share, and exchange information and ideas in real or virtual social networks. Typical social media includes blogs and microblogs (e.g., Twitter), content-sharing communities (e.g., Flickr, YouTube), social networks (e.g., Facebook) and collaborative projects (e.g., Wikipedia). During the last decade, social media sites have become increasingly popular information distribution tools that allow users to share and exchange their status, pictures and videos, ideas, and knowledge. Compared with traditional government situation reports or news reports for disaster management, social media provide faster access to a wider range of disaster information due to their real-time nature and crowd intelligence.

3.3.1. Overview. The main advantages of crowdsourcing via social media in disaster relief are the following [Gao et al. 2011; Yin et al. 2012]: (1) using social media status

reports and user requests are collected almost immediately after a disaster; (2) data collected from social media can be analyzed and categorized by simple tools to identify, for example, most urgent requests; and (3) geotag information associated with social media messages can help relief organizations locate requests for help. On the other hand, the authors also point out several limitations and shortcomings of social media, including: (1) no common mechanism is provided for collaboration and coordination between disparate relief organizations; (2) data from social media do not always provide the right and accurate information; and (3) there are not enough security features for relief organizations and operations.

In the past few years, many researchers conducted case studies of particular disaster events to showcase how data gathered from social media benefit disaster SA.

- By studying a dataset of 50 million tweets from 218,860 frequent users from March 7 to March 24, 2011 during and after the Great East Japan Earthquake in 2011, Sakaki et al. [2011] report that (1) the frequency of tweets was affected by planned blackouts, which were performed several days after the earthquake; (2) the frequency of tweets in heavily damaged areas decreased during the first several days that followed the earthquake; and (3) shortly after the earthquake, tweet frequencies of smartphones and feature phones were dominant in heavily damaged areas while the frequency of PCs was dominant in less-damaged areas.
- In a similar study, Miyabe et al. [2012] find, from tweets after the Great East Japan Earthquake in 2011, that people in the disaster area tended to directly communicate with each other (reply-based tweet). In contrast, people in other areas preferred spreading the information from the disaster area by retweets.
- Twitter data between February 27, 2010 and March 2, 2010 on the 2010 Chile earthquake is collected in Mendoza et al. [2010]. A study on one of its subsets indicates that “false rumors tend to be questioned much more than confirmed truths.” This study suggests the possibility of identifying rumor tweets by people who tweet when there are enough similar tweets in the same category.
- Two Twitter datasets collected using keywords for the Red River Flood and Oklahoma wild fires, both of which happened in Spring 2009, are analyzed in Vieweg et al. [2010] in order to identify features of information generated during the disasters. On-topic tweets about the disaster events are manually labeled and can be categorized with situation categories, including “Warning,” “Preparatory Activity,” “Fire Line/Hazard Location,” “Flood Level,” “Weather,” “Wind,” “Visibility,” “Road Conditions,” “Advice,” “Evacuation Information,” “Volunteer Information,” “Animal Management,” and “Damage/Injury.”
- The messages posted at an online forum, Tianya, during the first week after the 2008 Sichuan earthquake in China are studied in Qu et al. [2009]. Related threads are manually selected and classified into “Information,” “Opinion,” “Action,” “Emotion,” and “Other” categories. The authors identify the corresponding major roles that Tianya played in the disaster relief efforts.

Palen and other researchers [Palen et al. 2007a; Palen and Liu 2007; Palen et al. 2007b; Shklovski et al. 2008; Vieweg et al. 2008; Palen et al. 2009; Shklovski et al. 2010] have extensively studied how ICT, most notably social media, help people during disasters to form new communities with spatially dispersed participants, to gather and distribute accurate information, and to work together and solve problems in a distributed fashion.

3.3.2. How to Collect Related Content from Social Media. To make use of social media in disaster relief and management, the first step is to collect related content from the universal space of social media. Twitter provides streaming API, which collects tweets

with a specified set of keywords. However, due to the ambiguity of a word and the diversity of context in which it may appear, tweets containing disaster keywords may not be talking about the disaster situation or may not even be disaster related. Typically, a two-step strategy is used to tackle this: tweets are first collected using a set of keywords and then a classifier is applied to filtered tweets.

Tweets on earthquake events are collected in Sakaki et al. [2011] first by a keyword set $Q = \{\text{“earthquake” and “shakes”}\}$, then followed by a classifier using a support vector machine (SVM). The classifier uses the following group of features:

Simple statistical features. The number of words in a tweet as well as the position of the queried word in a tweet;

Keyword features. The words in a tweets; and

Context word features. The words adjacent to the queried word.

The work of Aramaki et al. [2011] focuses on extracting tweets that indicate an influenza patient to detect influenza epidemics using Twitter. Tweets containing “influenza” or “flu” are first collected. However, these tweets may include tweets expressing suspicions or questions or just talking about flu news. To filter out such tweets, several classifiers are compared using the context word feature within a context window; SVM with polynomial kernel is found to achieve the best performance.

These results can collect on-topic tweets from the entire tweet stream about a particular type of disaster. However, some on-topic tweets – although mentioning the disaster events – do not contain information useful for a timely and informed decision. It is proposed in Verma et al. [2011] to sift tweets that can contribute to SA during disasters and apply NLP techniques to identify such tweets. Specifically, a classifier is built for on-topic tweet datasets to filter tweets, with single features such as uni-grams, bi-grams and part-of-speech tags, and linguistic features such as (1) whether the tweet is objective or subjective; (2) whether the tweet is formal or informal; and (3) whether the tweet is personal or impersonal. Experiments show that max-entropy classifiers outperform native Bayesian classifiers, and good performance in different disaster types can be achieved if classifiers are trained separately using all these single features and linguistic features.

3.3.3. Statistical Analysis. Once tweets on disaster events are collected, as the volume of such tweets can be very large, it still might be difficult to extract useful information to help disaster-related decision-making. One solution is visualization. A typical central component to visualize tweets is a map-based filter [Kumar et al. 2011; MacEachren et al. 2011], which displays tweets on a map according to their geotags or the values of some other attributes. To visualize topic change over time, TwitterTracker [Kumar et al. 2011] includes a date-based tweet filter that supports the analysis of older tweets, together with a play mode that helps visualize changes in tweets through hashtags and entities. TwitterTracker also includes an entity and keyword analyzer that ranks entities contained in tweets and displays frequent keywords as tag clouds as well as a keyword trending engine that generates trends of keywords specified by the user.

SensePlace2 [MacEachren et al. 2011] provides a compact representation of the frequency distribution of tweets and a control that filters tweets by time windows. SensePlace2 supports a task list view that allows users to label and store query results. These stored queries also record user-set parameters (e.g., place and time filtering and decisions) to promote highly relevant tweets or hide irrelevant ones.

3.3.4. Prediction. To make the best use of this large number of tweets, in addition to the analysis and visualization tools discussed earlier, prediction models provide an alternative solution.

Prediction of Influenza. In the United States, a component of the influenza surveillance program tracks reported influenza-like illness (ILI) during influenza seasons. Similar organizations exist in Europe (European Influenza Surveillance (EISS)) and Japan (Infection Disease Surveillance Center (IDSC)). These systems rely on both virology and clinical data and typically have a 1- to 2-week reporting lag.

To eliminate the lag, tweets about influenza are used to improve the prediction [Signorini et al. 2011; Chen et al. 2010; Achrekar et al. 2011]. Signorini et al. [2011] use SVM to build a regression model to predict ILI score using weekly term-usage statistics on tweets as features. In Chen et al. [2010] and Achrekar et al. [2011], ILI is predicted using past ILI data together with current and past social media data as a time series prediction problem, on which the Auto-Regressive Moving Average (ARMA) model is applied.

Early Detection of Disaster Events. A disaster event detection system using Twitter data is described in Sakaki et al. [2010], which managed to send out earthquake warnings in Japan earlier than the announcement broadcast by the Japan Meteorological Agency (JMA).

3.3.5. Disaster Recovery. Data on social media not only help agencies with SA but also provide specific information queries regarding the disaster.

A question-answering system in the disaster domain is introduced in Mizuno and Inui [2013], which retrieves answers from tweet sets for disaster-related questions such as “What is in short supply in Tokyo?” The core module of the question-answering system extends a pattern-based relation extraction method, which extracts “X causes Y” patterns from the data. It displays answers in two modes: semantic mode, in which the answers are clustered by their semantic meanings with different colors, and Google Maps mode, in which answers are located on Google Maps.

One difficulty of using social media in disaster relief is that aid request messages and their corresponding aid provider messages are often not matched. This is because it is hard for victims and the matching rescuers to exchange messages due to the overwhelmingly vast amount of information flow during a disaster. A new method to match such request–provider pairs is developed in Varga et al. [2013]. Their method is similar to question–answer pair retrieval in question-answering systems, as aid requests and aid providers are analogous to questions and answers, respectively. The system consists of two SVM recognizers that identify whether a tweet is a request report or an aid provider message, together with a request–provider tweet matching recognizer, which exploits the dependency relations between tweet pairs extracted from linguistic features.

A summary of utilizing social media for disaster management is provided in Table III.

3.4. Information Retrieval for Emergency Disaster Data Management

IR is the technique of searching for textual entities (e.g., reports) relevant to a user query. Unlike IE, which concerns how to produce structured data ready for postprocessing from unstructured texts, IR identifies relevant documents from a collection of documents. Usually, a user query can be expressed as a collection of keywords, and the retrieved results are returned in the form of a ranking list, ordered by the predefined relevance scores.

Traditionally, the returned textual entities are ranked according to *term frequency* and *inverse document frequency* of the query terms, where *term frequency* (tf) denotes the number of occurrences of a term in an entity and *inverse document frequency* (idf) denotes the uniqueness of the query term.

The disaster scenario poses additional challenges for IR as the data under consideration is usually a combination of both static data and streaming data, the query

Table III. The Summary of Utilizing Social Media for Disaster Management

Usage Area	Methods	Description: pros(+)/cons(-)
Relevant Content Collection	Keyword-based filtering	(+) Straightforward; (-) Not easy to choose proper keywords; (-) False-positive information
	Classification: SVM, Entropy maximization, Bayesian method	(+) Reduce false-positive information; (-) Time-consuming feature engineering; (+) More accurate
Statistical Analysis	Visualization: Map-based filter, TwitterTracker	(+) Geo-information display; (+) Dynamically track changes over time; (+) Generate trends of keyword
	SensePlace2	(+) User input allowed; (+) Compact representation; (-) Less user-friendly
Prediction	Influenza surveillance based on virology and clinical data	(-) 1-2 weeks reporting lag; (+) Accuracy
	SVM and ARMA based on past media data	(+) Reduce time lag; (-) Feature construction;
Disaster discovery	Question-answer using pattern-based relation extraction	(-) Template construction; (+) Straightforward
	Question-answer using two SVM recognizers	(+) Better match between question and answer; (-) Feature construction;

delay should be as short as possible, and the retrieved results should be personalized depending on the users – all due to the unique characteristic of the emergency disaster management situation.

Some existing work on personalized web search [Sontag et al. 2012; Wang et al. 2013; Fathy et al. 2014] can serve as a basis for meeting such challenges. In the disaster management domain, a key ingredient for personalized IR is the domain knowledge, such as rules and ontologies of disasters. For example, a disaster management ontology may connect the terms “earthquake” and “tsunami” and make them logically correlated. Such a correlation can be used to expand a user query. Recent research effort has been devoted on domain-specific ontology query expansion [Chauhan et al. 2013, 2012; Hahm et al. 2014]. If the domain ontologies are available, these methods can be leveraged in disaster management IR.

Temporal and spatial locality is another important characteristic of emergency disaster data IR. This is because a disaster usually lasts for a short period of time and thus can only affect certain geographical areas. Recent research work on geospatial IE and monitoring cluster evolution [Li et al. 2012; Stefanidis et al. 2013; Tahrat et al. 2013] are examples that exploit such locality characteristics.

3.5. Disaster Storyline Generation

As an application of the techniques discussed earlier, we now introduce disaster storyline-generating systems [Lin et al. 2012; Wang et al. 2012; Zhou et al. 2014] which extract event summarization information from heterogeneous data sources.

In disaster management, people are interested in the developments of events as well as the differences among various phases as the events evolve over time. However, document summarization mostly engages in generating a compressed summary that delivers the major (or query-relevant) information of the original documents, while news topic detection and tracking usually aim at grouping news articles into clusters to present an event in a topic and then monitor future events related to the topic. As a result, these systems mainly focus on highlighting and summarizing events in a topic and lack the theme structure to capture the event evolution [Radev et al. 2000; Erkan



Fig. 9. Comparison between storyline generation and ground truth.

and Radev 2004; Shen and Li 2010; Wang et al. 2008]. A *disaster storyline* describes how a disaster evolves over time along the location attribute of the *events* and how the disaster affects different areas indicated by the description attributes. A disaster storyline offers a disaster evolutionary path not available in the perspective of either the temporal or geospatial dimension.

Here an *event* is denoted by a tuple (t, ℓ, s) , where t is the time that an event occurs, ℓ is the event location, and s is the textual description about the event. For example, (08/27/2011, New York City, “The five main New York City-area airports will be closed to arriving flights”) represents an event in Hurricane Sandy. To generate the storyline of a disaster, technologies such as event extraction, document summarization, and mathematical tools such as linear programming are employed. Figures 9(a) and 9(b) present the disaster storyline generated by the storyline generator, which is contrasted with the one manually labeled by humans about *Katrina* disaster. The figures show that the aforementioned system generates almost identical paths as the true evolutionary ones, which can boost information absorption significantly and thus improve SA.

4. ADDRESSING USER’S PERSONALIZED INFORMATION NEEDS IN DISASTER MANAGEMENT

When a disaster strikes, people at different locations, of different occupations, or playing different roles in disaster relief will have different information needs. For instance, before the hurricane arrives, the residents living close to a sea coast care about the impact of the tsunami caused by the hurricane, while the residents living far away from the coast care more about the chance of flooding in their specific area. After the hurricane, enterprise administrators would like to know about the latest traffic conditions and the status of their business partners, while inhabitants wish to know whether local supermarkets are open and water and power supplies have been restored. Therefore, the disaster-related information delivered should be *personalized* based on the role and interests of the user. In the next section, we present techniques in automatic information delivery and information sharing within communities that address this issue.

4.1. Automatic Information Delivery

4.1.1. News Recommendation. Automatic information delivery can be seen as an approach of news recommendation in the disaster information system. The goal is to automatically recommend the most relevant news or reports to each user. The relevance is calculated based on the profile and historical behaviors of a user. For instance, if a user’s profile indicates that the user lives in the coastal area of Miami Beach,

Table IV. Summary of Data Mining Techniques Used for Information Personalization in Disaster Management Systems

Methods	Usage Area	Description
Linear Regression	Automatic information delivery; relevance ranking	(+) Straightforward and efficient; (-) Time-consuming feature engineering; (-) Cannot capture feature correlation
Matrix Factorization	Automatic information delivery; relevance ranking; preference prediction	(+) Dimension reduction; (+) Capture feature correlation; (+) Identify latent factors; (-) Not straightforward
Ontology Integration	Automatic information delivery; relevance ranking; domain knowledge representation	(-) Depend on existing method; (+) Incorporate domain knowledge; (+) Meaningful and interpretable result ; (-) Not straightforward to incorporate ontology

then the news or reports on Miami Beach have a higher relevance than that of other locations. On the other hand, if the user frequently explores the news and reports on Miami Beach in the past, then other unread news on Miami Beach would have a higher relevance.

The news recommendation problem may be formulated as a relevance scoring problem. For a collection of news or reports, if we are able to precisely predict their relevance to each user, then the recommended results are simply the top K most relevant ones among them. Thus, the core problem is how to precisely estimate the relevance score between each piece of news and each user. Table IV summarizes different algorithms for information personalization in disaster management.

4.1.2. Algorithms.

Linear Regression. A widely adopted approach to calculate the relevance scores is the linear regression method [Seber and Lee 2012]. Linear regression is a well-studied method in data mining that uses a linear function to represent the relation between features and response values [Hastie et al. 2009]. The features of the relevance scores problem include user profiles (such as user’s location, user’s professionals and so on) and the characteristics of news (such as keywords, categories, and so on). The outputs of the regression models are the relevance scores. Specifically, if \mathbf{x} is the feature vector $(x_1, \dots, x_d)^T$ and \mathbf{w} is the parameter vector $(w_1, \dots, w_d)^T$, then the response value can be represented by a linear function of \mathbf{x} as $y = \mathbf{w}^T \mathbf{x}$. Using the database of historical data, the coefficient vector \mathbf{w} can be estimated through optimization.

Matrix Factorization. To capture the interaction between user features and news features, an aspect that is not addressed in the linear regression model, recommendation systems based on matrix factorization were recently proposed [Koren et al. 2009]. The data in the matrix factorization model is represented by a binary matrix \mathbf{D} , where an entry $\mathbf{D}_{i,j} = 1$ if the i th user clicked the j th report in the past, and $\mathbf{D}_{i,j} = 0$ otherwise. Certainly, there are many undefined entries in the matrix since not every user saw every piece of news. Thus, a recommendation system needs to solve a *matrix completion problem*, which is to determine those unfilled entries in \mathbf{D} . A reasonable assumption is that matrix \mathbf{D} can be approximately factored as $\mathbf{D} \approx \mathbf{UR}$, where \mathbf{U} and \mathbf{R} are the matrices consisting of the user vectors u_i and news items vectors r_j , respectively. Using the *alternative least squares* [Koren et al. 2009] optimization method, the missing entries in \mathbf{D} can be recovered.

4.1.3. Applications in Systems. The recommendation engine in the web-based prototype of BCIN in Zheng et al. [2010] has multiple processing steps, in which information from different data sources is integrated, ranked, and recommended based on users’ needs. The recommended information is then displayed in the dashboard module of

the system, which is a visualization panel facilitating information presentation and absorption. In a few practice sessions of the BCIN system the recommendation engine has gathered and recommended various news reports from government agencies and business companies. Recent recommendation systems [Li and Li 2014; Cantador et al. 2008] further incorporate ontology hierarchical information with news data and are able to generate high-quality, query-focused summarization, providing an easy access to structured and relevant information for users.

4.2. Disaster Information Sharing in Communities

Disaster data management requires updated and accurate information about the disaster, such as the locations and status of shelters, the geographical details of the affected areas, the availability of transportation means, and data about victims and relief personnel [Li et al. 2011, 2014]. These data, in addition to frequently changing, may belong to different organizations – for example, government/non-governmental organizations, or specific industries/enterprises – for example, shipping companies and retailing business. These public/private organizations and enterprises often interact with each other, forming explicit/implicit communities (e.g., members of the same industry sector or customers of the same enterprise) [Zheng et al. 2010]. A huge amount of information, originated from different communities and of various types or formats, may be collected and disseminated among communities during a disaster. Furthermore, operations on this information (e.g., collecting, integrating, storing, and querying) must be performed on an unprecedented scale [Li et al. 2014]. It is challenging to manage information processing tasks on such a large-scale, given the low latency requirement in disaster management.

To ensure appropriate information sharing within communities and facilitate information flow among different organizations, it is essential to identify how participants interact within communities in a disaster situation: that is, how the information is disseminated within a community or among multiple communities. This process may also reveal important information that is helpful in other disaster scenarios, such as early warning/recovery. Generally speaking, community information sharing in disaster management involves the following challenges:

- If the structure of communities is not available, how can we automatically detect the internal topology of each individual community and relations among different communities? What types of data can be utilized to perform such a community detection process?
- Once the structure of communities is identified, what is the optimal strategy to deliver the appropriate information to the right member in a community?
- Given that different communities have different interests and focuses, how can we get users within each community quickly involved in the communities? How can we reduce the user effort to identify individuals with similar interests or situations?

4.2.1. Community Detection in Disaster Management. Automated community discovery has been well studied in the context of IR [Xu et al. 2010b; Satuluri and Parthasarathy 2009]. The common strategy to detect communities involves utilizing social linking structure data among members. A myriad of approaches in this direction have been proposed, such as graph cut based partitioning and clique percolation methods [Fortunato 2010; Porter et al. 2009]. Another direction for community discovery is to use probabilistic models to capture the topics discussed by users within communities [Rosen-Zvi et al. 2010; Zhang et al. 2007]. For example, two generative models were proposed in Zhou et al. [2006] for community discovery in social networks, which combine probabilistic modeling with community detection. The sender recipient relations within social networks are regarded as the basis for modeling. A conditional model

for link analysis together with a discriminative model for content analysis is proposed in Yang et al. [2009]. By assuming that community memberships depend on topics of common interest and utilizing links between members in social graphs, Sachan et al. [2012] develop a generative model that can extract latent communities from social networks.

In disaster management, the techniques of detecting communities in social networks can also be applied to discover the information-sharing communities during disasters. The social links, as well as the textual content spreading among different organizations/enterprises/individuals in the disaster management system, are valuable information that can be exploited for community detection.

In addition to social links and textual content, the geo-locations of participants can be used to organize them into dynamic communities. For example, Zheng et al. [2010] consider the following two characteristics of geo-location information in the disaster recovery scenario: (1) the events that happen during the disaster are associated with one or several locations; and (2) similar disaster damage situations are often shared by spatially co-located members. Based on these observations, they develop a new method to detect communities using spatial clustering techniques. Another work that considers geographic information of community members is Ülgen [2005], in which three innovative applications of geospatial techniques are proposed, including (1) a neighborhood geographic information-sharing system, (2) a basic disaster awareness program monitoring facility, and (3) a school commute contingency program that tracks school-bus routes.

4.2.2. Information Sharing Strategies. During a disaster, large amounts of data – including status reports and news events – are generated through different information channels. It is imperative for a disaster management system to effectively and efficiently extract useful and pertinent information from raw data and dispatch that information to users. To achieve this, a lot of solutions utilizing distributed computing environments have been proposed. For instance, in Li et al. [2011], a community-based scalable cloud-computing infrastructure is presented for large-scale disaster management. The infrastructure coordinates various organizations and integrates massive amounts of heterogeneous data sources to effectively deploy personnel and logistics to people in need of aid. A strategy for forming a community-based virtual database was proposed in Li et al. [2014]. The proposed strategy connects local databases of suppliers to provide human and supply information for disaster management. The virtual database facilitates collaborative information sharing among organizations, enterprises, or individuals within communities. In a different direction, Yu et al. [2012] present a framework of community-based crowdsourcing, which consists of task-taker registration and task distribution components. The rationale is that community-based crowdsourcing promotes and better exploits the interactions among members in a community, thus can accomplish tasks more effectively and efficiently.

Another means of effectively disseminating information within a community or between communities is utilizing geospatial data of the members. However, it is often time-consuming to collect all the geo-locations of users, and sometimes the collected geospatial information is not accurate enough. In the past few years, a new source – messages in social media, in particular, microblogging – emerged as an alternative promising channel for collecting user geo-locations. A representative of such services is Twitter, in which the user's geospatial information (longitude/latitude) is contained in every post (tweet). In fact, social networks are playing increasingly more important roles in disaster management [Palen and Liu 2007; Shklovski et al. 2008; Vieweg et al. 2008; White et al. 2009; Li and Goodchild 2010]: from crowdsourcing disaster geospatial information, collectively generating and disseminating data, forming new

communities among spatially dispersed members, to distributively solving problems (see Section 3.3 for more discussions).

4.2.3. Community Recommendations in Disaster Management. Social media allow a user to broadcast within seconds messages to a large group of community members. A *community recommendation* (or user recommendation) system, on the other hand, offers a user suggestions regarding which other users in the network to connect with or which communities to join.

Community recommendation from a network perspective amounts to “finding missing edges in a user network.” Approaches to this problem usually make use of the node connections as well as the network structure – for example, network topological features-based methods [Liben-Nowell and Kleinberg 2007], supervised learning methods [Al Hasan et al. 2006], and relational learning methods [Popescul and Ungar 2003]. In social networks, the user-generated content (such as user relationships or posts) is a valuable and useful information source for modeling users’ preferences. Recently, several methods have been proposed for user recommendation in social media using the latent Dirichlet allocation (LDA) model for users’ stream [Pennacchiotti and Gurumurthy 2011].

Community recommendation helps users identify user groups with similar interests or situations. In the area of IR, researchers have developed several community recommendation approaches based on the characteristics of social media data. Specifically, a personalized community recommendation method based on collaborative filtering was introduced in Chen et al. [2008]. The proposed method considered different types of co-occurrences (e.g., semantic and user information) in social data. In Chen et al. [2009], association rule mining was first used to discover similarities between sets of communities, and LDA [Blei et al. 2003] was then applied to model user–community co-occurrences. A new principled framework of recommending users and communities in social media was proposed [Li et al. 2013] by employing a generative topic model to simultaneously discover community-oriented and user-oriented topics.

To expedite the information-spreading process within and among communities, a reasonable step is to automatically “push” newly published information to users. In this way, the time-consuming user queries are reduced and the latest information is dispatched to users in a more timely fashion. Along this direction, Zheng et al. [2011] present a disaster situation browsing system on mobile devices. For different pieces of information in a disaster situation, the system can automatically and interactively recommend potential recipients. In addition, a convenient interface is provided for users to explore other users’ recommendations and to share reports with others.

5. GENERAL DATA MINING AND LEARNING TECHNIQUES IN DISASTER MANAGEMENT

In this section, we present more applications of data-mining and machine-learning techniques in many other aspects of data-driven disaster management systems.

5.1. Natural Disaster Predictions

Natural disasters, such as hurricanes and earthquakes, are complex physical phenomena that result in enormous economic and human losses. One cannot overemphasize the importance of successful prediction, or even early warning, of these natural disasters. Early disaster predictions mostly rely on statistical analysis and involve complex models and strong assumptions. With recent advances in data acquisition techniques – such as global positioning systems (GPS) and high-resolution remote sensing – and the development of information technologies which makes, for example, a large volume of Internet-based volunteered geographic data available, increasingly advanced machine-learning and data-mining techniques have been developed for disaster

Table V. Summary of Prediction Methods in Disaster Management

Methods	Usage Area	Pros (+) / Cons (-)
Neural Network	Magnitude, location, and time of earthquakes; radon concentration	(+) Model non linear relations (-) Cannot predict time, epicentral location (-) Slow convergence
Decision Tree	Radon concentration, sequence data, premonitory factors	(+) Attribute construction (+) Analyze premonitory factors (-) Less accurate
Clustering	Precursory events for seismic activities, dynamics of earthquakes, tropical cyclone tracks	(+) Unsupervised (-) Sensitive to initialization
Association Rule	Intensity change of tropical cyclone	(+) Understand the coupling relationships of the physical processes (-) Sensitive to initialization
Others	Parameters of earthquakes, intensity of hurricanes	N/A

predictions. These techniques include neural networks, decision trees, clustering, and association rule mining and have been reported to be effective especially for short-term predictions. In this section, we discuss some recent progress of these prediction methods and use earthquakes and hurricanes as examples. A summary of these methods is presented in Table V.

5.1.1. Neural Networks–Based Prediction Methods. The neural networks–based approaches aim to model the nonlinear relationship between physical phenomena and various environmental parameters. Three different neural networks – feed-forward Levenberg-Marquardt backpropagation (LMBP) neural network (NN), radial-basis function (RBF) NN, and recurrent NN – are investigated in Panakkat and Adeli [2007] to predict the magnitude of the largest seismic event in the following month with eight physical and mathematical earthquake parameters called *seismicity indicators* as NN inputs. These NN models are trained and tested using data from Southern California and the San Francisco bay region. It is reported in Panakkat and Adeli [2007] that the recurrent NN model outperforms other models and works the best for earthquakes of magnitude between 4.5 and 6.0. Similar studies have been carried out by the authors to predict earthquakes of magnitude greater than 6.0 [Adeli and Panakkat 2009] as well as the location and time of occurrence of future earthquakes [Panakkat and Adeli 2009]. More recently, a population-based algorithm was proposed in Shah and Ghazali [2011] to improve the training process of the Multilayer Perceptron (MLP) model, which is used for earthquake predictions. A layered neural network (LNN) model is proposed in Negarestani et al. [2002] to “estimate the radon concentration in soil related to the environmental parameters.” Such a relationship can be used to distinguish abnormal variations caused by anomaly phenomena in the earth (e.g., earthquakes) from those normal variations due to environmental changes. The method is shown to be superior to the linear computational technique, as it can better “estimate radon variations related to environmental parameters that may have a non-linear effect on the radon concentration in soil.”

5.1.2. Decision Trees–Based Prediction Methods. A decision-tree model is used in Zmazek et al. [2003] to study the correlation of radon concentration in soil gas to environmental parameters. It is found that the decision-tree model can predict radon concentration of correlation 0.8 with observed data. However, such a correlation decreases before earthquakes with local magnitude 0.8–3.3. The general problem of early prediction of sequence data – whose many applications include early prediction of earthquakes and tsunamis – is investigated by Xing et al. [2008], who study two methods solving the

problem of supervised learning on sequence data. The first is a sequential classification rule (SCR) method, which extracts discriminative features from the training set sequences first and then mines a set of sequential classification rules to form a classifier. The second is a generalized sequential decision tree (GSDT) method, which constructs each attribute using a small set of features that have high early prediction utility. The proposed approaches are able to use only very short prefixes of the sequences and achieve very good accuracy. In Sikder and Munakata [2009], rough-set and decision-tree methods are compared for identifying the premonitory factors for earthquake. These two machine-learning techniques are considered particularly suitable for earthquake prediction because they are nonparametric and less restrictive on model assumptions.

5.1.3. Clustering-Based Prediction Methods. A novel two-level resolution cluster method, which also employs feature extraction and data visualization techniques, is introduced in Dzwiniel et al. [2005a, 2005b] to recognize precursory events for seismic activities around the Japanese island areas between 1997 and 2003. Specifically, local clustering on original observed seismic activity data and nonhierarchical clustering on the data in the feature space (whose coordinates are nonlinear functions of the original measurements) are combined to study the dynamics of earthquakes. In Kim et al. [2011], a fuzzy *c*-means clustering method is proposed to classify tropical cyclone tracks in the western North Pacific between 1965 and 2006. The optimal number of clusters are determined by four scalar validity measures: partition coefficient, partition index, separation index, and Dunn index. Their results show that 855 tropical cyclone tracks are clustered into 7 groups, each with distinctive features, such as lifetime, landfall region, intensity, and distribution of rainfall. Such a classifying result is very useful for predicting the behaviors of future tropical cyclones in that area. The authors later propose a seasonal tropical cyclone forecast model by combining the fuzzy *c*-means clustering method with statistical dynamics [Kim et al. 2012].

5.1.4. Association Rule Mining–Based Prediction Methods. To predict the intensity change of a tropical cyclone (TC) in the next 6 hours, the Apriori algorithm [Agrawal and Srikant 1994] is applied in Tang et al. [2005], which uses 13 attributes – such as date and hour of observation, distance between TC center to nearest land, and average sea surface temperature – from the raw data to identify 5 rules on how a TC would change its intensity. The same prediction problem is formulated as a supervised data-mining problem in Chatzidimitriou and Sutton [2005], in which the objective function is set to be the prediction accuracy with regard to changes in wind speed. The problem of hurricane trajectory prediction is studied in Dong and Pi [2013] by a data-mining approach, which applies an association analysis on the historical data of hurricane movements.

5.1.5. Other Methods. In addition to these developments, other machine-learning and data-mining techniques have been developed for disaster predictions [Azam et al. 2014]. For example, fuzzy logic is applied to synthetic seismic data to predict the parameters of earthquakes [Aydin et al. 2009] for earthquake prediction, and a hurricane intensity prediction model is developed in Su et al. [2010] by integrating feature weight learning (WFL) and the Extensible Markov Model (EMM). Song et al. [2014b] and Fan [2014] have developed an intelligent system as well as related probabilistic techniques to automatically discover, analyze, and simulate population evacuations during disasters such as the Great East Japan Earthquake and the Fukushima Daiichi nuclear accident.

Table VI. Summary of Risk and Insurance Modeling

Usage Area	Methods	Description; Pros(+)/cons(-)
Risk Analysis	Bayesian Network	(+) Capture hidden interaction between parameters; (+) Domain knowledge integration; (-) High complexity of training Bayesian networks
	Back Propagation Neural Networks	(+) Capture nonlinear relationship; (+) Agree more with actual situation; (-) Feature construction
	Other Methods of Risk analysis from images: J48, LSSVM, and others	(-) Labelled data for supervised learning; (+) More accurate; (-) Feature construction

Disaster type	Insurance model is composed of components	Assessment results
Earthquake	(1) Stochastic event generator (2) Hazard analysis procedure (3) Vulnerability analysis procedure (4) Financial analysis procedure	- Annual Exceeding Probability Curve - Average Annual Loss - Probable Maximum Loss
Flood	(1) Rainfall event module (2) Hydraulic module (3) Vulnerability module (4) Financial loss module	- Total Loss - Insured Loss - Loss Exceeding Probabilities
Hurricane	(1) Atmospheric Component (2) Engineering Component (3) Actuarial Component	- County-wise loss and the total loss for the entire state of Florida - Average Annual Loss

Fig. 10. List of insurance models for different kinds of disasters.

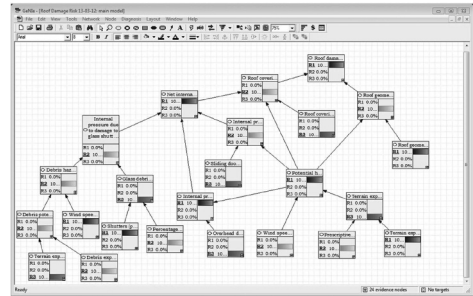


Fig. 11. A Bayesian network for predicting windstorm-induced damage to roof structures.

5.2. Risks and Insurance Modeling

As large-scale disasters such as hurricanes, earthquakes, and floods usually cause devastating damage, it is important to have experts from multidisciplinary backgrounds build risk assessment and insurance models that estimate and predict the damage ratio. With risk analysis and accurate assessment of damage caused by different disasters, governments can adopt suitable measures to reduce the risk, plan risk control strategies accordingly, and legitimately prevent the insurance companies from demanding unreasonable insurance policy rates.

A well-established risk assessment model or a reliable, validated insurance model is usually composed of several components, as illustrated in Figure 10 for natural disasters such as earthquakes [Tsai 2010; Hsu et al. 2006; Tsai and Chen 2010; Hsu et al. 2013], floods [Hsu et al. 2011], or hurricanes [Hamid et al. 2010]. A summary of risk and insurance modeling is given in Table VI.

To accurately assess the damage caused by unpredictable earthquakes, a framework to perform catastrophe modeling and risk assessment is proposed in Hsu et al. [2006], which includes the following four major components: (1) A stochastic event generator, which utilizes historical earthquake data to simulate and produce a list of stochastic incidents; (2) A hazard analysis procedure, which calculates the expected intensity of the earthquakes for specific locations; (3) A vulnerability analysis procedure, which computes the likely loss caused by a building’s exterior and interior damage, interruption of transportation services, and so on. This loss assessment computation is performed based on the fragility curve computed for the building or probability of various damage caused by earthquakes with a given intensity; and (4) A financial analysis procedure, which calculates the economic loss due to the damage of buildings for all the involved parties. A similar structure of the insurance model is also built for other disasters. Usually, stochastic records are generated based on historical information by simulation models such as the Monte Carlo method, with parameters tuned accordingly, and

possible damage is considered from different perspectives, then later converted into financial loss.

Within this insurance loss model, much effort has been devoted to the following questions: How do we associate disaster factors with overall damage? How do disaster factors interact with each other and lead to different levels of damage? How do we identify the most important factors in the risk assessment model and insurance model? To tackle these questions, data-mining methods have been applied on huge datasets with strong nonlinearity, outliers, and noise to capture the correlation between disaster factors and estimated losses.

5.2.1. Bayesian Network in Risk Analysis. Risk analysis is computationally complex due to the large number of parameters involved and interdependencies among them are difficult to understand and predict. Bayesian networks have been proved to be good at leveraging the uncertain expert knowledge in decision analysis and graphically representing the interdependencies among underlying parameters [Heckerman 2008]. To study the risk analysis on windstorm-induced damage to roof structures, Durgaprasad and Subba Rao [2012] employ a Bayesian network to model the hidden interaction between parameters and their unknown mechanism to cause damage. Figure 11 illustrates how a Bayesian network models the interdependencies between windstorm-induced damage and roof structures. Another framework to build Bayesian networks, which leverages the integration of domain knowledge and spatial data, appears in Li et al. [2010]. First, a nonparametric unsupervised learning method called Kernel Density Approximation (KDA) is used to preprocess the data so that the influence of the geospatial features can be extracted. Then, two well-known methods – Minimum Description Length (MDL) and Principal Component Analysis (PCA) – are applied for discretization and feature selection, respectively. The method is evaluated for flood disaster study and compared against nine other methods, such as C4.5 decision tree, random forest, and logistic regression, and is shown to excel in the detection of high-risk, precision, and receiver operating characteristic (ROC) areas.

5.2.2. Back Propagation Neural Networks in Risk Analysis. It is well known that usually the coupling between a tropical cyclone and the disaster factors that it incurs is strongly nonlinear. This suggests that the Back Propagation (BP) artificial NN, which was developed for fuzzy problems, is a good candidate for cyclone risk analysis. By applying PCA, a BP network is built in Lou et al. [2012] over the selected and highly correlated assessment factors so that an effective risk assessment model can be constructed to estimate economic losses caused by tropical cyclones. The BP network model (adopting the Levenberg-Marquardt (LM) algorithm) and entropy-based Attribute Recognition (AR) model are evaluated in Yang et al. [2009] for flood disaster grade analysis. The latter model is reported to agree more with actual situations.

5.2.3. Other Methods. Optical satellite images are combined with a digital terrain model (DTM) and hydrological network in Peter et al. [2013], together with several other machine-learning methods to distinguish flooded areas from unflooded areas. Among all the methods, decision tree algorithm J48 outperforms other methods in detecting the flooded area for the Ljubljana moor floods in 2010. In Cheng and Hoang [2012], an evolutionary fuzzy least-squares support-vector machine (LSSVM) inference model is applied to evaluate the risk of bridge damage. In addition, eigenvalue and projection pursuit methods are used in Hexiang and Qingjuan [2007] to forecast the risk of flood; a fuzzy classification model and linear weighted mean model are used in a two-level fuzzy synthetic evaluation system in Liu et al. [2009b] to evaluate the fire risk ranking of buildings; and a dynamic assessment model based on vague sets,

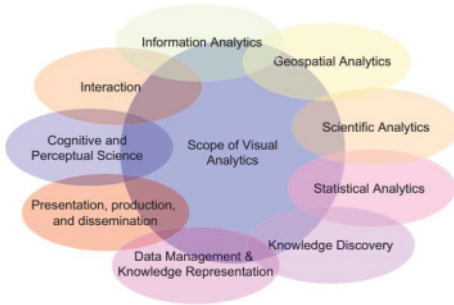


Fig. 12. Knowledge integration in visual analytics.

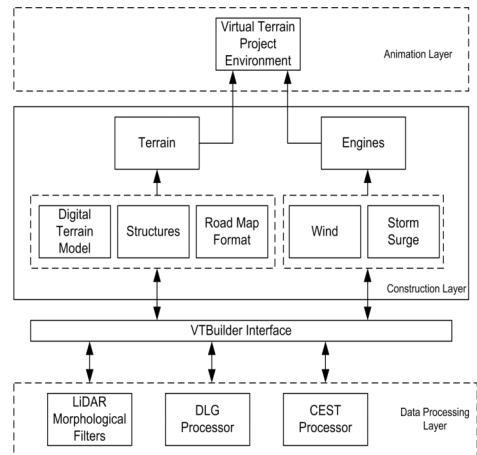


Fig. 13. System architecture of a 3D storm surge simulation.

information axiom, and comprehensive assessment method is proposed in Liu and Zuo [2015] for evaluating urban multihazard, integrated natural disaster risk.

5.3. Visual Analytics

Visual analytics is “the science of analytical reasoning facilitated by visual interactive interfaces” [Thomas and Cook 2006], which can attack certain problems whose size, complexity, and need for closed coupled human-machine analysis may make them otherwise intractable. Visual analytics is a multidisciplinary field that combines data analysis, data management, information visualization, scientific visualization, human cognition and perception, and knowledge discovery [Keim et al. 2008] (see Figure 12). People apply visual analytics techniques to synthesize information in novel ways in order to gain insight or identify patterns from massive, complicated, and sometimes even conflicting data.

5.3.1. Visual Analytics Systems. Conceptually, the guidelines for designing a visual analytics system include the following [Cook and Thomas 2005; Thomas and Cook 2006]: (1) support integration and transformation of different data formats into a unified representation; (2) support the understanding of both spatial and temporal data; (3) support the understanding of massive and dynamic data; (4) support the understanding of ambiguous, incomplete, or even conflicting data; (5) support adaptedness to user needs and task situations; and (6) facilitate interactions between users and the system to support cognition, perception, and analytic reasoning. Scholtz [2006] suggests five aspects of visual analytic environments as metrics for evaluation: SA, collaboration, interaction, creativity, and utility.

5.3.2. Why Disaster Management Needs Visual Analytics. Disaster management is a very complex task that involves processing large amounts of heterogeneous data. Visual analytics is a proven approach that interactively combines human intuition and perception, methods from knowledge discovery in databases (KDD), and tools in statistics and mathematics to derive patterns, knowledge, and insight, ultimately leading to better decision-making. Turning the information overload into opportunities is the overarching driving vision of visual analytics. In the domain of emergency management, depending on the specific cases, visual analytics is helpful in determining the extend

of damage, identifying objectives, assigning priorities, and facilitating effective coordination among various organizations for more efficient operations, such rescue and evacuation in the disaster zone. Moreover, a visual analytics system can visually and interactively display the incoming data, thus helping commanders better comprehend the situation and make the right decisions.

5.3.3. Visual Analytics in Disaster Management. Tomaszewski and colleagues [Tomaszewski 2008; Tomaszewski and MacEachren 2010] initiate a conceptual approach for modeling geo-historical context (GHC) that can be obtained through entity extraction and relevance ranking of retrieved documents. GHC encompasses three aspects of context – spatial (geographical), temporal (historical), and conceptual – and provides a framework in which the spatial, temporal, and conceptual context information of text documents can be visually represented to facilitate decision-making in disaster management. Another disaster management framework, SoKNOS [Döweling et al. 2009; Paulheim et al. 2009; Babitski et al. 2011], is an ontology-based, user-centric, interactive system that aggregates heterogeneous information and provides visualizations of the situation on different levels of generalization.

An important application of visual analytics in disaster management is to enhance SA. For example, visual analytics techniques have been applied to render rainfall accumulations in real time at the resolution of hydrologic units and provide decision-makers a powerful tool for assessing neighborhood flooding situation [Liu et al. 2009a] and to perform risk assessment or security vulnerability assessment for the critical infrastructure of a city [Kulawiak and Lubniewski 2014]. Uncertainties in disaster development sometimes complicate the planning of mitigation efforts, as there may be many possible scenarios, each calling for different response plans. An integrated solution for floor management is introduced in Waser et al. [2014], which integrates scalable multidimensional ensemble simulations and logistics optimization with rich visualization capabilities. The system is able to generate and simulate many disaster scenarios as well as the interactive 3D views of the corresponding response plans.

Social media offer informative and timely information, which make them an indispensable information resource for disaster relief. As users tend to update more frequently, microblogging, such as Twitter, is a faster channel for information sharing than traditional blogging. On the other hand, the geospatial information of tweets provided by the Twitter API makes tweets a large data source of volunteered geographic information, which has been exploited recently for event exploration [Marcus et al. 2011], operational crisis management [Terpstra et al. 2012], and discovering significant spatiotemporal events [Thom et al. 2012]. With respect to disaster management, TwitterHitter [White and Roth 2010] is an application that utilizes geovisual analytics and information visualization techniques to gain spatiotemporal insights from tweets and can be applied in a variety of tasks, such as criminal investigation and disaster management. Another similar application [MacEachren et al. 2011], which is also based on geovisual analytics on tweets to support crisis management, focuses more on information foraging and sense-making by leveraging the place, time, and concept characteristic of tweets.

Disaster management often involves coordinating personnel and operations from multiple locations. To synchronize distributed users under a visualization environment, one solution is to use an adaptive time adjusting algorithm to modify the output time of different units [Natarajan and Ganz 2009] or to use open-source web standards developed by the Open Geospatial Consortium (OGC) and WorldWideWeb Consortium (W3C), the latter of which is built on the web-based client-server models to serve shared content and geospatial information [Heard et al. 2014]. A system that is based on a visualization technique called “Framy” and designed specifically for mobile devices is presented in Paolino et al. [2010].

Analysis of public behaviors during and after disasters is important for evacuation planning. Visual analytics tools have been applied to the data collected through fieldwork and questionnaires in Gismondi and Huisman [2012] to represent and analyze residents' movement paths over time during a post-earthquake period, and applied to tweets in Chae et al. [2014] to provide an interactive spatiotemporal visualization and facilitate temporal pattern analysis, spatial analysis, and decision support.

5.4. 3D Simulation of Disaster Impacts

The deadliest natural disasters, such as hurricanes, are a big threat to the United States' East and Gulf Coasts. Accounting for 90% of hurricane-related fatalities, the rise of ocean water due to a hurricane's strong winds, or storm surge, is the greatest cause of human life and property losses. Furthermore, saltwater flooding is the major origin of damage to buildings and infrastructure [Pava et al. 2010a]. As a hurricane approaches a coastal area, various storm-surge models are used to predict the potential height of the storm surge in the projected rainfall area. Typically, the models are presented to coastal residents as color-coded, two-dimensional maps. However, for an untrained eye, it is hard to relate such two-dimensional maps to the daily three-dimensional (3D) real world. Much research effort has been carried out to develop 3D storm surge visualization systems that better illustrate the disaster situation and facilitate residents to make potentially life-saving evacuation decisions [Singh et al. 2005; Zhang et al. 2006; Pava et al. 2010a; Zhang et al. 2006]. In this section, we will focus on the 3D simulation of hurricanes and storm-surge flooding.

Visualization of the storm surge plays an important role in creating an interactive 3D animation environment for real-time modeling of hurricanes. Existing graphic animation engines can be used to animate objects such as buildings, trees, human figures, vehicles, and their behaviors in a hurricane situation to portray realistic storm-surge effects. The open-source Virtual Terrain Project (VTP) is usually selected as the 3D visualization environment platform for storm-surge animations. The VTP allows efficient animations of storm surge, wind, rain, traffic, vegetation, lightning, and more using OpenGL [Khronos Group 1997].

5.4.1. 3D Storm Surge Simulation System Architecture. Figure 13 presents the high-level view of the architecture of a 3D storm surge simulation system [Zhang et al. 2006]. The data-processing layer provides the facilities to process Light Detection and Ranging (LiDAR), the Coastal and Estuarine Storm Tide (CEST) model [Pava et al. 2010a; Zhang et al. 2008], and Digital Line Graph (DLG) data into formats that can be recognized by VTBuilder [Discoe 2002]. The output of VTBuilder is then used to construct 3D representation of terrain, buildings, and roads in the construction layer. Figure 14 shows a screen capture of FIU building models generated by VTBuilder. Wind and storm surge data from the CEST model are used to drive the individual graphics animation engines for wind, storm surge, vegetation, rain, traffic, and ocean waves. Access to the VTBuilder is bidirectional so that 3D models may be modified after the input data have been processed. The graphics animation engines and 3D models are then passed on as input for final rendering. This last layer provides the user with an interactive and 3D navigation of the storm surge visualization, along with animated storm surge, rain, lightning, clouds, vegetation, traffic, sun, and so forth.

5.4.2. Ocean Water and Rain. The ocean animation is directly tied to the storm surge engine, which, based on the latest storm-surge data, computes the height of the flat mesh used to represent the ocean. The mesh of the ocean is then animated to model ocean waves. In visualizations, the ocean is modeled as a flat mesh whose height may rise or fall depending on the latest storm-surge data. The surface of the mesh is animated by an OceanWave engine using Fournier's model for ocean waves [Fournier

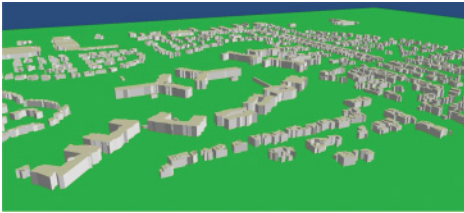


Fig. 14. The FIU building model.

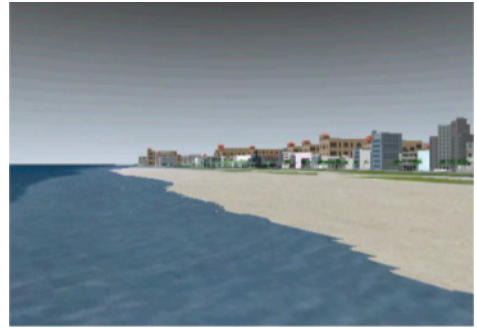


Fig. 15. Ocean wave animation.

and Reeves 1986]. Figure 15 shows a screen capture of the water animation using the model described. To produce more realistic scenery, an ocean water spray effect is added to simulate when ocean waves crash against the shore. Such a spray effect is modeled using a particle system, in which each water droplet resulting from the ocean spray is represented by a particle. The emission of spray from the ocean wave must meet two conditions: (1) the particle's speed has to be greater than the wave's phase speed and (2) the curvature of the surface has to be high.

The rain engine used to render precipitation is also based on a particle system. Each raindrop is modeled as a particle whose size and current position on the screen are stored in an array. At every frame in the animation, the array is updated to reflect the next position of the raindrop as it falls from the sky. The position of a rain particle is a function of the current wind speed and direction. These data are then used to render raindrops in the animation.

5.4.3. Clouds, Trees, and Lightning. A combination of static and dynamic techniques is utilized to enhance the animation effects of clouds, trees, and lighting.

Two cloud patterns are used to represent the sky in normal and hurricane conditions: a normal sky is clear with few clouds, while the hurricane sky is filled with dark and thick clouds to provide users with a sense of severe weather.

Tree animation is implemented by a billboard model in which each tree is represented by planar-texture-mapped geometry with various angles. Such a billboard model is easy to implement and requires only moderate computing resources. A vegetation engine is designed to simulate tree bending, rotation, and damage in hurricane storm-surge scenarios. For example, trees shake along the wind direction to an extent that is determined by the wind velocity.

Lightning is modeled as a sequence of lines and is approximated by a tree growing inversely from the sky to the ground. The root of the lightning tree is randomly generated at an arbitrary position in the sky. The subbranch of the lightning tree is also generated randomly from the main branch once the main branch is rendered.

Traffic is a very important engine in the 3D simulation system. A traffic animation system based on road segmentation is introduced in Li et al. [2009]. The system simulates the impacts of storm surge on the road systems in the coastal regions of southern Florida. To reduce computation cost, simple geometric shapes with added realistic effects are used to represent objects. The traffic system consists of two kinds of objects: road objects and vehicle objects.

A road is represented with an array of connected, adjacently nonparallel line segments. Such a representation, as illustrated in Figure 16, facilitates a simple animation for vehicles to move on the roads. The road segment attributes include coordinates and

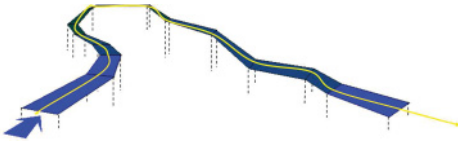


Fig. 16. 3D simulation of road segment.

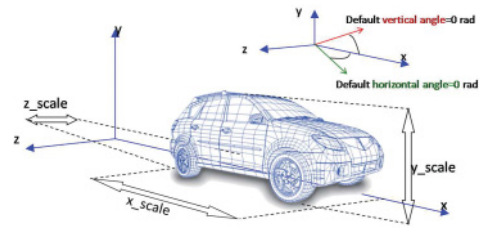


Fig. 17. Constant vehicle parameters.

length and direction angles, which are used to simulate vehicles moving on the roads and are stored in the road segment objects.

A vehicle in the traffic system is represented by a vehicle object, which stores attributes necessary for traffic simulation. There are two categories of attributes, constant and dynamic: constant attributes are assigned values during every system initialization and do not change after that, while dynamic attributes are updated continuously during an animation. These attributes, illustrated in Figure 17, usually specify the relationship between a vehicle and a road segment.

To render high-quality animation of storm-surge impact on traffic, three additional features are implemented to render interactions within or beyond the traffic system: interactions between vehicles, vehicle responses to surge flooding, and automatic distribution of lampposts along roads.

6. CONCLUSION

Data-driven disaster information management is an emerging and interdisciplinary research area that calls for collaborations between researchers with various backgrounds. In this article, we present a comprehensive survey on recent efforts in leveraging advanced data management and analysis techniques to build better disaster information management systems. To this end, we review and categorize results from a variety of research areas. Our hope is that this survey not only provides an overview of existing data-driven techniques for disaster management but also sheds light on specific application tasks related to disaster information management.

We end the survey with a compiled list of research problems. We classify the problems into the following three categories: (1) information and knowledge discovery in disaster management, (2) integration with the Geographic Information System (GIS), and (3) disaster data analysis platform.

- Information and Knowledge Discovery.** Within this category, the following problems require further investigation: (a) Support intelligent ad hoc or continuous queries on heterogeneous, multisource streaming disaster data. To achieve this, effective and efficient information and knowledge discovery algorithms must be developed considering both the context and the stakeholder's profile. In addition, domain-specific parameterized queries should be efficiently supported. (b) Facilitate real-time queries and discover patterns and information snippets spanning across multiple streams through developing high-throughput indexes and advanced knowledge discovery techniques for streams. (c) Build customized information extraction algorithms that learn from the interaction between domain experts and the system. (d) Incorporate uncertainty and possible adversariness in handling and delivering of data.
- Integration with Geographic Information System (GIS).** A Geographic Information System [Goodchild 2009] is an information system this is able to integrate,

store, edit, analyze, share, and display geographically referenced information. GIS applications are useful in many disaster management–related activities, such as creating hazard inventory maps, locating critical facilities, managing associated databases, and performing vulnerability assessment. An important research direction is the integration between a GIS and other components in disaster management systems. Such an integration will greatly enhance user experience and improve information and knowledge discovery in disaster management. Specifically, the following problems are worth further investigating: (a) automate or crowdsource the linkage construction between information/data and geomap in real-time – once life, property, and environmental data records are combined with hazards, emergency management personnel can formulate mitigation, preparedness, response, and recovery program needs [Horita et al. 2013]; (b) provision intelligent alerts and location broadcasting when people enter a dangerous zone or an unusual location ; (c) utilize social networks not only as data sources but also tools for help and locating people in danger.

—**Disaster Data Analysis Platform.** Disaster data are in general from different sources and are heterogeneous in nature. With regard to data analysis in disaster management, effective and efficient methods are required to analyze an ever increasing amount of data, discover the interdependencies of data, and extract useful information and knowledge for problem-solving and decision-making. An additional challenge is how to integrate data with great diversity and provide an aggregated view of the application domain. Such a diversity may exist because data are from heterogeneous sources with different levels of redundancy, accuracy, and uncertainty or may be due to different characteristics of data (e.g., structured/unstructured, real-time streams/static data). As a result, an integrated platform is often necessary for disaster data analysis. In our prior work, we have developed a distributed computing platform FIU-Miner [Zeng et al. 2013] to support efficient data analysis and an online geospatial data analysis and visualization system called “TerraFly Geocloud” [Zhang et al. 2015; Zeng et al. 2014]. Some interesting research directions include: (1) further integrating these types of platforms into disaster management systems; (2) providing a unified storage pipeline for heterogeneous, multisource disaster data; (3) building a computing pipeline capable of real-time services; and (4) enabling efficient information digestion and quick response for users in disaster situations.

REFERENCES

- Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. 2011. Predicting flu trends using Twitter data. In *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 702–707.
- Hojjat Adeli and Ashif Panakkat. 2009. A probabilistic neural network for earthquake magnitude prediction. *Neural Networks* 22, 7, 1018–1024.
- Rakesh Agrawal and Ramakrishnan Srikant. 1994. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases, (VLDB’94)*, Vol. 1215. 487–499.
- M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki. 2006. Link prediction using supervised learning. In *SDM’06: Workshop on Link Analysis, Counter-terrorism and Security*. SIAM, Maryland.
- Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, ACM, New York, 1568–1576.
- Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli. 2010. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems* 16, 6, 345–379.
- Ilhan Aydin, Mehmet Karakose, and Erhan Akin. 2009. The prediction algorithm based on fuzzy logic using time series data mining method. *World Academy of Science, Engineering and Technology* 51, 27, 91–98.
- Faisal Azam, Muhammad Sharif, Musarat Yasmin, and Sajjad Mohsin. 2014. Artificial intelligence based techniques for earthquake prediction: A review. *Science International Lahore* 4, 26, 1495–1502.

- Grigori Babitski, Simon Bergweiler, Olaf Grebner, Daniel Oberle, Heiko Paulheim, and Florian Probst. 2011. SoKNOS—using semantic technologies in disaster management software. In *The Semantic Web: Research and Applications*, 183–197.
- Bahareh Bagheri, Maryam Pourmahyabadi, and Hossein Nezamabadi-pour. 2013. A novel content based image retrieval approach by fusion of short term learning methods. In *5th Conference on Information and Knowledge Technology (IKT'13)*. IEEE, Shiraz, Iran, 355–358.
- BCIN. 2006. Business Continuity Information Network. Retrieved February 5, 2017 from <http://www.bizrecovery.org/>.
- Christian Becker and Christian Bizer. 2008. DBpedia mobile—a location-aware semantic web client. *Proceedings of the Semantic Web Challenge* 13–16.
- Christian Bizer and Richard Cyganiak. 2009. Quality-driven information filtering using the WIQA policy framework. *Web Semantics: Science, Services and Agents on the World Wide Web* 7, 1, 1–10.
- C. Bizer, T. Heath, and T. Berners-Lee. 2009. Linked data: The story so far. International. *Journal on Semantic Web and Information Systems* 5, 3, 2–9.
- D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.
- Iván Cantador, Alejandro Bellogín, and Pablo Castells. 2008. Ontology-based personalised and context-aware recommendations of news items. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, IEEE, New Brunswick, NJ, 562–565.
- Mifan Careem, Chamindra De Silva, Ravindra De Silva, Louiqa Raschid, and Sanjiva Weerawarana. 2006. Sahana: Overview of a disaster management system. In *International Conference on Information and Automation (ICIA'06)*. 361–366.
- CERCS. 2013. *Big Data and Disaster Management: A Report from the JST/NSF Joint Workshop*. Technical Report. Georgia Institute of Technology, enter for Experimental Research in Computer Systems, 266 Ferst Drive, Atlanta, GA 30332-0765, USA.
- Junghoon Chae, Dennis Thom, Yun Jang, SungYe Kim, Thomas Ertl, and David S. Ebert. 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics* 38, 51–60.
- Deepayan Chakrabarti and Kunal Punera. 2011. Event summarization using tweets. *ICWSM* 11, 66–73.
- Chia Hui Chang, Mohammed Kayed, Moheb R. Girgis, and Khaled F. Shaalan. 2006. A survey of web information extraction systems. *IEEE Transactions on Knowledge and Data Engineering* 18, 10, 1411–1428.
- Kyriakos Chatzidimitriou and Andrew Sutton. 2005. Alternative data mining techniques for predicting tropical cyclone intensification. In *American Association for Artificial Intelligence*, Vol. 37. Citeseer, New Jersey, 99–128.
- Rashmi Chauhan, Rayan Goudar, Rohit Rathore, Priyamvada Singh, and Sreenivasa Rao. 2012. Ontology based automatic query expansion for semantic information retrieval in sports domain. In *Eco-friendly Computing and Communication Systems*. Springer, Berlin 422–433.
- Rashmi Chauhan, Rayan Goudar, Robin Sharma, and Atul Chauhan. 2013. Domain ontology based semantic search for efficient information retrieval through automatic query expansion. In *2013 International Conference on Intelligent Systems and Signal Processing (ISSP'13)*. IEEE, Gujarat, India, 397–402.
- Lingji Chen, Harshavardhan Achrekar, Benyuan Liu, and Ross Lazarus. 2010. Vision: Towards real time epidemic vigilance through online social networks: Introducing SNEFT—social network enabled flu trends. In *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*. ACM, New York, NY, 4.
- W. Y. Chen, J. C. Chu, J. Luan, H. Bai, Y. Wang, and E. Y. Chang. 2009. Collaborative filtering for Orkut communities: Discovery of user latent behavior. In *Proceedings of the 18th International Conference on World Wide Web*. ACM, Madrid, Spain, 681–690.
- W. Y. Chen, D. Zhang, and E. Y. Chang. 2008. Combinational collaborative filtering for personalized community recommendation. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 115–123.
- Min-Yuan Cheng and Nhat-Duc Hoang. 2012. Risk score inference for bridge maintenance project using evolutionary fuzzy least squares support vector machine. *Journal of Computing in Civil Engineering* 28, 3, 1–9.
- Gong Cheng and Yuzhong Qu. 2009. Searching linked objects with falcons: Approach, implementation and evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5, 3, 49–70.

- J. Chidambaram, C. Prabhu, P. Rao, Rajeev Wankar, C. Sreevallabh Aneesh, and Arun Agarwal. 2008. A methodology for high availability of data for business continuity planning/disaster recovery in a grid using replication in a distributed database. In *TENCON 2008/2008 IEEE Region 10 Conference*. IEEE, 1–6.
- Scott Clark, Syrus Mesdaghi, Dan Palumbo, and Chris Holmes. 2013. Rapid open geospatial user-driven enterprise (rogue) joint capability technology demonstration white paper. LMN Solutions, Reston, VA.
- The Conference Board. 2006. Preparing for the worst: A guide to business continuity planning for mid-markets. In *Executive Action Series*. The Conference Board of Canada, Ottawa, Canada.
- 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.
- Gerald Corey. 2012. *Theory and Practice of Counseling and Psychotherapy*. Brooks Cole, Boston, MA.
- Jim Cowie and Wendy Lehnert. 1996. Information extraction. *Communications of the ACM* 39, 1, 80–91.
- CNN Wire Staff. 2011. Japan: 3 Nuclear Reactors Melted Down 'C News Story 'C KTVZ Bend. Ktvz.com. Archived from the original on 28 July 2011.
- DDST. 2015. The Puerto Rico Disaster Decision Support Tool (DDST). Retrieved February 5, 2017 from <http://www.udel.edu/DRC/DDST/>. (2015).
- Gerald DeJong. 1982. An overview of the FRUMP system. *Strategies for Natural Language Processing* 113, 149–176.
- Ben Discoe. 2002. Virtual terrain project. Retrieved February 5, 2017 from <http://vterrain.org/Doc/VTBuilder/overview.html>.
- Lars Döhling and Ulf Leser. 2011. Equatornlp: Pattern-based information extraction for disaster response. In *Terra Cognita 2011 Workshop, Foundations, Technologies and Applications of the Geospatial Web*, Vol. 38. Springer, Bonn, Germany, 127–138.
- X. Dong and D. C. Pi. 2013. Novel method for hurricane trajectory prediction based on data mining. *Natural Hazards and Earth System Science* 13, 12, 3211–3220.
- Sebastian Döweling, Florian Probst, Thomas Ziegert, and Knut Manske. 2009. Soknos!An interactive visual emergency management framework. In *GeoSpatial Visual Analytics*. Springer, Germany, 251–262.
- Janjanam Durgaprasad and Palivela Subba Rao. 2012. Modeling risk from large-scale disasters by integrating fragmented knowledge: Decision tool. *Leadership and Management in Engineering* 12, 3, 157–168.
- Witold Dzwinel, David A. Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka, and Takeo Ito. 2005a. Cluster analysis, data-mining, multi-dimensional visualization of earthquakes over space, time and feature space. In *Nonlinear Processes in Geophysics*, Vol. 12. Citeseer, New Jersey, 117–128.
- Witold Dzwinel, David A. Yuen, Krzysztof Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka, and Takeo Ito. 2005b. Nonlinear multidimensional scaling and visualization of earthquake clusters over space, time and feature space. *Nonlinear Processes in Geophysics* 12, 1, 117–128.
- EMSystems. 2005. EMSystems. Retrieved February 5, 2017 from <http://www.w3.org/2005/Incubator/eiif/wiki/EMSystems>.
- Günes Erkan and Dragomir R. Radev. 2004. LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research* 22, 457–479.
- Naglaa Fathy, Nagwa Badr, and Tarek Gharib. 2014. Personalized web search ranking based on different information resources. In *Signal Processing and Information Technology*. Springer, Germany, 211–214.
- FEMA. 2009. National incident management system. Retrieved February 5, 2017 from <http://www.fema.gov/about/programs/disastermanagement/>.
- Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3, 75–174.
- Alain Fournier and William T. Reeves. 1986. A simple model of ocean waves. In *ACM Siggraph Computer Graphics*, Vol. 20. ACM, New York, 75–84.
- Huiji Gao, Geoffrey Barbier, and Rebecca Goolsby. 2011. Harnessing the crowdsourcing power of social media for disaster relief. *Intelligent Systems, IEEE* 26, 3, 10–14.
- GeoVISTA. 2010. Retrieved February 5, 2017 from <http://www.geovista.psu.edu>. (2010).
- Matteo Gismondi and Otto Huisman. 2012. Spatio-temporal behaviour after an earthquake: A case study of Kawaguchi town, Japan. *Disaster Prevention and Management: An International Journal* 21, 3, 370–385.
- Michael F. Goodchild. 2009. Geographic information system. In *Encyclopedia of Database Systems*. Springer, Secaucus, NJ, 1231–1236.
- Ralph Grishman, Silja Huttunen, and Roman Yangarber. 2002. Real-time event extraction for infectious disease outbreaks. In *Proceedings of the 2nd International Conference on Human Language Technology Research*. Morgan Kaufmann Publishers Inc., ACM, CA, 366–369.

- Laura Haas. 2006. Beauty and the beast: The theory and practice of information integration. In *Database Theory (ICDT'07)*. Springer, Berlin, 28–43.
- Gyeong June Hahm, Mun Yong Yi, Jae Hyun Lee, and Hyo Won Suh. 2014. A personalized query expansion approach for engineering document retrieval. *Advanced Engineering Informatics* 28, 4, 344–359.
- Shahid Hamid, Golam Kibria, Sneha Gulati, Mark Powell, Bachir Annane, Steve Cocke, Jean-Paul Pinelli, Kurt Gurley, and Shu-Ching Chen. 2010. Predicting losses of residential structures in the state of Florida by the public hurricane loss evaluation model. *Statistical Methodology* 7, 5, 552–573.
- T. Hastie, R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*. Springer, Berlin.
- Tuukka Hastrup, Richard Cyganiak, and Uldis Bojars. 2008. Browsing linked data with Fenfire. In *Proceedings of WWW 2008 Workshop on Linked Data on the Web (LDOW2008)*. <http://www2008.org/>.
- Jefferson Heard, Sidharth Thakur, Jessica Losego, and Ken Galluppi. 2014. Big board: Teleconferencing over maps for shared situational awareness. *Computer Supported Cooperative Work* 23, 1 51–74.
- Tom Heath. 2008. *Information-seeking on the Web with Trusted Social Networks—from Theory to Systems*. Ph.D. Dissertation. The Open University, Milton Keynes, UK, 2008.
- Tom Heath and Enrico Motta. 2008. Revyu: Linking reviews and ratings into the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web* 6, 4, 266–273.
- David Heckerman. 2008. A tutorial on learning with Bayesian networks. In *Innovations in Bayesian Networks*. Springer, Secaucus, NJ, 33–82.
- Jessica Heinzelman and Carol Waters. 2010. *Crowdsourcing Crisis Information in Disaster-affected Haiti*. US Institute of Peace, Washington, DC.
- Liu Hexiang and Xu Qingjuan. 2007. Fuzzy valuation and forecast of regional flood and waterlogging risk. *Journal of Catastrophology* 4, 008.
- Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Volume 1*, 541–550.
- Aidan Hogan, Andreas Harth, Jürgen Umrich, and Stefan Decker. 2007. Towards a scalable search and query engine for the web. In *Proceedings of the 16th International Conference on World Wide Web*. ACM, 1301–1302.
- Flávio Eduardo Aoki Horita, Livia Castro Degrossi, Luiz Fernando Gomes de Assis, Alexander Zipf, and João Porto de Albuquerque. 2013. The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: A systematic literature review. In *Proceedings of the Nineteenth Americas Conference on Information Systems*. Chicago, Illinois.
- Arnold M. Howitt, Haruo Hayashi, Hiromi Akiyama, David W. Giles, and Dutch Leonard. 2010. An incident management system for Japan? *Crisis Response Journal* 9, 1, 17–19.
- Vagelis Hristidis, Shu-Ching Chen, Tao Li, Steven Luis, and Yi Deng. 2010. Survey of data management and analysis in disaster situations. *Journal of Systems and Software* 83, 10, (2010), 1701–1714.
- Wen-Ko Hsu, Wei-Ling Chiang, Qiang Xue, Dung-Mou Hung, Pei-Chun Huang, Cheng-Wu Chen, and Chung-Hung Tsai. 2013. A probabilistic approach for earthquake risk assessment based on an engineering insurance portfolio. *Natural Hazards* 65, 3, 1559–1571.
- Wen-Ko Hsu, Pei-Chung Huang, Ching-Cheng Chang, Cheng-Wu Chen, Dung-Moung Hung, and Wei-Ling Chiang. 2011. An integrated flood risk assessment model for property insurance industry in Taiwan. *Natural Hazards* 58, 3, 1295–1309.
- Wen-Ko Hsu, Dung-Mou Hung, W. L. Chiang, C. P. Tseng, and C. H. Tsai. 2006. Catastrophe risk modeling and application-risk assessment for Taiwan residential earthquake insurance pool. In *The 17th IASTED International Conference on Modelling and Simulation*. ACTA, Quebec, Canada, 338–342.
- Tao Hu, Xinyan Zhu, Kehua Su, and Lian Duan. 2011. Collection and remote management of disaster information with mobile positioning terminals. In *19th International Conference on Geoinformatics*. IEEE, Shanghai, China, 1–5.
- Muhammad Imran, Shady Elbassouni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013. Practical extraction of disaster-relevant information from social media. In *Proceedings of the 22nd International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, ACM, Switzerland, 1021–1024.
- Intermedix Corp. 2013. WebEOC. (2013). Retrieved February 5, 2017 from <https://www.intermedix.com/product/product-webeoc/index.php>.
- T. S. Jayram, Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2006. Avatar information extraction system. *IEEE Data Engineering Bulletin* 29, 1, 40–48.

- Yexi Jiang, Chang-Shing Perng, and Tao Li. 2011. Natural event summarization. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, New York, 765–774.
- Yexi Jiang, Chang-Shing Perng, and Tao Li. 2014. META: Multi-resolution framework for event summarization. In *SIAM International Conference on Data Mining*. SIAM, Pennsylvania, 605–613.
- Naim Kapucu, Tolga Arslan, and Fatih Demiroz. 2010. Collaborative emergency management and national emergency management network. *Disaster Prevention and Management: An International Journal* 19, 4, 452–468.
- Daniel A. Keim, Florian Mansmann, Jörn Schneidewind, Jim Thomas, and Hartmut Ziegler. 2008. Visual analytics: Scope and challenges. In *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Lecture Notes in Computer Science, Vol. 4404. Springer, Secaucus, NJ, 76–90.
- Khronos Group. 1997. OpenGL. Retrieved February 5, 2017 from <https://www.opengl.org/>.
- Hyeong-Seog Kim, Chang-Hoi Ho, Joo-Hong Kim, and Pao-Shin Chu. 2012. Track-pattern-based model for seasonal prediction of tropical cyclone activity in the Western North Pacific. *Journal of Climate* 25, 13, 4660–4678.
- Hyeong-Seog Kim, Joo-Hong Kim, Chang-Hoi Ho, and Pao-Shin Chu. 2011. Pattern classification of typhoon tracks using the fuzzy c-means clustering method. *Journal of Climate* 24, 2, 488–508.
- Jin Ki Kim, Raj Sharman, H. Raghav Rao, and Shambhu Upadhyaya. 2007. Efficiency of critical incident management systems: Instrument development and validation. *Decision Support Systems* 44, 1, 235–250.
- Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* 42, 8, 30–37.
- Marcin Kulawiak and Zbigniew Lubniewski. 2014. SafeCity1A GIS-based tool profiled for supporting decision making in urban development and infrastructure protection. *Technological Forecasting and Social Change* 89, 174–187.
- Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. 2011. TweetTracker: An analysis tool for humanitarian and disaster relief. In *ICWSM*. Springer, Secaucus, NJ, 661–662.
- Juan Li, Qingrui Li, Samee Ullah Khan, and Nasir Ghani. 2011. Community-based cloud for emergency management. In *6th International Conference on System of Systems Engineering (SoSE'11)*. IEEE, 55–60.
- Juan Li, Qingrui Li, Chao Liu, Samee Ullah Khan, and Nasir Ghani. 2014. Community-based collaborative information system for emergency management. *Computers & Operations Research* 42, 116–124.
- Linna Li and Michael F. Goodchild. 2010. The role of social networks in emergency management: A research agenda. *International Journal of Information Systems for Crisis Response and Management* 2, 4, 48–58.
- Lei Li and Tao Li. 2014. An empirical study of ontology-based multi-document summarization in disaster management. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44, 2, 162–171.
- Lei Li, Wei Peng, Saurabh Kataria, Tong Sun, and Tao Li. 2013. FRec: A novel framework of recommending users and communities in social media. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*. ACM, 1765–1770.
- Lianfa Li, Jinfeng Wang, Hareton Leung, and Chengsheng Jiang. 2010. Assessment of catastrophic risk using Bayesian network constructed from domain knowledge and spatial data. *Risk Analysis* 30, 7, 1157–1175.
- Tao Li, Wubai Zhou, Chunqiu Zeng, Qing Wang, Qifeng Zhou, Dingding Wang, Jia Xu, Yue Huang, Wentao Wang, Minjing Zhang, Steven Luis, Shu-Ching Chen, and Naphtali Rische. 2016. DI-DAP: An efficient disaster information delivery and analysis platform in disaster management. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, (CIKM'16), Indianapolis, IN, October 24-28, 2016*. 1593–1602.
- Wenwen Li, Michael F. Goodchild, Richard L. Church, and Bin Zhou. 2012. Geospatial data mining on the web: Discovering locations of emergency service facilities. In *Advanced Data Mining and Applications*. Springer, Berlin, 552–563.
- Yudan Li, Kasturi Chatterjee, Shu-Ching Chen, and Keqi Zhang. 2009. A 3-D traffic animation system with storm surge response. In *11th IEEE International Symposium on Multimedia (ISM'09)*. IEEE, New Brunswick, NJ, 257–262.
- D. Liben-Nowell and J. Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58, 7, 1019–1031.
- Chen Lin, Chun Lin, Jingxuan Li, Dingding Wang, Yang Chen, and Tao Li. 2012. Generating event storylines from microblogs. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 175–184.

- Chaofeng Liu and Xiangjun Zuo. 2015. A study on dynamic evaluation of urban integrated natural disaster risk based on vague set and information axiom. *Natural Hazards* 78, 1501–1506.
- M. Liu, S. M. Lo, B. Q. Hu, and C. M. Zhao. 2009b. On the use of fuzzy synthetic evaluation and optimal classification for computing fire risk ranking of buildings. *Neural Computing and Applications* 18, 6, (2009), 643–652.
- Yong Liu, David Hill, Luigi Marini, Rob Kooper, Alejandro Rodriguez, and Jim Myers. 2009a. Web 2.0 geospatial visual analytics for improved urban flooding situational awareness and assessment. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, Seattle, Washington, 554–555.
- Wei-ping Lou, Hai-yan Chen, Xin-fa Qiu, Qi-yi Tang, and Feng Zheng. 2012. Assessment of economic losses from tropical cyclone disasters based on PCA-BP. *Natural Hazards* 60, 3, 819–829.
- Alan M. MacEachren, Anthony C. Robinson, Anuj Jaiswal, Scott Pezanowski, Alexander Savelyev, Justine Blanford, and Prasenjit Mitra. 2011. Geo-Twitter analytics: Applications in crisis management. In *25th International Cartographic Conference (ICC'11)*, Paris, France, 3–8.
- Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. 2011. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver, BC, Canada, 227–236.
- Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: Can we trust what we RT?. In *Proceedings of the First Workshop on Social Media Analytics*. ACM, NY, 71–79.
- Mai Miyabe, Asako Miura, and Eiji Aramaki. 2012. Use trend analysis of Twitter after the great East Japan earthquake. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*. ACM, New York, NY, 175–178.
- Junta Mizuno and Kentaro Inui. 2013. NICT Disaster information analysis system. In *6th International Joint Conference on Natural Language Processing (IJCNLP'13)*, Nagoya, Japan, 29.
- MUC DATA SET 2001. MUC Data Sets. Retrieved February 5, 2017 from http://www-nlpir.nist.gov/related_projects/muc/muc_data/muc_data_index.html. (2001).
- Sriram Natarajan and Aura Ganz. 2009. Distributed visual analytics for collaborative emergency response management. In *Engineering in Medicine and Biology Society (EMBC'09). Annual International Conference of the IEEE*, 1714–1717.
- NC4. 2011. E-Teams. Retrieved February 5, 2017 from <http://www.nc4.us/ETeam.php>.
- A. Negarestani, S. Setayeshi, M. Ghannadi-Maragheh, and B. Akashe. 2002. Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction. *Journal of Environmental Radioactivity* 62, 3, 225–233.
- ABC News. 2011. US breaks with Japan over power plant warnings. <http://www.washingtonpost.com/wp-dyn/content/article/2011/03/16/AR2011031602816.html>.
- Kyoto News. 2012. Grandparents stifle grief to raise orphaned boy. <http://www.japantimes.co.jp/news/2012/02/23/national/grandparents-stifle-grief-to-raise-orphaned-boy/>.
- Jens Ortmann, Minu Limbu, Dong Wang, and Tomi Kauppinen. 2011. Crowdsourcing linked open data for disaster management. In *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web in Conjunction with the ISWC*. 11–22.
- M. Tamer Özsu and Patrick Valduriez. 2011. *Principles of Distributed Database Systems*. Springer Science & Business Media, New York, NY.
- Woojin Paik, Elizabeth D. Liddy, Jennifer Heverin Liddy, Ian Harcourt Niles, Eileen E. Allen, and others. 2000. Information extraction system and method using concept relation concept (CRC) triples. (June 13 2000). US Patent 6,076,088.
- Leysia Palen, Kenneth M. Anderson, Gloria Mark, James Martin, Douglas Sicker, Martha Palmer, and Dirk Grunwald. 2010. A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*.
- Leysia Palen, Starr Roxanne Hiltz, and Sophia B. Liu. 2007a. Online forums supporting grassroots participation in emergency preparedness and response. *Communications of the ACM* 50, 3, 54–58.
- Leysia Palen and Sophia B. Liu. 2007. Citizen communications in crisis: Anticipating a future of ICT-supported public participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 727–736.
- Leysia Palen, Sarah Vieweg, Sophia B. Liu, and Amanda Lee Hughes. 2009. Crisis in a networked world features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review* 27, 4, 467–480.

- Leysia Palen, Sarah Vieweg, Jeannette Sutton, Sophia B. Liu, and Amanda L. Hughes. 2007b. Crisis informatics: Studying crisis in a networked world. In *Proceedings of the 3rd International Conference on E-Social Science*. International Conference on e-Social Science, Michigan.
- Ashif Panakkat and Hojjat Adeli. 2007. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *International Journal of Neural Systems* 17, 01, 13–33.
- Ashif Panakkat and Hojjat Adeli. 2009. Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. *Computer-Aided Civil and Infrastructure Engineering* 24, 4, 280–292.
- Luca Paolino, Marco Romano, Monica Sebillio, and Giuliana Vitiello. 2010. Supporting the on-site emergency management through a visualisation technique for mobile devices. *Journal of Location Based Services* 4, 3-4, 222–239.
- Heiko Paulheim, Sebastian Döweling, Karen H. L. Tso-Sutter, Florian Probst, and Thomas Ziegert. 2009. Improving usability of integrated emergency response systems: The SoKNOS approach. In *GI Jahrestagung*. Citeseer, Germany, 1435–1349.
- Jairo Pava, Fausto Fleites, Fang Ruan, Kasturi Chatterjee, Shu-Ching Chen, and Keqi Zhang. 2010a. A three-dimensional geographic and storm surge data integration system for evacuation planning. In *IEEE International Conference on Information Reuse and Integration (IRP'10)*. IEEE, New Brunswick, NJ, 181–188.
- Yi Peng, Yong Zhang, Yu Tang, and Shiming Li. 2011. An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems* 51, 2, 316–327.
- M. Pennacchiotti and S. Gurusurthy. 2011. Investigating topic models for social media user recommendation. In *Proceedings of WWW*. ACM, NY, 101–102.
- Lamovec Peter, Mikoš Matjaž, and Oštir Kristof. 2013. Detection of flooded areas using machine learning techniques: Case study of the Ljubljana moor floods in 2010. *Disaster Advances* 6, 7, 4–11.
- Carol Peters, Martin Braschler, and Paul Clough. 2012. *Multilingual Information Retrieval: From Research to Practice*. Springer Science & Business Media, Secaucus, NJ.
- A. Popescul and L. H. Ungar. 2003. Statistical relational learning for link prediction. In *IJCAI Workshop on Learning Statistical Models from Relational Data*, Vol. 2003. Citeseer, Mexico.
- M. A. Porter, J. P. Onnela, and P. J. Mucha. 2009. Communities in networks. *Notices of the AMS* 56, 9, 1082–1097.
- Yan Qu, Philip Fei Wu, and Xiaoqing Wang. 2009. Online community response to major disaster: A study of Tianya forum in the 2008 Sichuan earthquake. In *42nd Hawaii International Conference on System Sciences (HICSS'09)*. IEEE, Hawaii, 1–11.
- Enrico L. Quarantelli. 1997. Ten criteria for evaluating the management of community disasters. *Disasters* 21, 1, 39–56.
- Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization*. Association for Computational Linguistics, ACM, Pennsylvania, 21–30.
- Stephane Roche, Eliane Propeck-Zimmermann, and Boris Mericskay. 2013. GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal* 78, 1, 21–40.
- M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers. 2010. Learning author-topic models from text corpora. *ACM Transactions on Information Systems* 28, 1, 4.
- M. Sachan, D. Contractor, T. A. Faruque, and L. V. Subramaniam. 2012. Using content and interactions for discovering communities in social networks. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, Lyon, France, 331–340.
- Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, New York, 851–860.
- Takeshi Sakaki, Fujio Toriumi, and Yutaka Matsuo. 2011. Tweet trend analysis in an emergency situation. In *Proceedings of the Special Workshop on Internet and Disasters*. ACM, NY, 3.
- Akira Sakuraba, Satoshi Nodas, Tomoyuki Ishida, Yasuo Ebara, and Yoshitaka Shibata. 2013. Tiled display environment to realize GIS based disaster information system. In *16th International Conference on Network-Based Information Systems (NBIS'13)*. IEEE, Gwangju, South Korea, 311–316.
- Khalid Saleem, Steven Luis, Yi Deng, Shu-Ching Chen, Vagelis Hristidis, and Tao Li. 2008. Towards a business continuity information network for rapid disaster recovery. In *Proceedings of the 2008 International Conference on Digital Government Research*. Digital Government Society of North America, ACM, 107–116.

- El Hamali Samiha, Nadia Nouali-TAboudjinent, and Omar Nouali. 2011. Knowledge extraction by Internet monitoring to enhance crisis management. In *Proceedings of the 8th International ISCRAM Conference-Lisbon*, Vol. 1. CERIST, Lisbon, Portugal, 1–5.
- Sunita Sarawagi. 2008. Information extraction. *Foundations and Trends in Databases* 1, 3, 261–377.
- Venu Satuluri and Srinivasan Parthasarathy. 2009. Scalable graph clustering using stochastic flows: Applications to community discovery. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, 737–746.
- Ted Schmitt, Jon Eisenberg, Ramesh R. Rao, and others. 2007. *Improving Disaster Management: The Role of IT in Mitigation, Preparedness, Response, and Recovery*. National Academies Press, Washington, D.C.
- Jean Scholtz. 2006. Beyond usability: Evaluation aspects of visual analytic environments. In *IEEE Symposium On Visual Analytics Science and Technology*. IEEE, Maryland, 145–150.
- George A. F. Seber and Alan J. Lee. 2012. *Linear Regression Analysis*. Vol. 936. John Wiley & Sons, Hoboken, NJ.
- Satoshi Sekine and Chikashi Nobata. 2004. Definition, dictionaries and tagger for extended named entity hierarchy. In *LREC*. European Language Resources Association, Lisbon, Portugal, 1977–1980.
- Habib Shah and Rozaida Ghazali. 2011. Prediction of earthquake magnitude by an improved ABC-MLP. In *Developments in E-systems Engineering (DeSE'11)*. IEEE, New Brunswick, NJ, 312–317.
- Chao Shen and Tao Li. 2010. Multi-document summarization via the minimum dominating set. In *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, ACM, Pennsylvania, 984–992.
- Irina Shklovski, Moira Burke, Sara Kiesler, and Robert Kraut. 2010. Technology adoption and use in the aftermath of hurricane Katrina in New Orleans. *American Behavioral Scientist* 53, 8, 1228–1246.
- Irina Shklovski, Leysia Palen, and Jeannette Sutton. 2008. Finding community through information and communication technology in disaster response. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. ACM, New York, 127–136.
- Alessio Signorini, Alberto Maria Segre, and Philip M. Polgreen. 2011. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS One* 6, 5 e19467.
- Iftikhar U. Sikder and Toshinori Munakata. 2009. Application of rough set and decision tree for characterization of premonitory factors of low seismic activity. *Expert Systems with Applications* 36, 1, 102–110.
- Thushari Silva, Vilas Wuwongse, and Hitesh Nidhi Sharma. 2013. Disaster mitigation and preparedness using linked open data. *Journal of Ambient Intelligence and Humanized Computing* 4, 5, 591–602.
- Peter A. Singh, Na Zhao, S.-C. Chen, and Keqi Zhang. 2005. Tree animation for a 3D interactive visualization system for hurricane impacts. In *IEEE International Conference on Multimedia and Expo (ICME'05)*. IEEE, New Brunswick, NJ, 598–601.
- Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Teerayut Horanont, Satoshi Ueyama, and Ryosuke Shibasaki. 2013. Intelligent system for human behavior analysis and reasoning following large-scale disasters. *IEEE Intelligent Systems* 28, 4, 35–42.
- Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, and Ryosuke Shibasaki. 2014a. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 5–14.
- Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, and Ryosuke Shibasaki. 2014b. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 5–14.
- David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Susan Dumais, and Bodo Billerbeck. 2012. Probabilistic models for personalizing web search. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, NY, 433–442.
- Anthony Stefanidis, Andrew Crooks, and Jacek Radzikowski. 2013. Harvesting ambient geospatial information from social media feeds. *GeoJournal* 78, 2, 319–338.
- Yu Su, Sudheer Chelluboina, Michael Hahsler, and Margaret H. Dunham. 2010. A new data mining model for hurricane intensity prediction. In *IEEE International Conference on Data Mining Workshops (ICDMW'10)*. IEEE, New Brunswick, NJ, 98–105.
- Sabiha Tahrat, Eric Kergosien, Sandra Bringay, Mathieu Roche, and Maguelonne Teisseire. 2013. Text2Geo: From textual data to geospatial information. In *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics*. ACM, New York, NY, 23.
- J. Tang, R. Yang, and M. Kafatos. 2005. Data mining for tropical cyclone intensity prediction. In *6th Conference on Coastal Atmospheric and Oceanic Prediction and Processes*. Session 7, Tropical Cyclones.
- T. Terpstra, R. J. P. Stronkman, A. de Vries, and G. L. Paradies. 2012. Towards a realtime Twitter analysis during crises for operational crisis management. In *9th International Conference on Information Systems*

- for *Crisis Response and Management (ISCRAM'12)*. Simon Fraser University, Vancouver, BC, Canada, 171–180.
- Dennis Thom, Harald Bosch, Steffen Koch, Michael Worner, and Thomas Ertl. 2012. Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages. In *IEEE Pacific Visualization Symposium (PacificVis'12)*. IEEE, Songdo, Korea, 41–48.
- James J. Thomas and Kristin A. Cook. 2006. A visual analytics agenda. *IEEE Computer Graphics and Applications* 26, 1, 10–13.
- Brian Tomaszewski. 2008. Producing geo-historical context from implicit sources: A geovisual analytics approach. *The Cartographic Journal* 45, 3, 165–181.
- Brian Tomaszewski and Alan M. MacEachren. 2010. Geo-historical context support for information foraging and sensemaking: Conceptual model, implementation, and assessment. In *IEEE Symposium on Visual Analytics Science and Technology (VAST'10)*. IEEE, Utah, 139–146.
- Chung-Hung Tsai. 2010. Earthquake risk assessment information for tourism industry. In *3rd International Conference on Information Sciences and Interaction Sciences (ICIS'10)*. IEEE, Chengdu, China, 485–490.
- Chung-Hung Tsai and Cheng-Wu Chen. 2010. An earthquake disaster management mechanism based on risk assessment information for the tourism industry—A case study from the island of Taiwan. *Tourism Management* 31, 4, 470–481.
- UCI. 2015. Rescue Disaster Portal. Retrieved February 5, 2017 from <http://rescue-ibm.calit2.uci.edu/DisasterPortalProject/disasterportal.html>. (2015).
- Suha Ülgen. 2005. Public participation geographic information sharing systems for community based urban disaster mitigation. In *Geo-information for Disaster Management*. Springer, Berlin, 1427–1434.
- Ushahidi. 2012. Retrieved February 5, 2017 from <http://www.ushahidi.com/>. <http://www.ushahidi.com/>. (2012).
- István Varga, Motoki Sano, Kentaro Torisawa, Chikara Hashimoto, Kiyonori Ohtake, Takao Kawai, Jong-Hoon Oh, and Stijn De Saeger. 2013. Aid is out there: Looking for help from tweets during a large scale disaster. In *ACL (1)*. ACL, Sofia, Bulgaria, 1619–1629.
- Sudha Verma, Sarah Vieweg, William J. Corvey, Leysia Palen, James H. Martin, Martha Palmer, Aaron Schram, and Kenneth Mark Anderson. 2011. Natural language processing to the rescue?: Extracting “situational awareness” tweets during mass emergency. In *ICWSM*. Citeseer, New Jersey.
- Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1079–1088.
- Sarah Vieweg, Leysia Palen, Sophia B. Liu, Amanda L. Hughes, and Jeannette Sutton. 2008. Collective intelligence in disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings. In *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM'08)*. ISCRAM, Washington DC.
- Dingding Wang, Tao Li, and Mitsunori Ogihara. 2012. Generating pictorial storylines via minimum-weight connected dominating set approximation in multi-view graphs. In *AAAI*. Citeseer, New Jersey, 683–689.
- Dingding Wang, Tao Li, Shenghuo Zhu, and Chris Ding. 2008. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 307–314.
- Hongning Wang, Xiaodong He, Ming-Wei Chang, Yang Song, Ryen W. White, and Wei Chu. 2013. Personalized ranking model adaptation for web search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 323–332.
- J. Waser, A. Konev, B. Sadransky, Z. Horváth, H. Ribičić, R. Carnecky, P. Kluding, and B. Schindler. 2014. Many plans: Multidimensional ensembles for visual decision support in flood management. *Computer Graphics Forum* 33, 3, 281–290.
- Connie White. 2010. Social media and meta-networks for crisis mapping: Collaboratively building spatial data for situation awareness in disaster response and recovery management. In *Specialist Meeting Spatio-Temporal Constraints on Social Networks*. The National Center for Geographic Information and Analysis, California, 14–15.
- Connie White, Linda Plotnick, Jane Kushma, and Starr Roxanne Hiltz. 2009. An online social network for emergency management. *International Journal of Emergency Management* 6, 3, 369–382.
- Jeremy J. D. White and Robert E. Roth. 2010. TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information. In *Proceedings of GIScience*, Vol. 2010. Springer, Zurich, Switzerland.

- Michael White, Tanya Korelsky, Claire Cardie, Vincent Ng, David Pierce, and Kiri Wagstaff. 2001. Multidocument summarization via information extraction. In *Proceedings of the 1st International Conference on Human Language Technology Research*. Association for Computational Linguistics, ACM, PA, 1–7.
- Wikipedia. 2011. Tohoku Earthquake. Retrieved February 5, 2017 from http://en.wikipedia.org/wiki/2011_Tohoku_earthquake_and_tsunami.
- Zhengzheng Xing, Jian Pei, Guozhu Dong, and S. Yu Philip. 2008. Mining sequence classifiers for early prediction. In *SDM*. SIAM, Pennsylvania, 644–655.
- Hua Xu, Shane P. Stenner, Son Doan, Kevin B. Johnson, Lemuel R. Waitman, and Joshua C. Denny. 2010a. MedEx: A medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association* 17, 1, 19–24.
- Kaikuo Xu, Changjie Tang, Chuan Li, Yexi Jiang, and Rong Tang. 2010b. An MDL approach to efficiently discover communities in bipartite network. In *Database Systems for Advanced Applications*. Springer, Secaucus, NJ, 595–611.
- T. Yang, R. Jin, Y. Chi, and S. Zhu. 2009. Combining link and content for community detection: A discriminative approach. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 927–936.
- Xiaoling Yang, Jianzhong Zhou, Jiehua Ding, Weiping Deng, and Yongchuan Zhang. 2009. Study on evaluation methods of flood disaster grade attribute recognition analysis and BP neural network. In *6th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'09)*. Vol. 4. 386–390.
- Yimin Yang, Wenting Lu, Jesse Domack, Tao Li, Shu-Ching Chen, Steven Luis, and Jainendra K. Navlakha. 2012. MADIS: A multimedia-aided disaster information integration system for emergency management. In *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'12)*. 233–241.
- Jie Yin, Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power. 2012. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems* 27, 6, 52–59.
- Liyang Yu. 2011. Linked open data. In *A Developer's Guide to the Semantic Web*. Springer, 409–466.
- Zhiyong Yu, Daqing Zhang, Dingqi Yang, and Guolong Chen. 2012. Selecting the best solvers: Toward community based crowdsourcing for disaster management. In *IEEE Asia-Pacific Services Computing Conference (APSCC'12)*. IEEE, Guilin, China, 271–277.
- Chunqiu Zeng, Yexi Jiang, Li Zheng, Jingxuan Li, Lei Li, Hongtai Li, Chao Shen, Wubai Zhou, Tao Li, Bing Duan, and others. 2013. Fiu-miner: A fast, integrated, and user-friendly system for data mining in distributed environment. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, 1506–1509.
- Chunqiu Zeng, Hongtai Li, Huibo Wang, Yudong Guang, Chang Liu, Tao Li, Mingjin Zhang, Shu-Ching Chen, and Naphtali Rishe. 2014. Optimizing online spatial data analysis with sequential query patterns. In *Proceedings of the 15th IEEE International Conference on Information Reuse and Integration (IRI'14)*. 253–260.
- H. Zhang, B. Qiu, C. L. Giles, H. C. Foley, and J. Yen. 2007. An LDA-based community structure discovery approach for large-scale social networks. In *Proceedings of IEEE International Conference on Intelligence and Security Informatics*. IEEE, New Brunswick, NJ, 200–207.
- Keqi Zhang, S.-C. Chen, Peter Singh, Khalid Saleem, and Na Zhao. 2006. A 3d visualization system for hurricane storm-surge flooding. *Computer Graphics and Applications, IEEE* 26, 1 (2006), 18–25.
- Keqi Zhang, Chengyou Xiao, and Jian Shen. 2008. Comparison of the CEST and SLOSH models for storm surge flooding. *Journal of Coastal Research* 24, 2, 489–499.
- Keqi Zhang, Jianhua Yan, and Shu-Ching Chen. 2006. Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing* 44, 9, 2523–2533.
- Mingjin Zhang, Huibo Wang, Yun Lu, Tao Li, Yudong Guang, Chang Liu, Erik Edrosa, Hongtai Li, and Naphtali Rishe. 2015. TerraFly GeoCloud: An online spatial data analysis and visualization system. *ACM Transactions on Intelligent Systems and Technology* 6, 3, 34.
- Li Zheng, Chao Shen, Liang Tang, Tao Li, Steve Luis, and Shu-Ching Chen. 2011. Applying data mining techniques to address disaster information management challenges on mobile devices. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, 283–291.
- Li Zheng, Chao Shen, Liang Tang, Tao Li, Steve Luis, Shu-Ching Chen, and Vagelis Hristidis. 2010. Using data mining techniques to address critical information exchange needs in disaster affected public-private networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 125–134.

- Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li, Steven Luis, and Shu-Ching Chen. 2013. Data mining meets the needs of disaster information management. *IEEE Transactions on Human-Machine Systems*, 43, 5, 451–464.
- Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li, Steven Luis, Shu-Ching Chen, and Jainendra K. Navlakha. 2012. Disaster SitRep-A vertical search engine and information analysis tool in disaster management domain. In *IEEE 13th International Conference on Information Reuse and Integration (IRI)*. IEEE, Las Vegas, NV, 457–465.
- D. Zhou, E. Manavoglu, J. Li, C. L. Giles, and H. Zha. 2006. Probabilistic models for discovering e-communities. In *Proceedings of the 15th International Conference on World Wide Web*. ACM, Edinburgh, Scotland, 173–182.
- Wubai Zhou, Chao Shen, Tao Li, Shu-Ching Chen, and Ning Xie. 2014. Generating textual storyline to improve situation awareness in disaster management. In *IEEE 15th International Conference on Information Reuse and Integration (IRI'14)*. IEEE, California, 585–592.
- Ningning Zhu. 2003. Data versioning systems. Technical Reports (TR-131), Computer Science Department, Stony Brook University, Stony Brook, NY 11794, February 2003.
- Zipei Fan, Xuan Song, Ryosuke Shibasaki. 2014. CitySpectrum: A non-negative tensor factorization approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'14)*. ACM.
- Boris Zmazek, Ljupčo Todorovski, Sašo Džeroski, Janja Vaupotič, and Ivan Kobal. 2003. Application of decision trees to the analysis of soil radon data for earthquake prediction. *Applied Radiation and Isotopes* 58, 6, 697–706.
- Matthew Zook, Mark Graham, Taylor Shelton, and Sean Gorman. 2010. Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy* 2, 2, 7–33.

Received October 2015; revised July 2016; accepted November 2016