Processing Social Media Messages in Mass Emergency: A Survey

MUHAMMAD IMRAN and CARLOS CASTILLO, Qatar Computing Research Institute
FERNANDO DIAZ, Microsoft Research
SARAH VIEWEG, Qatar Computing Research Institute

Social media platforms provide active communication channels during mass convergence and emergency events such as disasters caused by natural hazards. As a result, first responders, decision makers, and the public can use this information to gain insight into the situation as it unfolds. In particular, many social media messages communicated during emergencies convey timely, actionable information. Processing social media messages to obtain such information, however, involves solving multiple challenges including: parsing brief and informal messages, handling information overload, and prioritizing different types of information found in messages. These challenges can be mapped to classical information processing operations such as filtering, classifying, ranking, aggregating, extracting, and summarizing. We survey the state of the art regarding computational methods to process social media messages and highlight both their contributions and shortcomings. In addition, we examine their particularities, and methodically examine a series of key subproblems ranging from the detection of events to the creation of actionable and useful summaries. Research thus far has, to a large extent, produced methods to extract situational awareness information from social media. In this survey, we cover these various approaches, and highlight their benefits and shortcomings. We conclude with research challenges that go beyond situational awareness, and begin to look at supporting decision making and coordinating emergency-response actions.

Categories and Subject Descriptors: H. [**Information Systems**]; H.4 [**Information Systems Applications**]; H.4.2 [**Types of Systems**]: Decision Support

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Social media, crisis computing, disaster management, mass emergencies

ACM Reference Format:
Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2015. Processing social media messages in mass emergency: A survey. ACM Comput. Surv. 47, 4, Article 67 (June 2015), 38 pages.
DOI: http://dx.doi.org/10.1145/2771588

1. INTRODUCTION

Crisis situations, such as disasters brought on by natural hazards, present unique challenges to those who study them, creating conditions that call for particular research methods [Stallings 2002]. In this article, we survey methods for studying disasters from the perspective of information processing and management, specifically methods for processing social media content.

Crisis situations—particularly those with little to no warning (known as "sudden-onset crises")—generate a situation that is rife with questions, uncertainties, and the need to make quick decisions, often with minimal information. When it comes to

For inquiries about this article, please contact mimran@qf.org.qa.
Authors' addresses: M. Imran, C. Castillo, and S. Vieweg, Qatar Computing Research Institute, West Bay, 5825, Doha, Qatar; emails: mimran@qf.org.qa, chato@acm.org, svieweg@qf.org.qa; F. Diaz, Microsoft Research, 641 Avenue of the Americas, Floor 7, New York NY 10011; email: diazf@acm.org.
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.
© 2015 ACM 0360-0300/2015/06-ART67 $15.00
DOI: http://dx.doi.org/10.1145/2771588

information scarcity, research in recent years has uncovered the increasingly important role of social media communications in disaster situations, and shown that information broadcast via social media can enhance situational awareness during a crisis situation [Vieweg 2012]. However, social media communications during disasters are now so abundant that it is necessary to sift through hundreds of thousands, even millions, of data points to find information that is most useful during a given event.

The goal of this survey is to provide computer science researchers and software developers with computational methods that they can use to create tools for formal response agencies, humanitarian organizations, and other end users, giving them a way to successfully identify, filter, and organize the overwhelming amount of social media data that are produced during any given crisis. Such tools can help stakeholders make time-critical—and potentially life-saving—decisions.

1.1. Social Media During Crisis Situations

Brief History. The use of Internet technologies to gather and disperse information in disaster situations, as well as to communicate among stakeholders, dates back to the late 1990s. Internet historians point to online newsgroups and email clients that were used to coordinate protests in Indonesia in 1998 [Poole et al. 2005]. In addition, there are cases of websites being set up in response to crises in 2003 [Palen and Liu 2007]. To the best of our knowledge, 2004 is the first year in which a user-generated content website was used in response to a crisis; after the Indian Ocean Tsunami of December 26 that year, an electronic bulletin board was set up and moderated for 10 days.[1] In addition, in the aftermath of Hurricane Katrina, which struck the city of New Orleans in the United States in 2005, significant emergency response activity took place on MySpace [Shklovski et al. 2010]. One of the earliest known cases of people using the microblogging service Twitter in an emergency was during severe wildfires that took place near San Diego, California (in the United States) in 2007.[2] Since then, it has become common practice for affected populations and concerned others to use Twitter to communicate, ask questions, collect and spread information, and organize response efforts (among other tasks) [Starbird et al. 2010; Vieweg et al. 2010; Sarcevic et al. 2012; Starbird 2013; Imran et al. 2014a; Cobb et al. 2014].

Today. The growing adoption of social media during disasters has created opportunities for information propagation that would not exist otherwise. Emergency response agencies routinely post information such as emergency alerts and advice through these channels,[3] but social media enables much more than "top-down" communications. People post situation-sensitive information on social media related to what they experience, witness, and/or hear from other sources [Hughes and Palen 2009]. This practice allows both affected populations and those outside the impact zone to learn about the situation *first-hand* and in near real time.

We know that information posted to social media platforms in time- and safety-critical circumstances can be of great value to those tasked with making decisions in these fraught situations. Previous research has shown that information that contributes to situational awareness is reported via Twitter (and other social media platforms) during mass emergencies [Vieweg et al. 2010; Vieweg 2012; Imran et al. 2014a]. Now, those tasked with formal response efforts—from local fire departments[4] to international aid agencies—are working to incorporate information broadcast on social

[1]http://www.thefreelibrary.com/www.p-h-u-k-e-t.com+Has+Served+Its+Purpose+After+the+Tsunami\%3B+Site...-a0126803919.
[2]Eric Frost, personal communication.
[3]See e.g. https://blog.twitter.com/2013/twitter-alerts-critical-information-when-you-need-it-most
[4]http://edition.cnn.com/2012/11/01/tech/social-media/twitter-fdny/.

media platforms into their processes and procedures. Many emergency responders and humanitarian officials recognize the value of the information posted on social media platforms by members of the public (and others), and are interested in finding ways to quickly and easily locate and organize information that is of most use to them [Hughes 2012].[5] Some agencies have even begun to formally incorporate social media monitoring and communication during mass emergency situations. The American Red Cross (ARC), in a survey, reported the effectiveness of social media and mobile apps.[6] In 2014 ARC opened their Social Media Digital Operations Center for Humanitarian Relief. The goals of the center are to "source additional information from affected areas during emergencies to better serve those who need help; spot trends and better anticipate the public's needs; and connect people with the resources they need, like food, water, shelter or even emotional support."[7] Though the ARC is currently one of the few (possibly the only) formal agencies to support such a center, it is likely that similar operations will begin within other organizations. Among other similar examples, the Australia Crisis Tracker, which is a machine-learning–based tool to filter spam and categorize data into different event types, is also deployed with the Australian Red Cross.[8]

Though formal response agencies express interest in incorporating social media into their processes, obstacles exist. For example, a recent survey by the US Congressional Research Service cites administrative cost as a significant barrier to adopting social media during emergencies: "The number of personnel required to monitor multiple social media sources, verify the accuracy of incoming information, and respond to and redirect incoming messages is also uncertain ... Responding to each message in a timely manner could be time consuming and might require an increase in the number of employees ..." [Lindsay 2011]. Others have expressed concerns including issues related to roles, responsibilities, and liabilities; difficulties evaluating the veracity, trustworthiness, and reliability of information; and information overload in general [Vieweg et al. 2014; Hughes et al. 2014a]. Computational methods can help overcome some of these obstacles, by reducing the amount of information to be examined by humans. Automatic methods are necessary when human computation is limited; in the following sections, we detail what those methods entail.

1.2. Background Reading

Sociologists began researching human behavior in mass emergency situations long before the Internet, or even modern computing. The purpose of this section is not to provide the reader with an exhaustive list of sociology of disaster literature; we highlight a few foundational readings that are helpful for the computer science, information science, technology, and social media scholars to gain quick insight into the rich and varied field of the sociology of disaster.

E. L. Quarantelli's chapter "The Disaster Research Center (DRC) Field Studies of Organized Behavior in the Crisis Time Period of Disasters" (in Stallings [2002]) provides a brief history of one of the foremost disaster research institutes in the United States. Quarantelli gives background on the Disaster Research Center, and explains the strategic as well as academically oriented decisions that were made in order to highlight the importance of studying the social science aspects of disaster.

In Mileti [1999], he and the contributing authors aim to reach a general (i.e., nonacademic) audience and provide background on disasters caused by natural hazards.

[5]Andrej Verity, personal communication.
[6]http://www.redcross.org/news/press-release/More-Americans-Using-Mobile-Apps-in-Emergencies
[7]http://www.redcross.org/news/press-release/The-American-Red-Cross-and-Dell-Launch-First-Of-Its-Kind-Social-Media-Digital-Operations-Center-for-Humanitarian-Relief
[8]http://www.research.ibm.com/articles/crisis-tracker.shtml

The volume is comprised of "synthesized statements of what is known, collectively, about hazards and human coping strategies." Mileti and colleagues point to causes of disaster, which happen when three major systems—the physical, social, and built environments—interact in complex ways. The authors' goal is to give the reader a way to understand how to study disaster situations, with a final goal of helping members of the public create more resilient communities.

When it comes to combining studies of disaster with the use of information technology, including social media, the Harvard Humanitarian Initiative [2011] presents an in-depth analysis of the response to the earthquake in Haiti in January 2010. With a stronger focus on social media, a recent survey by Hughes et al. [2014a] considers the motivating factors of emergency responders regarding their use of social media data. The authors describe the challenges they face, best practices regarding the adoption of social media by formal response organizations, and also touch on instances of integrated, end-to-end systems that are currently being built to meet these needs. In addition, an article by Palen and Liu [2007] was one of the first to provide an early assessment regarding how information and communication technology can support the participation of the public during crisis situations. Since then, many articles that focus on the role of social media in disaster have been published, but the two we mention here provide a good "first glance" to readers who are new to the field.

Our brief overview of foundational reading would not be complete without mentioning the much-discussed issue of trust and the use of social media. A recent ACM Computing Survey looks at this very topic [Sherchan et al. 2013]. The authors review the various definitions of "trust" from a variety of academic disciplines, discuss the factors that contribute to notions of trust, and combine the complex and much-scrutinized idea of trust with computing and social network research.

1.3. Scope and Organization

The overarching problem we aim to confront in this article is that of extracting time-critical information from social media that is useful for emergency responders, affected communities, and other concerned populations in disaster situations.

We note that social media analysis has been used for a number of applications in the domains of economics, politics (e.g., Hong and Nadler [2011]) and public health (e.g., Aramaki et al. [2011]); for a survey, see Weber and Mejova [2013]. Even if we consider only applications to time-critical settings, we note that often the same methods described in this survey can be applied to the analysis of social media during mass converge events, such as large political conventions, concerts, or sports events; or for monitoring the performance of a media campaign or a televised debate, among similar applications. However, while many of the methods and algorithms that we describe can be used for other purposes, we explain them from the perspective of their applications to mass emergencies, a topic that has a specific scientific and technical community, and that targets a particular set of use-cases.

The following two sections briefly describe our target end-user audience, and their information needs (Section 2), and end-to-end integrated systems (Section 3). The subsequent sections form the main technical part of this survey and present a systematic analysis of the computational methods we cover.

—Section 4 starts with a general characterization of social media messages broadcast during disasters. Next, it introduces methods for programmatically acquiring data from social media, and for preprocessing the data.
—Section 5 covers methods for the detection of new events, which involves detecting the first message on a given topic or subtopic. The section also covers how to track these events, that is, how to collect further messages belonging to the same topic.

—Section 6 outlines methods to mine and aggregate information. These methods include unsupervised classification (and clustering), supervised classification, information extraction, and summarization.
—Section 7 presents how semantic technologies can be applied in this domain. This corresponds, first, to enrich the content with semantic information, and second, to use an ontology for disaster management to describe the content of the messages.

The final section concludes the survey, and outlines current research directions.

2. USERS AND INFORMATION NEEDS

Much of the research we present here focuses on the computational aspects of processing social media messages in time- and safety-critical situations. It is also important to consider the end users of these technological solutions: those who benefit from having curated information that describes a disaster or crisis and enhances situational awareness, including formal response agencies, humanitarian organizations, and members of the public.

2.1. Public Participation in Crises

Ideally, to understand how the public participates in social media during crises, we should start by asking how the public reacts in general to crises.

Contrary to Hollywood renditions of disaster situations, human response to crises is not one of panic and mayhem [Mitchell et al. 2000]. Victims of disaster do not lose control, run amok, or flee the area in fear. Instead, they make quick decisions based on the information available to them at the time, which oftens allow them to save their own lives as well as help those around them [Mitchell et al. 2000]. Neighbors, friends, and other members of the public are the first to respond when a disaster strikes. They rush to the scene to perform search-and-rescue operations, administer first aid, and perform critical tasks necessary in the first moments of response. Often, these "first responders" are victims of the disaster themselves [Dynes 1970]. The role of the public in disaster response efforts is critical; with the growing use of social media to gather and disperse information, organize relief efforts, and communicate, members of the public who can play a valuable role in these situations are no longer limited to those in the area of impact.

As Dynes [1994] explains, emergencies do not render victims incapable of helping themselves and others, nor do they create a situation in which they are unable to make intelligent, personally meaningful decisions. What emergencies *do* create is an environment in which new and perhaps unexpected problems are presented, which members of the public are called upon to solve. Research in recent years on the use of social media in disasters shows how members of the public, formal response agencies, and other stakeholders have taken to online outlets to perform tasks such as communicating about hospital availability [Starbird 2013], coordinating medical responses [Sarcevic et al. 2012], and communicating with the public during various crises [Cobb et al. 2014], among many others. These users interact in complex ways including producing, distributing, and organizing content [Starbird et al. 2010].

2.2. Differences in Information Needs

The recognition that social media communications are a valid and useful source of information throughout the disaster lifecycle (preparation, impact, response, and recovery) is increasing among the many stakeholders who take action in disaster situations. In particular, members of the public; formal response agencies; and local, national, and international aid organizations are all aware of the ability to use social media to gather

and disperse timely information in the aftermath of disaster. The specific information they seek—and their ability to put it to use—may differ, however [Vieweg 2012].

Depending on the circumstances of the disaster, and what roles and duties the various stakeholders are responsible for, their specific information needs will vary. For example, in a wildfire situation that affects a community, members of a formal response organization such as local police or area firefighters can benefit from information such as where people are smelling smoke, what precautions they are taking (e.g., clearing brush, watering yards), and what traffic patterns look like. In a large-scale, sudden-onset disaster such as a typhoon or earthquake, humanitarian agencies, such as the various branches of the United Nations, benefit from information that details the current situation "on the ground," such as where electricity has been disabled, or where people are without food and water. In any disaster situation, members of the public play a variety of roles and take on many tasks; the information they find valuable may be very personal-that is, hearing that a friend or loved one is safe, or it may be more broadly applicable, such as the status of a certain neighborhood or town.

Overall, the information any individual, group, or organization finds useful and seeks out in a disaster will depend on their goals. Is it a group interested in providing food to children? Is it an organization that can set up a field hospital? Is it an individual living in a foreign country who is concerned about one's family? The different types of information sought by these different stakeholders may be broadcast on Twitter, but to find it quickly, users rely on technological methods to sift through the millions of tweets broadcast at any given time to find useful information. Further information and a deeper perspective on users of social media in disaster can be found in Hughes et al. [2014a] and Hughes and Palen [2012].

3. SYSTEMS FOR CRISIS-RELATED SOCIAL MEDIA MONITORING

Table I provides examples of existing systems described in the literature that extract crisis-relevant information from social media.[9] The systems we include have varying degrees of maturity; some have been deployed in real-life situations, while others are still under development.

Most existing systems are built around the concept of a *dashboard*, or a set of visual displays that provide a summary of social media during the crisis according to temporal, spatial, and thematic aspects. Common elements in these displays include:

—Lists/timelines showing recent or important messages, sometimes grouping the messages into clusters or categories
—Time series graphs representing the volume of a hashtag, word, phrase, or concept over time, and sometimes marking peaks of activity
—Maps including geotagged messages or interpolated regions, possibly layered according to different topics
—Pie charts or other visual summaries of the proportion of different messages

These visual elements are powered by computational capabilities that include:

—Collections of social media messages matching a given criterion, from one or multiple social media and/or Short Message Service (SMS) streams, typically with a focus on Twitter (described in Section 4)
—Natural Language Processing (NLP), including Named Entity Recognition (NER) and linking of named entities to concepts (described in Section 4.3)

[9]The list is not extensive, and does not include tools such as Radian6 (http://www.salesforcemarketingcloud.com/products/social-media-listening/) that have not been described in the literature, but might be relevant for other reasons-for example, in the case of Radian6, because it is used by the American Red Cross.

Table I. Example Systems Described in the Academic Literature that Extract Crisis-Relevant Information from Social Media

System name Data; example capabilities	Reference and URL
Twitris	[Sheth et al. 2010; Purohit and Sheth 2013]
Twitter; semantic enrichment, classify automatically, geotag	http://twitris.knoesis.org/
SensePlace2	[MacEachren et al. 2011]
Twitter; geotag, visualize heat-maps based on geotags	http://www.geovista.psu.edu/SensePlace2/
EMERSE: Enhanced Messaging for the Emergency Response Sector	[Caragea et al. 2011] http://emerse.ist.psu.edu/
Twitter and SMS; machine-translate, classify automatically, alerts	
ESA: Emergency Situation Awareness	[Yin et al. 2012; Power et al. 2014]
Twitter; detect bursts, classify, cluster, geotag	https://esa.csiro.au/
Twitcident	[Abel et al. 2012]
Twitter and TwitPic; semantic enrichment, classify	http://wis.ewi.tudelft.nl/twitcident/
CrisisTracker	[Rogstadius et al. 2013]
Twitter; cluster, annotate manually	https://github.com/jakobrogstadius/crisistracker
Tweedr	[Ashktorab et al. 2014]
Twitter; classify automatically, extract information, geotag	https://github.com/dssg/tweedr
AIDR: Artificial Intelligence for Disaster Response	[Imran et al. 2014a]
Twitter; annotate manually, classify automatically	http://aidr.qcri.org/

—Extraction of information from the messages, including geotagging (described in Section 4.4)
—Monitoring the volume of messages (or sets of messages) to detect or help detect subevents within a crisis, and sometimes the crisis itself (described in Section 5), possibly including the generation of user-defined alerts when certain conditions are met
—Clustering or automatic grouping of similar messages (described in Section 6.1.3)
—Classification of messages or groups of messages manually or automatically (see Section 6.1.2)
—Automatic translation of messages

Though some of these systems are based on input or feedback provided by emergency responders and other officials, we note that to a large extent they are framed as a way to process social media data during crisis situations; their goal is not to address specific needs of emergency responders or other stakeholders. This focus on *processing* social media data possibly impacts the adoption of these systems by the practitioner community. Methodologies such as participatory design have been proposed to improve the matching between the needs of public information officers during a crisis and the tools built by researchers and developers, for example [Hughes 2014].

When the goal of meeting the specific need of users *is* stated explicitly in the design of systems, it often revolves around *enhancing situational awareness*, defined in Endsley [1995] as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

For instance, *ESA* [Yin et al. 2012; Power et al. 2014] aims at enhancing situational awareness with respect to crises induced by natural hazards, particularly earthquakes. This is done by presenting information in time (frequency series) and space (maps), which is achieved by performing event detection, text classification, online clustering, and geotagging. Similarly, *SensePlace2* [MacEachren et al. 2011] is presented as a *geovisual analytics* system, which filters and extracts geographical, temporal, and thematic

information from tweets in order to present them in a layered map. Data from social media can also be presented along with data from physical sensors, for instance, to overlay earthquakes detected by seismic sensors on a map presenting social media data [Avvenuti et al. 2014; Musaev et al. 2014].

In parallel to approaches that use NLP techniques to enhance situation awareness, *Crisis Mapping* emerged as an alternative type of system by employing digital volunteers to collect, classify, and geotag messages [Okolloh 2009; Meier 2015], and eventually by using the input from those volunteers to train machines to perform these tasks automatically [Imran et al. 2014a].

4. DATA CHARACTERIZATION, ACQUISITION, AND PREPARATION

Both academics and practitioners gather social media data during crisis events. In this section, we describe the common practices used to collect, represent, and process these data.

4.1. Characteristics of Messages Broadcast on Social Media in Disaster

Social media is a general term that encompasses a variety of platforms on which user-generated content can be disseminated and consumed, and where users can connect with others. This definition currently includes blogging and microblogging, social networking sites, social media sharing platforms, and *wikis* [van Dijck 2013].

Activities such as staying in touch with friends and family, and connecting with others, have driven the growth of social media platforms.[10] Currently, different social networking sites are used for different purposes, but commonalities do exist. For instance, the top 3 activities on Twitter are to (1) post about daily activities, (2) upload and share photos, and (3) comment on posts of others; on Facebook they are to (1) upload and share photos, (2) message with friends on a one-on-one basis, and (3) comment on posts of friends [GlobalWebIndex 2013].

In any of these platforms, an increase in social media communications can be triggered by a variety of causes, which can be divided into *endogenous* and *exogenous* [Crane and Sornette 2008]. Endogenous causes refer to phenomena in which an idea or "meme" gains popularity by a process of *viral contagion* or *information cascade*, in which content spreads rapidly through a network, potentially reaching a significant fraction of all the users [Chen et al. 2013].

Exogenous causes refer to large-scale events, usually happening in the physical world, of wide interest to social media users. Emergencies and mass convergence events are examples of an exogenous cause; during such events, we know that communications activity increases. For instance, it has been observed that mobile network usage—both in terms of phone calls and SMS—increases in emergency situations [Gao et al. 2014]. The same is true for social media usage, which "rises during disasters as people seek immediate and in-depth information" [Fraustino et al. 2012].[11]

To illustrate the types of information that affected populations broadcast specifically on the popular microblogging platform Twitter, we turn to some example messages that have been highlighted in previous literature:

—"OMG! The fire seems out of control: It's running down the hills!" (bush fire near Marseilles, France, in 2009, quoted from Twitter in De Longueville et al. [2009])

[10]http://www.pewinternet.org/2011/11/15/why-americans-use-social-media/.

[11]For instance, activity on Facebook was observed to increase significantly in the areas most affected by the August 2014 earthquake in California (https://www.facebook.com/notes/facebook-data-science/on-facebook-when-the-earth-shakes/10152488877538859).

—"Red River at East Grand Forks is 48.70 feet, +20.7 feet of flood stage, −5.65 feet of 1997 crest. #flood09" (automatically-generated tweet during Red River Valley floods in 2009, quoted from Twitter in Starbird et al. [2010])
—"Anyone know of volunteer opportunities for hurricane Sandy? Would like to try and help in any way possible" (Hurricane Sandy 2013, quoted from Twitter in Purohit et al. [2013])
—"My moms backyard in Hatteras. That dock is usually about 3 feet above water [photo]" (Hurricane Sandy 2013, quoted from Reddit in Leavitt and Clark [2014])
—"Sirens going off now!! Take cover ... be safe!" (Moore Tornado 2013, quoted from Twitter in Blanford et al. [2014])

Though these are only a few examples, they convey a sense of the types of information posted during an event, and show that it is varied. Vieweg [2012] points to this variation in her research that is based on a detailed study of four crisis events, in which she identifies in Twitter 35 types of information in three broad categories defined by Mileti [1999]: social environment, built environment, and physical environment. She points out that social environment messages describe anything having to do with people and their reactions to the crisis, built environment messages correspond to information and updates about property and infrastructure, and physical environment messages include updates about the hazard agent, weather, and other environmental factors (see Section 6.1.1 for details on different ways of categorizing this information).

Quantifying the amount of information found in social media based on type is even more difficult than locating that information in the first place. Important variations have been observed across crises (even for similar events) and at different stages of a crisis [Blanford et al. 2014; Imran et al. 2013a]. Olteanu et al. [2014] looked at the prevalence of three broad categories of information in tweets related to six crisis events. The results show large variabilities in the number of tweets reporting the negative consequences of an event (20%–60%), those offering or asking for donations (15%–70%) and those warning about risks or providing advice (5%–20%).

Using tweets as an indication of spatial zones during a disaster is also possible. For instance, Acar and Muraki [2011] examine the use of Twitter during an earthquake in Japan and observed that tweets from affected areas include more requests for help and more warnings, while tweets from other areas which are far from the disaster epicenter tend to mostly include other types of information, such as concern and condolences.

4.2. Data Acquisition

Most large social media platforms provide programmatic access to their content through an Application Programming Interface (API). However, the details of these APIs vary substantially from one platform to another, and change over time.

APIs to access social media data typically belong to one of two types: those allowing queries of an archive of past messages (also known as *search* APIs) and those allowing data collectors to subscribe to a real-time data feed (also known as *streaming* or *filtering* APIs). Both types of APIs typically allow data collectors to express an information need, including one or several of the following constraints: (i) a time period; (ii) a geographical region for messages that have GPS coordinates (which are currently the minority); or (iii) a set of keywords that must be present in the messages, which requires the use of a query language whose expressiveness varies across platforms. In the case of archive/search APIs, messages are returned sorted by relevance (a combination of several factors, including recency) or just by recency. In the case of real-time/streaming/filtering APIs, messages are returned in order of their posting time.

Data collection strategies impact the data obtained and analytic results. For instance, selecting messages in the geographical region affected by a disaster versus selecting messages based on a keyword-based query may return datasets having different characteristics [Olteanu et al. 2014].

Data availability. In addition to issues regarding network connectivity during disasters [Jennex 2012], access to social media data is in general limited, which is a serious obstacle to research and development in this space [Reuter and Scholl 2014].

First, historical, archived social media can typically be queried only through *search APIs*, which are limited in terms of number of queries or data requested per unit of time. The exceptions are (i) a 1% data sample collected from Twitter by the Internet Archive,[12] and (ii) collections created by researchers using automatic data collection methods (e.g., web crawling), which are against the terms of services of most social media platforms.

Second, access to real-time data is quite limited. Twitter offers a public streaming API providing a random sample of 1% of all postings, plus the possibility of filtering all public postings by keywords.[13] This is in contrast with most large social media platforms, which do not offer this level of data access publicly.[14]

The consequence of these limitations is that most work on crisis-related messages is done using Twitter data (with few exceptions, e.g., Facebook [Bird et al. 2012] and Reddit [Leavitt and Clark 2014]), which provides an incomplete view of a crisis situation, as there are many different online social media sites that might be used for different purposes. Limitations in the amount of data that can be collected and in general a dependency on a small set of data providers or APIs, further reduces the efficiency and effectiveness of tools for handling crisis-related social media messages.

4.3. Data Preprocessing

Most researchers and practitioners prepare social media data by *preprocessing* it, using some of the methods outlined to follow, before performing the actual analysis. Many preprocessing techniques are available, and the choice depends on the type of data at hand and the goals of the analysis.

Natural Language Processing (NLP). The text of the messages can be preprocessed by using an NLP toolkit. Typical operations include tokenization, part-of-speech tagging (POS), semantic role labeling, dependency parsing, named entity recognition, and entity linking. A number of off-the-shelf implementations of these operations are available online, for example, the Stanford NLP[15] or NLTK for Python.[16] Social media–specific NLP toolkits can also be used. For instance, ArkNLP [Owoputi et al. 2013], which is trained on Twitter data, is able to recognize Internet idioms such as "ikr" (*I know, right?*) and assign them the correct POS tag (interjection, in this case). Additionally, higher-level operations can be applied, including applying sentiment analysis methods to infer aspects of the emotion conveyed by a piece of text [Pang and Lee 2008].

Feature extraction. For many automatic information-processing algorithms (e.g., machine learning), each data item must be represented accurately as an information record. The representation of choice for text is typically a numerical vector in which each position corresponds to a word or phrase—this is known as the vector space model

[12]Available at https://archive.org/details/twitterstream.
[13]Available at https://dev.twitter.com/docs/api/streaming.
[14]As of November 2014, we found no publicly-available equivalent of Twitter's *streaming API* in Sina Weibo, Facebook, YouTube, Google Plus, or Tumblr.
[15]http://www-nlp.stanford.edu/software/.
[16]http://www.nltk.org/.

in information retrieval. The value in each position can be a binary variable, indicating the presence or absence of that word or phrase in the message, or a number following some weighting scheme, such as TF-IDF, which favors terms that are infrequent in the collection [Baeza-Yates and Ribeiro-Neto 2011]. To avoid having too many variables, textual features can be discarded by removing stopwords and functional words, or by normalizing words using stemming or lemmatization (e.g., considering "damaging" and "damage" as equivalent), or other means.

Additionally, other text-based features can be added, such as the length of the text in words or characters, and the number of questions or exclamation marks. If some NLP preprocessing is performed on the text—such as POS tagging—a feature such as *noun:fire* (instead of *verb:fire*) can be used to distinguish that the word "fire" is being used in a message as a noun ("I heard a fire alarm") instead of a verb ("They should fire him").

In the case of tweets, characteristics such as the presence of user mentions ("@user"), URLs, or hashtags can be included as features. In the case of images or video, content-based features such as colors, textures, and shapes can be included (see, e.g., Liu et al. [2007] for a survey). Additionally, features such as the date of a message, tags associated with it, the number of views/comments it has received, or information about its author are often available in a platform-dependent manner.

Obviously, one can spend a great deal of time constructing features by hand. In order to guide this exploration, both researchers and practitioners should prioritize the development and understanding of features likely to be correlated with the target variable (e.g., tweet classification). By the same token, features valuable to one target variable may not be important at all for a different target variable. Although often underappreciated, feature engineering is perhaps the most important part of a modeling exercise.

Deduplication. Further reduction of the amount of data to be processed can be achieved by removing near-duplicate messages. Given that in many social media platforms the number of people reposting a message can be interpreted as a measure of its importance, whenever removing near-duplicates it is advisable to save the number of near-duplicates that have been found (e.g., as done in Rogstadius et al. [2013] to prioritize highly reposted stories). Deduplication can be done by applying a clustering method (see Section 6.1.3).

Filtering. A fraction of the messages collected will not be relevant for a given crisis. This fraction depends on the specific collection method used, as discussed by Olteanu et al. [2014], and on other factors, such as the presence of off-topic messages using the same tags or keywords as the on-topic ones [Qu et al. 2011]. These messages can be post-filtered using human labeling or crowdsourcing, keyword-based heuristics, or automatic classification.

Additionally, many messages are posted automatically on social media for financial gain, exploiting the attention that a certain hashtag has received. These unsolicited commercial messages are known as *spam* [Gupta and Kumaraguru 2012; Uddin et al. 2014]; there are well-studied methods that can remove a substantial portion of them [Benevenuto et al. 2010]. Finally, in some cases we might want to also remove messages posted by automatic agents or social media *bots*. Their identification is similar to that of spammers.

4.4. Geotagging and Geocoding

Attaching geographical coordinates to a message (a process known as *geotagging*) is useful for a number of tasks in disaster response [Graham et al. 2014; Ikawa et al.

2013; Lingad et al. 2013]. Geotagging allows the retrieval of information about a *local event*, by filtering the messages corresponding to a particular geographical region. It allows the visualization of information about an event on a map, possibly making it more actionable for emergency responders. Geotagging can also be used for higher-level tasks, such as helping predict epidemic transmission of diseases based on geographical proximity [Brennan et al. 2013].

The availability of machine-readable location information in social media messages, in the form of metadata, depends on the user's device having the capacity to know its location (e.g., via Global Positioning System (GPS)), on the specific client software having the capability to read this from the device, and most important, on the user enabling this feature explicitly (opting in). In practice, only 2% of crisis-related messages include machine-readable location information [Burton et al. 2012].

However, while explicit metadata about locations may be absent, many messages in social media do contain implicit references to names of places (e.g., "The Christchurch hospital is operational"). *Geocoding* refers to finding these geographical references in the text, and linking them to geographical coordinates. This can be done by using a named entity extractor to extract potential candidates, and then comparing those candidates with a list of place names. This is the approach used by MacEachren et al. [2011], which uses Gate[17] for the first task and Geonames[18] for the second.

While building a comprehensive database of geospatial information, including place names, is an important component of geotagging (see, e.g., Middleton et al. [2014]), geotagging is not merely a dictionary look-up process because of ambiguities. These ambiguities are known as "*geo/non-geo*" and "*geo/geo*." A geo/non-geo ambiguity occurs, for instance, in the message "Let's play Texas Hold 'em," that does not refer to the state of Texas in the United States. A geo/geo ambiguity is found in the message "There is a fire in Paris," which may refer to the capital of France, or to any of more than a dozen places on Earth sharing the same name. For instance, Sultanik and Fink [2012] proposed an unsupervised approach to extract and disambiguate location mentions in Twitter messages during crisis situations.

In general, geotagging is done through probabilistic methods (see, e.g., Cheng et al. [2010]), often exploiting contextual clues. These clues may include the general location of a crisis, information about nearby places, and location information indicated by users in their profiles [Gelernter and Mushegian 2011].

4.5. Archived versus Live Data Processing

Depending on the urgency with which the output of an analysis is required, data may be provided to an algorithm either as an archive, for *retrospective* analysis, or as a live data feed, for *real-time* analysis. These correspond to two standard concepts in computer science: offline processing and online processing.

Retrospective data analysis (*offline processing*) starts with a batch of data relevant to an event, usually containing messages over the entire time range of interest. For example, we might consider reconstructing a timeline of events in the aftermath of an earthquake by looking at all of the tweets from the moment of the earthquake up to two weeks after. In deciding how to build the timeline, we have the complete context of events during this two-week window.

Live data analysis (*online processing*) is done over a stream of data relevant to the event, usually provided in real time or with a short delay. For example, we might consider constructing a timeline of the events in the aftermath of an earthquake *as*

[17]https://gate.ac.uk/.
[18]http://geonames.org/.

we observe new tweets; in deciding how to build the timeline, we have an incomplete context of the events and their future repercussions.

It is possible for algorithms to lie in between, operating on small batches of data at regular intervals (e.g., hourly, daily). The trade-off between retrospective and live data is a matter of accuracy versus latency. Retrospective data analysis maximizes our context and, as a result, gives us an accurate picture of the data. However, because we have to wait for the data to accumulate, we incur latency between when an event happens and when it is processed. Live data analysis, on the other hand, minimizes the latency; but, because we have partial information, we may incur lower accuracy. The choice of collection methodology depends on the use case. Crisis responders may want lower latency in order to better respond to a developing situation; forensic analysts may want higher accuracy and have the benefit of waiting for data to be collected.

While developing an algorithm, we can use retrospective data to *simulate* live data. This is a standard experimental methodology that has been used in the past for information-filtering tasks [Voorhees and Harman 2005] and, more recently, for crisis informatics [Guo et al. 2013; Aslam et al. 2013].

4.6. Challenges

There are a number of challenges associated with the processing of social media messages. In this section, we group them into two high-level categories: scalability and content. We defer specific challenges (e.g., challenges to event detection) to their respective technical sections to follow.

Scalability issues. Large crises often generate an explosion of social media activity. Data size may be an issue, as for crises that last several days, millions of messages may be recorded. While the text of each message can be sorted, a data record for a Twitter message (140 characters of text), for example, is around 4KB when we consider the metadata attached to each message. Thus, a Twitter collection for a crisis is on the order of several hundred megabytes to a few gigabytes. In addition, multimedia objects such as images and videos may significantly increase storage space requirements.

Data velocity may be a more challenging issue, especially considering that data does not flow at a constant rate, but experiences drastic variations. The largest documented peak of tweets per minute during a natural hazard that we are aware of is 16,000 tweets per minute.[19]

Finally, redundancy, which is commonly cited as a scalability challenge, to some extent cannot be avoided in this setting. Repeated (reposted/retweeted) messages are common in time-sensitive social media, even encouraged, as in some platforms messages that gain more notoriety are those that are simply repeated more.

Content issues. Microblog messages are brief and informal. In addition, this type of messaging is often seen by users to be more akin to speech, as opposed to a form of writing, which—compounded with technological, cross-lingual and cross-cultural factors—implies that on the Internet "we find language that is fragmentary, laden with typographical errors, often bereft of punctuation, and sometimes downright incoherent" [Baron 2003]. This poses significant challenges to computational methods, and can lead to poor and misleading results by what is known as the "garbage in, garbage out" principle.

Messages are also highly heterogeneous, with multiple sources (e.g., traditional media sources, eyewitness accounts), and varying levels of quality. Quality itself is an important and complex question for crisis managers, encompassing a number of attributes

[19]During Hurricane Sandy in 2012: http://www.cbsnews.com/news/social-media-a-news-source-and-tool-during-superstorm-sandy/.

including objectivity, clarity, timeliness, and conciseness, among others [Friberg et al. 2011]. Additionally, different languages can be present in the same crisis and sometimes in the same message—a phenomena known as "borrowing," and "code switching." This makes it difficult for both machines and humans (e.g., content annotators) to understand or classify messages.

Finally, brief messages sent during a crisis often assume a shared context from which only a minor part is sometimes made explicit. The area of study in linguistics known as *pragmatics* focuses on "communication in context," and explains how people are able to infer the meaning of the communications because humans are very adept at understanding context. Thus, in the case of Twitter communications, a reader can understand the tweet author's intent because that reader knows the context within which that tweet is being broadcast. Current computational methods are not able to make the same inferences humans do, thus cannot achieve the same level of understanding [Vieweg and Hodges 2014].

Privacy issues. Social media, since it is, by definition, content directly created by end users, might carry personally identifiable information (PII). Researchers and practitioners should be mindful of any explicit or inferred PII in the data. For example, an individuals may not explicitly reveal PII such as location for privacy reasons; inferring users' location may be in conflict with their expectation of privacy. As a result, academic work may have to be approved by a human subjects review process (ethics approval). Practitioners, on the other hand, may be subject to a response organization's own standards and policies (see, e.g., Gilman and Baker [2011] and ICRC [2013]).

5. EVENT DETECTION AND TRACKING

Most systems for social media processing during crises start with *event detection*. An event is the occurrence of something significant that is associated with a specific time and location [Brants et al. 2003]. However, due to the online nature of social media communications, events as they play out in social media may or may not be necessarily associated with a physical location. In the context of social media, Dou et al. [2012] define an event as: "An occurrence causing changes in the volume of text data that discusses the associated topic at a specific time. This occurrence is characterized by topic and time, and often associated with entities such as people and location."

Crisis and emergency situations typically fall into two broad categories: *predicted* (or *forewarned*) and *unexpected*. Some disaster events can be predicted to a certain level of accuracy based on meteorological or other data (e.g., this is the case with most large storms and tornadoes), and information about them is usually broadcast publicly before the actual event happens. Sometimes an event may not be explicitly anticipated, for example, as with a mass protest, but still may be forecast from social media and other data [Ramakrishnan et al. 2014]. Other events cannot be predicted (e.g., earthquakes); in this case, an automatic detection method is useful to find out about them as quickly as possible once they happen. In this section, we study techniques available for the automatic detection of both predicted and unexpected events.

Historically, methods for event detection and tracking in social media in the context of crises and emergencies are adaptations of methods to perform these tasks with a more general scope: that of detecting news. These methods, in turn, are adaptations of methods to find new topics in general document collections.

5.1. Background on Event Detection and Discovery

A well-studied problem in Information Retrieval is detecting events in a stream of documents (see, e.g., Allan [2002]). These documents can be news articles from traditional media sources or posts on social media (e.g., tweets, Facebook posts, Flickr

images). Traditionally, the *Topic Detection and Tracking* (TDT) research community uses newswires as source data streams for event detection.

Various techniques are employed in TDT including story segmentation, topic detection, new event detection, link detection, and topic tracking. *Story segmentation* focuses on determining story boundaries from streaming speech recognition output, usually from radio or television broadcasts. *Topic detection* groups relate documents together into cohesive topics. *New event detection* processes each new document to decide whether it describes a previously unseen story. *Link detection* detects whether two documents are similar. Finally, *event tracking* follows the evolution of an event/topic to describe how it unfolds.

Event detection on social media is different from the traditional event detection approaches that are suitable for other document streams. Social media data emerge more quickly and in larger volumes than traditional document streams. In addition, social media data are composed of short, noisy, and unstructured content that often require a different approach than what is used with traditional news articles. Considering the unique characteristics of social media streams, we focus the remainder of this section on new event detection and event tracking. Nevertheless, techniques and evaluation metrics from the TDT community provide insight into methods that might work for the Twitter domain.

5.2. New Event Detection (NED)

In the context of mass emergencies, New Event Detection (NED) refers to the task of discovering the first message related to an event by continuously monitoring a stream of messages. NED decides whether a message is about something new that has not been reported in previous messages [Yang et al. 2009]. "New" is normally operationalized as sufficiently different according to a similarity metric. Hellinger similarity, Kullback-Leibler divergence, and cosine similarity are among the metrics commonly used in NED [Kumaran et al. 2004].

As observed in Section 4.5, we can do retrospective (offline) or online analysis [Yang et al. 1998]. While the most useful NED systems for emergencies are those who perform this analysis online, they are often adaptations or improvements of offline methods, which we discuss next.

5.2.1. Retrospective New Event Detection. Retrospective NED refers to the process of identifying events using messages or documents that have arrived in the past. Methods for retrospective NED involve the creation of clusters of documents or messages based on a suitable definition of similarity between them, which may involve more than one dimension of them, for example, using similar words, involving similar groups of people, occurring close to each other in time and/or space, and so forth.

For instance, Zhao et al. [2007] introduce a retrospective NED method that uses textual, social, and temporal characteristics of the documents to detect events on social streams such as weblogs, message boards, and mailing lists. They build multigraphs using social textual streams, where nodes represent social actors, and edges represent the flow of information between actors. Clustering techniques and graph analysis are combined to detect an event.

Sayyadi et al. [2009] introduce a retrospective NED approach that overlays a graph over the documents, based on word co-occurrences. They assume that keywords co-occur between documents when there is some topical relationship between them. Next, a community detection method over the graph is used to detect and describe events. Pohl et al. [2012] describe a two-phase clustering approach to identify crisis-related subevents in the photo-hosting site Flickr and video-hosting site YouTube. During the first phase, which is to identify subevents, clusters are formed by using only items that

contain geographical coordinates. These coordinates are added automatically by the device used to capture the photo or video, or are added later by its author/uploader. Next, they calculate term-based centroids of the identified clusters using cosine distance to describe the identified subevents further.

Another retrospective NED approach is presented in Chen and Roy [2009], with experiments run on Flickr. It uses photos, user-defined tags, and other metadata including time and location to detect events. This approach simultaneously analyzes the temporal and geographical distribution of tags, and determines the event type (e.g., whether it is recurring) to form clusters. Finally, for each tag cluster, the corresponding photos are retrieved.

Ritter et al. [2012] extract significant events from Twitter by focusing on certain types of words and phrases. In their system, called *TwiCal*, they extract event phrases, named entities, and calendar dates. To extract named entities, they use a named entity tagger trained on 800 randomly selected tweets. To extract event mentions, they use a Twitter-tuned POS tagger [Ritter et al. 2011]. The extracted events are classified retrospectively into event types using a latent variable model that first identifies event types using the given data, and then performs classification.

Li et al. [2012b] introduce *Twevent*, a system that uses message segments instead of individual words to detect events. The authors claim that a tweet segment, which represents one or more consecutive words in tweets, contains more meaningful information than unigrams. The *Twevent* approach works in phases. First, the individual tweet is segmented, and bursty segments are identified using the segments' frequency in a particular time window. Next, identified segments are retrospectively clustered using K-Nearest Neighbors (KNN) clustering. Finally, a post-filtering step uses Wikipedia concepts to filter the detected events.

5.2.2. Online New-Event Detection. Online new-event detection does not use previously seen messages or any prior knowledge about the events to be identified. Online NED is typically performed with low latency (in real time), in the sense that the time between seeing a document corresponding to a new event and reporting that a new event has been detected is relatively short.

Methods based on keyword burst. A straightforward approach is to assume that words that show sharp frequency increases over time are related to a new event. For instance, Robinson et al. [2013a] introduce a system to detect earthquakes using Twitter. The earthquake detector, which is based on the Emergency Situation Awareness (*ESA*) platform [Power et al. 2014], checks for the keywords "`earthquake`" and "`#eqnz`" in the real-time Twitter stream, and applies a burst detection method to analyze word frequencies in fixed-width time windows and compare them to historical word frequencies. Unusual events are identified if the observed frequencies are much higher than those recorded in the past. Earle et al. [2012] have compared simple keyword-based approaches with data from seismological sensors, finding that while many earthquakes are not detected by Twitter users, detections are fast, "considerably faster than seismographic detections in poorly instrumented regions of the world."

Marcus et al. [2011] introduced *TwitInfo*, a system for detecting, summarizing, and visualizing events on Twitter. *TwitInfo* collects tweets based on a user-defined query (e.g., keywords used to filter the Twitter stream). It then detects events by identifying sharp increases in the frequency of tweets that contain the particular user-defined query as compared to the historical weighted running average of tweets that contain that same query. Further, tweets are obtained from the identified events to identify and represent an aggregated sentiment (i.e., classifying tweets into positive and negative classes). The authors evaluated the system on various events such as earthquakes and popular football games.

A similar system, *TwitterMonitor* [Mathioudakis and Koudas 2010], also collects tweets from the Twitter stream and detects trends (e.g., emerging topics such as breaking news or crises) in real time. The trend detection approach proposed in their article works in two phases. During the first phase, *TwitterMonitor* identifies bursty keywords, which are then grouped based on their co-occurrences. Once a trend is identified, additional information from the tweets is extracted to analyze and describe the trend. For example, the system uses Grapevine's entity extractor [Angel et al. 2009], which identifies entities mentioned in the trends.

Another Twitter-specific event detection approach introduced by Petrović et al. [2010] uses Locality Sensitive Hashing (LSH) for hashing a fixed number of recent documents in a bounded space, and processed in a bounded time, to increase the performance of nearest neighbors search.

With so many event-detection systems, a comparison of how they function is of interest. McMinn et al. [2013] describe a corpus to evaluate event-detection methods, composed of 500 news events sampled over a four-week period, and including tweet-level relevance judgments for thousands of tweets referring to these events. While existing NED systems have not been evaluated against this corpus, we anticipate this calibration of systems to occur in the future.

Beyond keyword bursts. There are well-known problems of relying on increases in the frequency of a keyword (or a segment) to detect events. For instance, consider popular hashtags such as "#musicmonday," which is used to suggest music on Mondays, or "#followfriday/#ff," which are used to suggest people to follow on Fridays. In these cases, there should be big pseudo-events detected every Monday and every Friday.

To address this problem, Becker et al. [2011] present an approach to classify real-world events from nonevents using Twitter. They use four types of features, which are temporal, social, topical, and Twitter-specific, to identify real events using the Twitter stream in real time. First, based on temporal features (i.e., volume of messages posted during an hour), they form initial clusters using the most frequent terms in the messages. Clusters are then refined using social features (i.e., users' interactions like retweets, replies, mentions). Next, they apply heuristics, for example, a high percentage of retweets and replies often indicates a nonevent, whereas a high percentage of mentions indicates that there is an event. Further, cluster coherence is estimated using a cosine similarity metric between messages and cluster centroid. Finally, as the authors report that multiword hashtags (e.g., "#musicmonday" and "#followfriday") are highly indicative of some sort of Twitter-specific discussion and do not represent any real event, they check the frequency of such hashtags used in each cluster to further refine the results.

Weng et al. [2011] present an algorithm for event detection from tweets using clustering of wavelet-based signals. Their approach involves three steps. First, they use wavelet transformation and auto-correlation to find bursts in individual words, and keep only the words with high-signal auto-correlations as event features. Then, the similarity for each pair of event-features is measured using cross-correlation. Finally, they use a modularity-based graph partitioning algorithm to detect real-life events. One of the strong points of this approach over the traditional event detection approaches is the capability of differentiating real-life big events from trivial ones. This is achieved mainly by two factors: the number of words, and the cross-correlation among the words related to an event.

Unlike the approaches presented earlier, Corley et al. [2013] present a method to detect and investigate events through metadata analytics and topic clustering on Twitter. Various features such as retweets, usage of different terms, and hashtags are analyzed for a certain time period to determine a baseline and noise ratio. An event is detected

once a particular feature value exceeds its noise boundaries and expected threshold. Once an event has been detected, its related topics are identified using the topic clustering approach.

Domain-specific approaches. As in many natural-language processing applications, approaches that are specific to a certain domain generally perform better than the approaches that are open-domain or generic.

For instance, Phuvipadawat and Murata [2010] describe a method for detecting breaking news from Twitter. First, tweets containing the hashtag "#breakingnews" or the phrase "breaking news" are fetched from the Twitter streaming API. Grouping of the extracted tweets is then performed, based on content similarity, and using a variant of the TF-IDF technique. Specifically, the similarity variant assigns a high similarity score to hashtags and proper nouns, which they identify using the Stanford NER implementation.

Data from traditional media sources can also be used for detecting newsworthy events. Not surprisingly, traditional media and social media have different editorial styles, perceived levels of credibility, and response delays to events. Tanev et al. [2012] find news articles describing security-related events (such as gun fights), and use keywords in their title and first paragraph to create a query. This query is issued against Twitter to obtain tweets related to the event. Dou et al. [2012] describe *LeadLine*, an interactive visual analysis system for event identification and exploration. *LeadLine* automatically identifies meaningful events in social media and news data using burst detection. Further, named entities and geo-locations are extracted from the filtered data to visualize them on a map through their interface.

Another domain-specific event-detection method is based on prespecified rules and introduced in Li et al. [2012a]. Their system, *TEDAS*, detects, analyzes, and identifies relevant crime- and disaster-related events on Twitter. First, tweets are collected based on iteratively refined rules (e.g., keywords, hashtags) from Twitter's streaming API. Next, tweets are classified via supervised learning based on content as well as Twitter-specific features (i.e., URLs, hashtags, mentions). Additionally, location information is extracted using both explicit geographical coordinates and implicit geographical references in the content. Finally, tweets are ranked according to their estimated level of importance.

Sakaki et al. [2010] detect hazards and crises such as earthquakes, typhoons, and large traffic jams using temporal and spatial information. The authors consider three types of features associated with tweets: statistical features (i.e., number of words in a tweet, position of the query word within a tweet), keyword-based features (i.e., the actual words in a tweet), and contextual features (e.g., words appearing near the query term, for instance if "earthquake" is a query, terms such as "magnitude" and "rocks" would be features in the tweet "5.3 magnitude earthquake rocks parts of Papua New Guinea"). In order to determine if a tweet corresponds to one of these hazards or crises, they use Support Vector Machines (SVM)—a known supervised classification algorithm (more about supervised classification in Section 6.1.2). *LITMUS* [Musaev et al. 2014] detects landslides using data collected from multiple sources. The system, which depends on the USGS seismic activity feed provider, the TRMM (NASA) rainfall feed, and social sensors (e.g., Twitter, YouTube, Instagram), detects landslides in real time by integrating multisourced data using relevance-ranking strategy (Bayesian model). Social media data is processed in a series of filtering steps (keyword-based filtering, removing stop-words, geotagging, supervised classification) and mapped based on geo-information either obtained from metadata or from content. Events with high relevancy are identified as "real events."

5.3. Event Tracking and Subevent Detection

Event tracking. Event tracking refers to the task of studying how events evolve and unfold. For a general discussion on the subject, see Allan [2002] and Lee et al. [2013].

The way in which emergency response agencies deal with crisis events varies as a crisis unfolds. Emergency situations typically consist of four phases: warning, impact, emergency, and recovery [Killian 2002, page 51]. During the warning phase, the focus is on monitoring the situation. Impact is when the disaster agent is actually at work, while the emergency phase is the immediate post-impact period during which rescue and other emergency activities take place. Recovery is the period during which longer-term activities occur, such as reconstruction and returning to a "normal" state.

Various techniques have been proposed to identify event phases. For instance, Iyengar et al. [2011] introduce an approach to automatically determine different phases of an event on Twitter. The approach, which is mainly based on content-based features of tweets, uses an SVM classifier and a hidden Markov model. Various content-specific features such as bag of words and POS tags are used to automatically classify tweets into three phases of an event: *before*, *during*, and *after*. A disaster-specific lexicon of discriminative words for each phase of the event can also be employed [Chowdhury et al. 2013].

Subevent detection. The detection of large-scale events has been studied with more attention than the detection of small-scale "subevents" that happen as a crisis unfolds.

Pohl et al. [2012] show the importance of subevent detection during crisis situations. They use multimedia metadata (tags and title) associated with content found on social media platforms such as YouTube and Flickr. Their framework uses a clustering approach based on self-organizing maps to detect subevents. First, a preselection of the data is performed based on user-identified keywords. The selected data is then passed to a subevent detection module that performs clustering to further split the data into subevents.

Khurdiya et al. [2012] present a system for event and subevent detection using Conditional Random Fields (CRF) [Lafferty et al. 2001]. Their system consists of four main modules: (1) a CRF-based event extractor to first extract actor, action, date, and location; event titles are also extracted using CRFs; (2) an event resolution to find similar events; (3) an event compiler that characterizes events; and (4) an event-reporting module, which is the end-user interface used to browse events details.

Hua et al. [2013] introduce *STED*, a semi-supervised targeted-interest event and subevent detection system for Twitter. To minimize the human effort required for labeling, they introduce an automatic label creation and expansion technique, which takes labels obtained from newspaper data and transfers them to tweets. They also propagate labels using mentions, hashtags, and retweets. Next, they build mini-clusters using a graph partitioning method to group words related to the event, and use supervised learning to classify other tweets using the examples provided by each mini-cluster. A final step on the classified output is to perform a location estimation using information from geo-coded tweets.

Table II lists event-detection systems. The majority of them have been surveyed earlier; additional tools are covered in the following sections.

5.4. Challenges

In addition to the discussion of general data challenges in Section 4.6, the following are particularly relevant to event detection.

Table II. Some of the Event Detection Tools Surveyed

System/tool	Approach	Event types	Real-time	Query type	Spatio-temporal	Sub-events	Reference
Twitter Monitor	burst detection	open domain	yes	open	no	no	[Mathioudakis and Koudas 2010]
TwitInfo	burst detection	earthquakes+	yes	kw	spatial	yes	[Marcus et al. 2011]
Twevent	burst segment detection	open domain	yes	open	no	no	[Li et al. 2012b]
TEDAS	supervised classification	crime/disasters	no	kw	yes	no	[Li et al. 2012a]
LeadLine	burst detection	open domain	no	kw	yes	no	[Dou et al. 2012]
TwiCal	supervised classification	conflicts/politics	no	open	temporal	no	[Ritter et al. 2012]
Tweet4act	dictionaries	disasters	yes	kw	no	no	[Chowdhury et al. 2013]
ESA	burst detection	open domain	yes	kw	spatial	no	[Robinson et al. 2013a]

The table includes the types of events for which the tool is built (open domain or specific), Whether detection is performed in real time, the type of query (open or "kw" = keyword-based), and whether it has spatio-temporal or subevent detection capabilities. Sorted by publication year.

Inadequate spatial information. Spatial and temporal information are two integral components of an event. Most systems that rely on Twitter data for event detection face challenges to determine geographical information of tweets that lack GPS information. In this case, automatic text-based geotagging can be used, as described on Section 4.4.

Mundane events. People post mundane events on social media sites. These data points introduce noise, which creates further challenges for an event detection algorithm to overcome. For such cases, separation of real-life big events from trivial ones is required.

Describing the events. Creating a description or label for a detected event is in general a difficult task. Often the keywords that are more frequent during the event are presented as a description for the event in the form of a list of words (e.g., {*sandy*, *hurricane*, *new york*}), but that list does not constitute a grammatically well-formed description (e.g., "*Hurricane Sandy hits New York*"). We present one approach to address this in Section 6.2.2.

6. CLUSTERING, CLASSIFICATION, EXTRACTION, AND SUMMARIZATION

Once we have found social messages related to a crisis event or topic, there are several ways in which we might process them. In this section, we describe some approaches. Broadly, we separate techniques into those classifying the data item as a whole and those extracting useful information from the content of one or more data items.

6.1. Classifying Social Media Items

In many situations, we are interested in classifying social media items into one or more categories. In this section, we will describe three methods of classification with decreasing levels of manual human supervision. Text classification is a large field of research, thus we will cover literature relevant to disasters. Interested readers should refer to other sources for a broader discussion [Srivastava and Sahami 2009].

6.1.1. Content Categories. There is not a single standard or widely accepted way of categorizing crisis-related social media messages. While some crisis-related ontologies have been proposed (see Section 7.2), in general, different works use different approaches.

Table III summarizes various dimensions of social media content that different research articles have used to classify information:

(1) By factual, subjective, or emotional content: to separate between facts (or combinations of facts and opinions), from opinions, or expressions of sympathy
(2) By information provided: to extract particular categories of information that are useful for various purposes
(3) By information source: to select messages coming from particular groups of users, for example, eyewitness accounts or official government sources
(4) By credibility: to filter out messages that are unlikely to be considered credible
(5) By time: to filter messages that refer to different stages of an event when temporal boundaries for the event are unclear
(6) By location: to select messages according to whether they originate from or near the place that was affected by an event, or from areas that were not affected

Other classification dimensions can be envisioned. In general, we observe that the selection of the set of categories used by researchers and practitioners is usually driven by two main factors: the data that is present in social media during crises and the information needs of response agencies. None of these factors is static, as both can change substantially from one crisis situation to another. We also remark that different types of disaster elicit different distributions of messages [Olteanu et al. 2015].

6.1.2. Supervised Classification. When a set of example items in each category is provided, a *supervised classification* algorithm can be used for automatic classification. This type of algorithm "learns" a predictive function or model from features of these examples (see Section 4.3) in order to label new, unseen data items. This set of examples is referred to as the *training set*. After a model has been learned from the training data, it is evaluated using a different, hold-out set of labeled items, not used during the training. This second set of examples is referred to as the *testing set*.

Depending on the nature of data in hand, different preprocessing techniques can be used. In any case, the input items are transformed into feature vectors, following the methods described in Section 4.3.

Training examples. The number of training examples required to achieve good accuracy depends on many factors, including the number of categories into which messages have to be classified and the variability of messages inside each category. Typical sizes of training sets range from a few hundred [Yin et al. 2012] to a few thousand [Imran et al. 2014a]. More examples yield better results in general, with diminishing results after a certain point. In general, the accuracy of models created using training data from one crisis decreases when applied to a different crisis, or when applied to the same crisis but at a different point in time [Imran et al. 2014b].

Feature selection. Even if messages are brief, the feature space in which they are represented is typically high dimensional (e.g., one dimension for every possible term). This introduces a number of problems including the amount of computational resources required for the data analysis, and it also increases the chances of overfitting the training data. In this case, a *feature selection method* (e.g., mutual information) should be employed as a first step to discard input features that have little or no correlation with the given training labels. Feature selection is an active area of machine-learning research; state-of-the-art techniques can be found in modern textbooks or journals (e.g., Guyon and Elisseeff [2003]).

Learning algorithms. After features have been extracted and selected, a machine-learning algorithm can be applied. Supervised classification algorithms include, among others, naïve Bayes, SVMs, logistic regression, decision trees, and random forests. The

Table III. Classification of Various Dimensions of Content Posted on Social Media During High-Impact Events with Description and Related Work References

Classification dimension	Description/examples
By factual, subjective, or emotional content	
Factual information	*(Examples under "By information provided")*
Opinions	opinions, criticism (e.g., of government response)
Sympathy	condolences, sympathy [Kumar et al. 2013]; condolences [Acar and Muraki 2011]; support [Hughes et al. 2014b]; thanks, encouragement [Bruns 2014]; prayers [Olteanu et al. 2014]
Antipathy	*schadenfreude*, animosity against victims (e.g., because of long-standing conflict)
Jokes	jokes, trolling [Metaxas and Mustafaraj 2013]
By information provided	
Caution and advice	caution and advice [Imran et al. 2013b]; warnings [Acar and Muraki 2011]; hazard, preparation [Olteanu et al. 2014]; tips [Leavitt and Clark 2014]; advice [Bruns 2014]; status, protocol [Hughes et al. 2014b]
Affected people	people trapped, news [Caragea et al. 2011]; casualties, people missing, found, or seen [Imran et al. 2013b]; self-reports [Acar and Muraki 2011]; injured, missing, killed [Vieweg et al. 2010]; looking for missing people [Qu et al. 2011]
Infrastructure/ utilities	infrastructure damage [Imran et al. 2013b]; collapsed structure [Caragea et al. 2011]; built environment [Vieweg et al. 2010]; closure and services [Hughes et al. 2014b]
Needs and donations	donation of money, goods, services [Imran et al. 2013b]; food/water shortage [Caragea et al. 2011]; donations or volunteering [Olteanu et al. 2014]; help requests, relief coordination [Qu et al. 2011]; relief, donations, resources [Hughes et al. 2014b]; help and fundraising [Bruns 2014]
Other useful information	hospital/clinic service, water sanitation [Caragea et al. 2011]; help requests, reports about environment [Acar and Muraki 2011]; consequences [Olteanu et al. 2014]
By information source	
Eyewitnesses/ bystanders	members of public [Metaxas and Mustafaraj 2013], victims, citizen reporters, eyewitnesses [Diakopoulos et al. 2012; Olteanu et al. 2014; Bruns 2014]
Government	administration/government [Olteanu et al. 2014]; police and fire services [Hughes et al. 2014b]; government [Bruns 2014]; news organization and authorities [Metaxas and Mustafaraj 2013]
NGOs	nongovernmental organizations [De Choudhury et al. 2012]
News media	news organizations and authorities, blogs [Metaxas and Mustafaraj 2013], journalists, media, bloggers [De Choudhury et al. 2012]; news organizations [Olteanu et al. 2014]; professional news reports [Leavitt and Clark 2014]; media [Bruns 2014]
By credibility	
Credible information	newsworthy topics, credibility [Castillo et al. 2013]; credible topics [Canini et al. 2011]; content credibility [Gupta and Kumaraguru 2012]; users and content credibility [Gupta et al. 2014]; source credibility [Thomson et al. 2012]; fake photos [Gupta et al. 2013]
Rumors	rumor [Hughes et al. 2014b; Castillo et al. 2013]
By time	
Prephase/ preparedness	posted before an actual event occurs, helpful for the preparedness phase of emergency management [Petak 1985]; predisaster, early information [Iyengar et al. 2011; Chowdhury et al. 2013]
Impact-phase/ response	posted during the impact phase of an event, helpful for the response phase of emergency management [Petak 1985]; during disaster [Iyengar et al. 2011; Chowdhury et al. 2013]
Post-phase/ recovery	posted after the impact of an event, helpful during the recovery phase [Petak 1985]; post-disaster information [Chowdhury et al. 2013; Iyengar et al. 2011]
By location	
Ground-zero	information from ground zero (victims' reports, bystanders) [De Longueville et al. 2009; Ao et al. 2014]
Near-by areas	information originating close to the affected areas [De Longueville et al. 2009]
Outsiders	information coming from other parts of the world, sympathizers [Kumar et al. 2013]; distant witness (in the sense of [Carvin 2013]); not on the ground [Starbird et al. 2012]; location inference [Ikawa et al. 2012]

choice of a method is largely dependent on the specific problem setting. For instance, *ESA* [Yin et al. 2012; Power et al. 2014] uses naïve Bayes and SVM, *EMERSE* [Caragea et al. 2011] uses SVM, *AIDR* [Imran et al. 2014a] uses random forests, and *Tweedr* [Ashktorab et al. 2014] uses logistic regression. While in most cases algorithms are used to predict a single label for each element, adaptations of these algorithms that generate multiple labels for each element are sometimes employed (e.g., Caragea et al. [2011]).

Ensemble/stacked classification. In some cases, an explicit model of a certain factor is desired, as exemplified by the work of Verma et al. [2011]. They observe that messages that contribute the most to situational awareness are also those that are expressed using objective (as opposed to subjective) language. In this case, one can create a *stacked classifier* in which at one level certain characteristics of the message are modeled (e.g., by having a classifier that classifies messages as objective or subjective) and at the next level these characteristics are combined with other characteristics also modeled by specific classifiers (e.g., writing styles such as formal or informal), and with features from the message itself. Verma et al. [2011] find that this approach performs better than directly using the input features.

Comparison in terms of classification effectiveness/accuracy is a nontrivial task. This is due to the use of different datasets, different baselines, and different performance measures (e.g., AUC, F1, precision). However, classification accuracies reported by most of the systems we have reviewed for this survey range from 0.60 to 0.90. Recent collections available for research may contribute to making systems easier to compare.[20]

6.1.3. Unsupervised Classification. Clustering is an *unsupervised machine-learning method*, a family of methods that seek to identify and explain important hidden patterns in unlabeled data. Unsupervised machine-learning methods include clustering, dimensionality reduction (e.g., principal component analysis), and hidden Markov models, among others.

The process of performing clustering begins by ingesting a set of items (e.g., documents, tweets, images), which are then processed with the objective of grouping similar items together. In general, the goal is to form clusters in such a way that elements within a cluster are more similar to each other than to the elements that belong to other clusters. Many clustering algorithms have been developed based on different approaches; examples include *K-means* (centroid-based), *hierarchical clustering* (connectivity-based), and *DBSCAN* (density-based). For an overview of clustering methods, see Zaki and Meira [2014, Part III].

In the context of dealing with social media data during crises, clustering can help reduce the number of social media messages that need to be processed/examined by humans, for instance by displaying multiple equivalent messages as a single item instead of multiple ones. This is the approach used by *CrisisTracker* [Rogstadius et al. 2013], which is a crowdsourced social media curation system for disaster awareness. The system, which collects data from Twitter based on predefined filters (i.e., keywords, bounding box), groups these tweets into *stories*, which are clusters of tweets. These stories are then curated/classified by humans, whose effort can be greatly reduced if they are asked to classify entire clusters instead of single tweets. The specific clustering method employed in this case is locality-sensitive hashing, an efficient probabilistic technique that uses hash functions to detect near-duplicates in data. Another example is *SaferCity* [Berlingerio et al. 2013], which identifies and analyzes incidents related to public safety in Twitter, adopting a spatio-temporal clustering approach based on the modularity maximization method presented in Blondel et al. [2008] for event identification.

[20]http://crisislex.org/.

Clusters are then classified using a semantic labeling approach using a controlled vocabulary and based on a rank score provided by the Lucene library.[21]

In addition to clustering methods that partition the items into groups, there are *soft-clustering* methods that allow an item to simultaneously belong to several clusters with varying degrees. This is the case of topic-modeling methods, out of which Latent Dirichlet Allocation (LDA) is one of the most popular. In the crisis domain, Kireyev et al. [2009] use topics extracted using LDA, including a weighting scheme that accounts for the document frequency (number of tweets containing a word) of the words in the tweet, as well as for the length of the tweet. When applied to data from an earthquake in Indonesia in 2009, it detects topics that cover different aspects of the crisis such as {*tsunami*, *disaster*, *relief*, *earthquake*}, {*dead*, *bodies*, *missing*, *victims*} and {*aid*, *help*, *money*, *relief*}. There is free software available for creating these topic models, such as MALLET [McCallum 2002], which has been applied to crisis data by Karandikar [2010].

6.1.4. Discussion. In the context of social media data analysis, many factors influence the choice of a learning approach (i.e., supervised vs. unsupervised). For instance, approaches based on unsupervised learning certainly have advantages over supervised ones in cases in which obtaining training examples would be prohibitively costly, would introduce an unacceptable delay, or would simply not be possible. Unsupervised approaches are more useful for cases in which information classes are completely unknown from an information-seeker point of view.

However, in terms of usefulness, the output of an unsupervised approach might be considered less useful than an approach based on supervised learning. Emergency responders may be used to specific categorizations and methodologies (e.g., the MIRA framework used by the United Nations[22]) and may expect that software tools output the same categories.

6.2. Subdocument Analysis

In contrast to text classification schemes, which make predictions about data items, subdocument analysis techniques extract granular information from the *content* of the data items. In this section, we present two important subdocument analysis methods in the context of crisis situations: information extraction and text summarization.

6.2.1. Information Extraction. The task of automatically extracting structured information from unstructured (e.g., plain text documents) or semi-structured (e.g., web pages) data is known as Information Extraction (IE). The most common IE task is named entity extraction, which consists of detecting regions of a text referring to people, organizations, or locations [Liu et al. 2011; Ritter et al. 2011]. This is the first step toward semantic enrichment (see Section 7.1).

In the context of crisis-related social media, IE can be used, for instance, to transform tweets reporting injured people in natural language (e.g., "*5 injured and 10 dead in Antofagasta*") to normalized records such as {<people-affected=5, report-type=injury, location=Antofagasta, Chile>, <people-affected=10, report-type=fatal-casualty, location=Antofagasta, Chile>}. These records are machine-readable, which means that they can be easily filtered, sorted, or aggregated.

Information extraction from social media is a challenging task because of the informal writing style and presence of many ungrammatical sentences, as was noted in Section 4.6. State-of-the-art approaches to IE involve the use of probabilistic sequential models such as hidden Markov models, conditional Markov models, maximum-entropy

[21]http://lucene.apache.org/.

[22]https://docs.unocha.org/sites/dms/Documents/mira_final_version2012.pdf (accessed Nov. 2014).

Markov models, or conditional random fields. Heuristics based on regular expressions can also be applied to this problem, although these are in general less effective than probabilistic methods.

Varga et al. [2013] use linguistic patterns and supervised learning to find "trouble expressions" in social media messages. For instance, in a tweet such as "*My friend said infant formula is sold out. If somebody knows shops in Sendai-city where they still have it in stock, please let us know*," the nucleus of the problem is the sentence "*infant formula is sold out*." This extraction is then used to match tweets describing problems to tweets describing solutions to those problems.

Imran et al. [2013a, 2013b] apply conditional random fields to the extraction of information from tweets. Their method proceeds in two steps. First, tweets are classified to consider categories such as "infrastructure damage," "donations," and "caution and advice." Then, a category-dependent extraction is done, for which, for "infrastructure damage" tweets, the specific infrastructure reported damaged is extracted, while for "donations" tweets, the item being offered in donation is extracted.

6.2.2. Summarization. Another approach to dealing with information overload involves presenting users with a text-based representation of the evolving event. Text summarization refers to the generation of a text summary of a document or set of documents [Nenkova and McKeown 2011]. Text summarization systems are optimized to generate a text summary that contains only the core topics discussed in the set of documents. Most systems produce this summary by *extracting* key sentences from the input document. This is in contrast to systems that produce a summary by *abstracting* or generating new sentences.

During crisis events, text summarization must be done in an *incremental* and *temporal* manner. Incremental text summarization, also known as *update summarization*, refers to generating a summary given: (1) the set of documents to be summarized, and (2) a reference set of documents that the user has read. The objective for the system is to produce a summary only of data that the user has not already read [Dang and Owczarzak 2008]. Temporal text summarization refers to creating an extractive summary from a set of time-stamped documents, usually in retrospect [Allan et al. 2001; Nallapati et al. 2004; Feng and Allan 2009].

Drawing on the work from summarization research, the TREC Temporal Summarization track focuses on generating updates relating to unfolding crisis events immediately after their occurrence [Aslam et al. 2013]. Table IV presents an example of this type of summary. The focus of this initiative is to first define standardized metrics for the task and then to encourage the development of systems optimizing them. These metrics include time-sensitive versions of precision and recall, ensuring that systems are penalized for *latency*: delivering information about an event long after it occurred. In addition, the metrics include a redundancy penalty to prevent systems from delivering repetitive information. In order to optimize these metrics, systems can use staged text analysis with standard information retrieval measures [McCreadie et al. 2013]. Alternatively, systems can use regression-based combinations of features from classic text summarization literature [Guo et al. 2013; Xu et al. 2013]. Other methods are purely content-based, hierarchically clustering sentence text [Wang and Li 2010]. In the context of social media, Shou et al. [2013] propose a system for online update summarization based on incremental clustering. The performance is evaluated under experimental conditions different from the TREC track, making it difficult to compare with other results.

Although algorithms for summarization exist for crisis events, their development is still preliminary and several challenges remain. Foremost, the relative importance of

Table IV. Example Summary from the TREC 2013 Temporal Summarization Track.

time	update
Nov 21, 2012 10:52	Tel Aviv bus bombing; 13 injuries; reported on bus line 142; occurred on Shaul Hamelech street; No claims of responsibility; 3 badly hurt; occurred in the heart of Tel Aviv near military hdqtrs
Nov 22, 2012 20:49	occurred in an area with many office buildings; occurred in area with heavy pedestrian traffic; first notable bombing in Tel Aviv since 2006; At least 28 people were wounded; Hamas' television featured people praising the attack; Khaled Mashal, leader of Hamas, categorically rejected any connection of the bombing to his group; ...
Nov 26, 2012 04:33	an Israeli Arab man was arrested on charges of planting the explosive device on the bus; Suspect was reportedly connected to Hamas; Suspect was reportedly connected to the Islamic Jihad
Nov 26, 2012 14:49	The Romanian Foreign Minister condemned the bombing,
Nov 29, 2012 04:55	govt rep refers to attack as terrorist attack
Nov 30, 2012 05:22	Fears about Bus Bomb Before the Cease-Fire: Could derail peace talks
Nov 30, 2012 06:47	The suspect remotely detonated the explosive device; suspect hid device in advance on the bus; The explosive device contained a large quantity of metal shrapnel designed to cause maximum casualties; The suspect later on confessed to carrying out the bus attack; suspect prepared the explosive device; suspect chose the target of the attack; ...

Updates reflect new or updated information as it is reported.

different features is not well understood. To date, research has primarily used features from batch text summarization. A second challenge is scale. Many algorithms require aggressive intersentence similarity computation, a procedure that scales poorly.

6.3. Challenges

In addition to the discussion on generic data-related challenges in Section 4.6, the following are particularly relevant to mining:

Combining manual and automatic labeling. In a supervised learning setting, human labels are necessary, but they may be costly to obtain. This is particularly problematic in crises that attract a multilingual population, or for tasks that require domain knowledge (e.g., people who know informal, local place names in Haiti, and who speak Haitian Creole). Labeled data are not always reliable and may not be available at the time of the disaster; in this case, a hybrid approach that mixes human labeling and automatic labeling can be employed [Imran et al. 2013c]. The selection of items to be labeled by humans can be done using *active learning*, a series of methods to maximize the improvement in classification accuracy as new labels are received.

Domain adaptation. Best practice is to avoid having to retrain an automatic classifier every time a new crisis occurs. However, simply reusing an existing classifier trained on data from a previous crisis does not perform well in practice, as it yields a substantial loss in accuracy even when the two crises have several elements in common [Imran et al. 2013a].

In machine learning, *domain adaptation* (or *domain transfer*) is a series of methods designed to maximize the accuracy of a classifier trained on one dataset, adapting it to continue to perform well on a dataset with different characteristics. To the best of our knowledge, these methods have not been applied to crisis-related social media data.

7. SEMANTIC TECHNOLOGIES IN DISASTER RESPONSE

One of the main goals of semantic technologies is to allow users to easily search through complex information spaces, and to find, navigate, and combine information. In the context of social media use during crises and mass convergence events, this is achieved by

linking data elements to concepts in a machine-readable way, enabling the representation of a situation as a complex and interrelated set of elements.

7.1. Semantic Enrichment of Social Media Content

Semantic technologies are particularly useful in social media, because they provide a powerful method for dealing with the variety of expressions that can be used to refer to the same concept, and with the many relationships that can exist between concepts. For instance, suppose we are looking for messages related to infrastructure damage using a keyword-query search. We would think that searching for something such as "damage AND (airport OR port OR bridge OR building . . .)" would be sufficient, until we notice that it is not only difficult to cover every particular type of infrastructure (airport, port, bridge, building, etc.) but also to cover every particular instance of that type (e.g., there are tens of thousands of airports in the world).

Named entity linking is a widely used semantic technology that deals with this problem. It operates in two phases. First, a *named entity recognizer* module detects entities—such as names of persons, places, and organizations. Second, for each named entity that is found, a *concept* is located that more closely matches the meaning of that named entity in that context. Concepts are generally operationalized as Uniform Resource Locators (URLs). For instance, Zhou et al. [2010] link named entities to the URLs of articles on Wikipedia. In a phrase such as "*Terminal 2 of JFK was damaged*," the named entity corresponding to the segment "JFK" would be linked to the URL https://en.wikipedia.org/wiki/John_F._Kennedy_International_Airport. There are several free and commercial services that can be used to perform named entity linking, including Alchemy,[23] Open Calais,[24] and Zemanta.[25]

Once an element is linked to a concept, further automatic annotation can be done by following links from the concept. Returning to our example of "*Terminal 2 of JFK was damaged*," if we go to its Wikipedia page,[26] we can learn that "JFK" is an instance of the class "airport." The airport concept is represented in this case by the URL of a Wikipedia category page to which the JFK Airport page belongs (https://en.wikipedia.org/wiki/Category:Airports), which in turn belongs to the category of transport building and structures (https://en.wikipedia.org/wiki/Category:Transport_buildings_and_structures).

After named-entity linking, messages that have been semantically enriched can be used to provide *faceted search*, a popular approach to interactively search through complex information spaces. In faceted search, the information of interest can be found not only by specifying a related keyword, but also by specifying a concept or concepts associated with the items of interest. In the example, we could select "airport" from a list of buildings and structures and then find a series of social media messages that are relevant but do not necessarily include the specific word "airport" in them.

Abel et al. [2011] present an adaptive faceted search framework to find relevant messages on Twitter. The framework enriches tweets with semantics by extracting entities (i.e., persons, locations, organization), then finding and linking those entities with external resources to create facets. Each facet enables search and navigation of relevant semantically related content. In follow-up work [Abel et al. 2012], they introduce *Twitcident*, a system that supports semantic filtering, faceted search, and summarization of tweets. The semantic-based approach is also implemented in *Twitris 2.0* [Jadhav

[23]http://www.alchemyapi.com/.
[24]http://www.opencalais.com/.
[25]http://developer.zemanta.com/.
[26]Or to its related semantic resource, DBPedia http://dbpedia.org/.

Table V. Crisis Ontologies, Including Some of the Classes and Attributes they Cover, and the Format in which They are Specified

Ontology Name	Coverage	Format	Reference
SOKNOS	resources, damage, disasters	OWL	[Babitski et al. 2011]
HXL	damage, geography, organization, disasters	RDF	http://hxl.humanitarianresponse.info/
SIADEX	processes, resources, geography	RDF	[de la Asunción et al. 2005]
OTN	specific to infrastructure	OWL	[Lorenz et al. 2005]
MOAC	damage, disasters, processes, resources	RDF	http://observedchange.com/moac/ns/
FOAF	emergency management people	RDF	http://www.foaf-project.org/
AktiveSA	transportation, meteorology, processes, resources, people	OWL	[Smart et al. 2007]
IntelLEO	response organizations	RDF	http://www.intelleo.eu/
ISyCri	damages, processes, disasters	OWL	[Bénaben et al. 2008]
WB-OS	features, components and information to build crisis management websites	XML	[Chou et al. 2011]
EDXL-RM	data exchange language for resource management	XML	https://www.oasis-open.org/

(OWL: Ontology Web Language, RDF: Resource Description Framework).

et al. 2010] and *EDIT* [Traverso et al. 2014], which present event-related social media capturing semantics in terms of spatial, temporal, and thematic dimensions.

7.2. Ontologies for Disaster Management

Information technologies to support disaster response often involve interactions between software operated by different agencies and/or provided by different developers or vendors. Allowing computer systems to communicate information in a unified way is a key challenge in general, but especially during crisis events in which different agencies must address different dimensions of a problem in coordination with each other [Hiltz et al. 2011]. Interoperability at the semantic level requires centralized specifications describing machine-understandable common vocabularies of concepts and linkages between them. An effective way of achieving this is to use machine-understandable ontologies that define, categorize, and maintain relationships between different concepts to facilitate common understanding and unified communication.

Table V lists some of the ontologies that have been introduced in recent years. Some examples from this table:

—The *Humanitarian eXchange Language* (HXL)[27] is an ontology created in 2011 and 2012 and is currently under review; it describes 49 classes and 37 properties. The focus of HXL is mainly on four areas: organization (i.e., formal response agencies like military, charities, NGOs), disaster (i.e., classification of disasters such as natural, man-made), geography (i.e., event location, geo-location of displaced people), and damage (i.e., damages related to humans, infrastructure).

—*Management of A Crisis* (MOAC)[28] is an ontology with 92 classes and 21 properties covering four areas: disaster, damage, processes (i.e., rescue, search, evacuation processes), and resources (i.e., services, vehicles, tents). HXL and MOAC have elements in common: in both cases, the objective is to describe different aspects of a crisis, including its effects, the needs of those affected, and the response to the crisis.

[27]http://hxl.humanitarianresponse.info/.
[28]http://observedchange.com/moac/ns/.

—*Integrated Data for Events Analysis* (IDEA)[29] is a framework for coding social, economic and political events. It is used in the *Global Database of Events, Language and Tone* (GDELT),[30] which is a machine-generated list of event data extracted from news reports.
—*Service-orientierte ArchiteKturen zur Unterstützung von Netzwerken im Rahmen Oeffentlicher Sicherheit* ("Service-oriented architectures supporting networks of public security," SOKNOS)[31] is an ontology for information integration for resource planning, including damage and resource categorization during disasters.

To support communication among different ontology-based systems, the problem of ontology heterogeneity needs to be solved by performing an *ontology mapping*, which is the process of mapping the concepts of two ontologies from the same or overlapping domains. Many approaches have been proposed to perform ontology mapping. For instance, Tang et al. [2006] treated this as a decision-making problem and proposed an approach based on Bayesian decision theory. For a survey on ontology mapping techniques, see Noy [2004].

The ontologies in Table V are crisis-specific, but not social media–specific. However, they can be combined with ontologies describing social media concepts such as users, tagging, sharing, and linking. For instance, the "Semantically Interlinked Online Communities" (SIOC)[32] ontology, originally developed to model sites such as blogs and online forums, has recently been extended to support the modeling of microblogs by adding concepts such as *follower* or *follows*. An ontology specific to Twitter appears in Celino et al. [2011] and includes user sentiments and locations, while Passant and Laublet [2008] enables semantic tagging of social media data through an ontology called Meaning-Of-A-Tag (MOAT). For a survey on ontologies developed for social media, see Bontcheva and Rout [2012].

8. SUMMARY AND FUTURE RESEARCH DIRECTIONS

In a relatively short time period—roughly 4 to 6 years—the research community working on the topics we have covered here has achieved a fairly high degree of maturity with respect to filtering, classifying, processing, and aggregating social media data during crises.

However, the underlying (although sometimes explicitly stated) claim behind this line of work, that is, that this research is *useful* for the public and/or formal response agencies, that it has the potential to save lives and/or property during an emergency, remains to be seen. While there are notable exceptions including the American Red Cross, the US Federal Emergency Management Agency (FEMA), the UN Office for the Coordination of Humanitarian Affairs, and the Filipino Government, the use of social media is still experimental for many organizations and not yet part of their normal, day-to-day operations.

There are two main directions in which we advocate future research going. First, continue deepening the data processing capabilities that have been the main focus of computing research on this topic thus far. Second, engage more deeply with human-centered approaches toward making the computing research the foundation of viable systems that emergency responders can implement.

[29]http://vranet.com/IDEA.aspx.
[30]http://gdeltproject.org/.
[31]http://soknos.de/.
[32]http://sioc-project.org/.

8.1. Deepening Data Processing Capabilities

From situational awareness to decision support. The systems we described in Section 3 for processing social media during disasters have a strong focus on situational awareness, which is an important first step but might not be enough for emergency management. During an emergency, social media are used as an information source in order to make decisions. Therefore, next-generation systems should be designed and evaluated in terms of their decision-support capabilities. This might even include forecasting using signals from social media.

Extending to other types of media. Data from various sources should be processed and integrated: "The strategies of emergency services organizations must also recognize the significant interweaving of social and other online media with conventional broadcast and print media" [Bruns 2014]. There are some examples of the processing of other types of information items during crises, including short messages (SMS) [Melville et al. 2013], news articles in traditional news media and blogs [Leetaru and Schrodt 2013], and images [Abel et al. 2012].

Verifying information. Systems that receive user-generated content are always exposed to abuse, which can be countered by a mixture of automatic and manual methods. For instance, algorithms to detect false product reviews are deployed by most major online retailers [Liu 2012, Chapter 10]. In the crisis domain, determining the credibility of information posted on social media is a major concern for those who process information (e.g., computer scientists and software engineers) and for the information consumers (e.g., the public and formal response organizations) [Hiltz et al. 2011; Hughes et al. 2014a].

Automatic classification can be used to filter out content that is unlikely to be considered credible [Gupta and Kumaraguru 2012; Castillo et al. 2013] or to annotate messages seen by users with credibility scores automatically [Gupta et al. 2014]. Additionally, the public itself can be mobilized to confirm or discredit a claim through crowdsourcing [Popoola et al. 2013].

8.2. Beyond Data Processing

Designing with the users. Considering the number of systems that have been designed and built so far, there is little research on how usable and useful those systems are (with some exceptions, e.g., Robinson et al. [2013b]; Tucker et al. [2012]).

How should information be presented to users? How should users interact with it? The key to answering these question lies with the users themselves, who should be brought into the process of designing the systems, dashboards, and/or visualizations that they require to serve their needs. A highly regarded methodology for achieving this is *participatory design* [Hughes 2014].

Helping governments and NGOs communicate with the public. Three days *before* Typhoon Pablo made its landfall in the Philippines in 2012, government officials were already calling users to use the hashtag "#pabloph" for updates about the typhoon. An effective use of hashtags has also been encouraged by the United Nations Office for the Coordination of Humanitarian Affairs [UN OCHA 2014]. Computational methods can be used not only to help formal response agencies choose which hashtags to use but, more generally, to help them design and evaluate effective communication strategies in social media (see Veil et al. [2011] regarding best practices for crisis communications using social media).

The conversation between the public and formal organizations can also be conducted through platforms that, instead of passively waiting for people to post information, ask

them directly to answer certain questions that are relevant for the emergency response or relief operations [Ludwig et al. 2015].

Social media for coordinating actions. The final output of the processing of social media messages is not limited to the presentation of information in a given format. Computational methods can be applied to augment the information in a number of ways. For instance, Varga et al. [2013] match problem-tweets ("infant formula is sold out in Sendai") to solution-tweets ("Jusco supermarket is selling infant formula"), and Purohit et al. [2013] match tweets describing urgent need of resources ("we are coordinating a clothing drive for families affected") with tweets describing the intention to donate them ("I've a bunch of clothes I want to donate"). We regard these efforts as preliminary results toward the ability to use social media as a mechanism for coordination of action in future emergency situations. A recent special issue of the *Journal on Computer Supported Collaborative Work* explores various ways that computing can support collaboration and coordination during an emergency [Pipek et al. 2014].

ACKNOWLEDGMENTS

We are thankful to Jakob Rogstadius and Per Aarvik, who pointed us to historical information. We would like to thank our collaborators and co-authors Patrick Meier, Alexandra Olteanu, Aditi Gupta, and Hemant Purohit.

REFERENCES

Fabian Abel, Ilknur Celik, Geert-Jan Houben, and Patrick Siehndel. 2011. Leveraging the semantics of tweets for adaptive faceted search on twitter. In *The Semantic Web*. Springer, 1–17.

Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. 2012. Semantics + filtering + search = Twitcident. Exploring information in social web streams. In *Proc. of Hypertext*. 285–294.

Adam Acar and Yuya Muraki. 2011. Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities* 7, 3, 392–402.

James Allan. 2002. *Topic detection and tracking: event-based information organization*. Springer.

James Allan, Rahul Gupta, and Vikas Khandelwal. 2001. Temporal summaries of new topics. In *Proc. of SIGIR*. ACM, 10–18.

Albert Angel, Nick Koudas, Nikos Sarkas, and Divesh Srivastava. 2009. What's on the grapevine?. In *Proc. of SIGMOD*. ACM, 1047–1050.

Ji Ao, Peng Zhang, and Yanan Cao. 2014. Estimating the locations of emergency events from Twitter streams. *Procedia Computer Science* 31, 731–739.

Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using Twitter. In *Proc. of EMNLP*. Association for Computational Linguistics, 1568–1576.

Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. 2014. Tweedr: Mining Twitter to inform disaster response. In *Proc. of ISCRAM*.

Javed A. Aslam, Fernando Diaz, Matthew Ekstrand-Abueg, Virgil Pavlu, and Tetsuya Sakai. 2013. TREC 2013 temporal summarization. In *Proc. of TREC*.

Marco Avvenuti, Stefano Cresci, Andrea Marchetti, Carlo Meletti, and Maurizio Tesconi. 2014. EARS (Earthquake Alert and Report System): a real time decision support system for earthquake crisis management. In *Proc. of KDD*. ACM, 1749–1758.

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*. Springer, 183–197.

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 2011. *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd ed.). Addison-Wesley Professional.

N. S. Baron. 2003. Language of the internet. In *The Stanford handbook for language engineers*, A. Farghali (Ed.). CSLI publications, Stanford, CA, 59–127.

Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond trending topics: Real-world event identification on Twitter. *Proc. of ICWSM*. 438–441.

Frédérick Bénaben, Chihab Hanachi, Matthieu Lauras, Pierre Couget, and Vincent Chapurlat. 2008. A metamodel and its ontology to guide crisis characterization and its collaborative management. In *Proc. of ISCRAM*.

Fabrcio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virglio Almeida. 2010. Detecting spammers on Twitter. In *Proc. of CEAS*, Vol. 6, 12.

Michele Berlingerio, Calabrese Francesco, Lorenzo Giusy, Di, Dong Xiaowen, Gkoufas Yiannis, and Dimitrios Mavroeidis. 2013. SaferCity: A system for detecting and analyzing incidents from social media. In *Proc. of ICDMW*. 1077–1080.

Deanne Bird, Megan Ling, and Katharine Haynes. 2012. Flooding Facebook—the use of social media during the Queensland and Victorian floods. *The Australian Journal of Emergency Management* 27, 1.

Justine I. Blanford, Jase Bernhardt, Alexander Savelyev, Gabrielle Wong-Parodi, Andrew M. Carleton, David W. Titley, and Alan M. MacEachren. 2014. Tweeting and tornadoes. In *Proc. of ISCRAM*.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 10.

Kalina Bontcheva and Dominic Rout. 2012. Making sense of social media streams through semantics: a survey. *Semantic Web.*

Thorsten Brants, Francine Chen, and Ayman Farahat. 2003. A system for new event detection. In *Proc. of SIGIR*. ACM, 330–337.

Sean Brennan, Adam Sadilek, and Henry Kautz. 2013. Towards understanding global spread of disease from everyday interpersonal interactions. In *Proc. of IJCAI*. AAAI Press, 2783–2789.

Axel Bruns. 2014. Crisis communication. *The Media and Communications in Australia,* 351–355.

Scott H. Burton, Kesler W. Tanner, Christophe G. Giraud-Carrier, Joshua H. West, and Michael D. Barnes. 2012. "Right Time, Right Place" health communication on Twitter: Value and accuracy of location information. *Journal of Medical Internet Research* 14, 6, e156.

Kevin Robert Canini, Bongwon Suh, and Peter L. Pirolli. 2011. Finding credible information sources in social networks based on content and social structure. In *Proc. of PASSAT*. IEEE, 1–8.

Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, H. Kim, Prasenjit Mitra, Dinghao Wu, A. Tapia, Lee Giles, Bernard J. Jansen, and others. 2011. Classifying text messages for the Haiti earthquake. In *Proc. of ISCRAM*.

Andy Carvin. 2013. *Distant Witness*. CUNY Journalism Press.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2013. Predicting information credibility in time-sensitive social media. *Internet Research* 23, 5, 560–588.

Irene Celino, Daniele Dell'Aglio, Emanuele Della Valle, Yi Huang, Tony Kyung-il Lee, Stanley Park, and Volker Tresp. 2011. Making sense of location-based micro-posts using stream reasoning. In *Proc. of MSM*. 13–18.

Ling Chen and Abhishek Roy. 2009. Event detection from Flickr data through wavelet-based spatial analysis. In *Proc. of CIKM*. ACM, 523–532.

Wei Chen, Laks V. S. Lakshmanan, and Carlos Castillo. 2013. Information and influence propagation in social networks. *Synthesis Lectures on Data Management* 5, 4, 1–177.

Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: A content-based approach to geo-locating Twitter users. In *Proc. of CIKM*. ACM, 759–768. DOI:http://dx.doi.org/10.1145/1871437.1871535

Chen-Huei Chou, Fatemeh Zahedi, and Huimin Zhao. 2011. Ontology for developing Web sites for natural disaster management: Methodology and implementation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 41, 1, 50–62.

Soudip Roy Chowdhury, Muhammad Imran, Muhammad Rizwan Asghar, Sihem Amer-Yahia, and Carlos Castillo. 2013. Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In *Proc. of ISCRAM*.

Camille Cobb, Ted McCarthy, Annuska Perkins, Ankitha Bharadwaj, Jared Comis, Brian Do, and Kate Starbird. 2014. Designing for the deluge: Understanding and supporting the distributed, collaborative work of crisis volunteers. In *Proc. of CSCW*.

Courtney D. Corley, Chase Dowling, Stuart J. Rose, and Taylor McKenzie. 2013. SociAL sensor analytics: Measuring phenomenology at scale. In *Proc. of ISI*. IEEE, 61–66.

Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS* 105, 41, 15649–15653.

H. T. Dang and K. Owczarzak. 2008. Overview of the TAC 2008 update summarization task. In *Proc. of TAC*.

Munmun De Choudhury, Nicholas Diakopoulos, and Mor Naaman. 2012. Unfolding the event landscape on Twitter: Classification and exploration of user categories. In *Proc. of CSCW*.

Marc de la Asunción, Luis Castillo, Juan Fdez-Olivares, Óscar García-Pérez, Antonio González, and Francisco Palao. 2005. SIADEX: An interactive knowledge-based planner for decision support in forest fire fighting. *AI Communications* 18, 4, 257–268.

Bertrand De Longueville, Robin S. Smith, and Gianluca Luraschi. 2009. OMG, from here, I can see the flames!: A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proc. of LBSN*. ACM, 73–80.

Nicholas Diakopoulos, Munmun De Choudhury, and Mor Naaman. 2012. Finding and assessing social media information sources in the context of journalism. In *Proc. of CHI*.

Wenwen Dou, Xiaoyu Wang, Drew Skau, William Ribarsky, and Michelle X. Zhou. 2012. Leadline: Interactive visual analysis of text data through event identification and exploration. In *Proc. of VAST*. IEEE, 93–102.

Russell R. Dynes. 1970. *Organized Behavior in Disaster*. Heath LexingtonBooks.

Russell R. Dynes. 1994. Community emergency planning: False assumptions and inappropriate analogies. *International Journal of Mass Emergencies and Disasters* 12, 2, 141–158.

Paul S. Earle, Daniel C. Bowden, and Michelle Guy. 2012. Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics* 54, 6.

Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1, 32–64.

Ao Feng and James Allan. 2009. Incident threading for news passages. In *Proc. of CIKM*. ACM, 1307–1316.

Julia D. Fraustino, Brooke Liu, and Yan Jin. 2012. *Social Media Use during Disasters: A Review of the Knowledge Base and Gaps (Final Report, START)*. Technical Report. Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security.

Therese Friberg, Stephan Prödel, and Rainer Koch. 2011. Information quality criteria and their importance for experts in crisis situations. In *Proc. of ISCRAM*. 145–149.

Liang Gao, Chaoming Song, Ziyou Gao, Albert-Laszlo Barabasi, James P. Bagrow, and Dashun Wang. 2014. Quantifying information flow during emergencies. *Scientific Reports,* 4.

Judith Gelernter and Nikolai Mushegian. 2011. Geo-parsing messages from microtext. *Transactions in GIS* 15, 6, 753–773.

Daniel Gilman and Leith Baker. 2011. *Humanitarianism in the Age of Cyber-warfare: Towards the Principled and Secure Use of Information in Humanitarian Emergencies*. Technical Report. United Nations Office for the Coordination of Humanitarian Affairs.

GlobalWebIndex. 2013. Stream Social Survey 2013 Q2. Retrieved May 11, 2015 from https://www.globalwebindex.net/blog/stream-social-q2-2013.

Mark Graham, Scott A. Hale, and Devin Gaffney. 2014. Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer* 66, 4, 568–578.

Qi Guo, Fernando Diaz, and Elad Yom-Tov. 2013. Updating users about time critical events. In *Advances in Information Retrieval*. Springer, Berlin, 483–494.

Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events. In *Proc. of PSOSM*. ACM, 2.

Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. TweetCred: Real-time credibility assessment of content on Twitter. In *Proc. of SocInfo*. Springer, 228–243.

Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy. In *Proc. of WWW (Companion)*. IW3C2, 729–736.

Isabelle Guyon and André Elisseeff. 2003. Special issue on variable and feature selection. *Journal of Machine Learning Research,* 3.

Harvard Humanitarian Initiative. 2011. *Disaster Relief 2.0 Report: The Future of Information Sharing in Humanitarian Emergencies*. Washington, DC and Berkshire, UK.

Starr Roxanne Hiltz, Paloma Diaz, and Gloria Mark. 2011. Introduction: Social media and collaborative systems for crisis management. *ACM Transactions on Computer–Human Interaction (TOCHI)* 18, 4.

Sounman Hong and Daniel Nadler. 2011. Does the early bird move the polls?: The use of the social media tool 'Twitter' by US politicians and its impact on public opinion. In *Proc. of DGR*. ACM, 182–186.

Ting Hua, Feng Chen, Liang Zhao, Chang-Tien Lu, and Naren Ramakrishnan. 2013. STED: Semi-supervised targeted-interest event detection in Twitter. In *Proc. of KDD*. ACM, 1466–1469.

Amanda L. Hughes. 2012. *Supporting the Social Media Needs of Emergency Public Information Officers with Human-Centered Design and Development*. Ph.D. Dissertation. University of Colorado at Boulder.

Amanda L. Hughes. 2014. Participatory design for the social media needs of emergency public information officers. In *Proc. of ISCRAM*.

Amanda L. Hughes and Leysia Palen. 2009. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management* 6, 3, 248–260.

Amanda L. Hughes and Leysia Palen. 2012. The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security and Emergency Management,* 1–20.

Amanda L. Hughes, Steve Peterson, and Leysia Palen. 2014a. Social media in emergency management. In *Issues in Disaster Science and Management: A Critical Dialogue Between Scientists and Emergency Managers*. FEMA in Higher Education Program.

Amanda L. Hughes, Lise A. St. Denis, Leysia Palen, and Kenneth Anderson. 2014b. Online public communications by police & fire services during the 2012 Hurricane Sandy. In *Proc. of CHI*.

ICRC. 2013. *Professional Standards for Protection Work Carried out by Humanitarian and Human Rights Actors in Armed Conflict and Other Situations of Violence*. Technical Report. International Committee of the Red Cross.

Yohei Ikawa, Miki Enoki, and Michiaki Tatsubori. 2012. Location inference using microblog messages. In *Proc. of WWW*. ACM, 687–690.

Yohei Ikawa, Maja Vukovic, Jakob Rogstadius, and Akiko Murakami. 2013. Location-based insights from the social web. In *Proc. of SWDM*. IW3C2.

Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014a. AIDR: Artificial intelligence for disaster response. In *Proc. of WWW (Companion)*. IW3C2, 159–162.

Muhammad Imran, Carlos Castillo, Ji Lucas, M. Patrick, and Jakob Rogstadius. 2014b. Coordinating human and machine intelligence to classify microblog communications in crises. *Proc. of ISCRAM*.

Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013a. Practical extraction of disaster-relevant information from social media. In *Proc. of WWW (Companion)*. 1021–1024.

Muhammad Imran, Shady Mamoon Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. 2013b. Extracting information nuggets from disaster-related messages in social media. In *Proc. of ISCRAM*.

Muhammad Imran, Ioanna Lykourentzou, and Carlos Castillo. 2013c. Engineering crowdsourced stream processing systems. *CoRR* abs/1310.5463.

Akshaya Iyengar, Tim Finin, and Anupam Joshi. 2011. Content-based prediction of temporal boundaries for events in Twitter. In *Proc. of PASSAT*. IEEE, 186–191.

Ashutosh Jadhav, Hemant Purohit, Pavan Kapanipathi, Pramod Ananthram, Ajith Ranabahu, Vinh Nguyen, Pablo N. Mendes, Alan Gary Smith, and Amit Sheth. 2010. A.: Twitris 2.0: Semantically empowered system for understanding perceptions from social data. In *Proc. of SWC*.

Murray E. Jennex. 2012. Social media–truly viable for crisis response. *Proc. of ISCRAM,* 1–5.

Anand Karandikar. 2010. *Clustering short status messages: A topic model based approach*. Master's thesis. University of Maryland.

Arpit Khurdiya, Lipika Dey, Diwakar Mahajan, and Ishan Verma. 2012. Extraction and compilation of events and sub-events from Twitter. In *Proc. of WI-IAT*. IEEE Computer Society, 504–508.

Lewis M. Killian. 2002. Methods for disaster research: Unique or not? In *Methods of Disaster Research*, Robert A. Stallings (Ed.). Xlibris, Philadelphia, PA, 49–93.

Kirill Kireyev, Leysia Palen, and K. Anderson. 2009. Applications of topics models to analysis of disaster-related Twitter data. In *NIPS Workshop on Applications for Topic Models: Text and Beyond*, Vol. 1.

Shamanth Kumar, Fred Morstatter, Reza Zafarani, and Huan Liu. 2013. Whom should I follow?: Identifying relevant users during crises. In *Proc. of Hypertext*.

Girdhar Kumaran, James Allan, and Andrew McCallum. 2004. *Classification models for new event detection*. Technical Report. University of Massachusetts Amherst.

John Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Alex Leavitt and Joshua A. Clark. 2014. Upvoting Hurricane Sandy: Event-based news production processes on a social news site. In *Proc. of CHI*. ACM, 1495–1504.

Pei Lee, Laks V. S. Lakshmanan, and Evangelos E. Milios. 2013. Event evolution tracking from streaming social posts. arXiv preprint arXiv:1311.5978.

Kalev Leetaru and Philip A. Schrodt. 2013. GDELT: Global database of events, language, and tone. In *Proc. of ISA*. http://gdelt.utdallas.edu/.

Chenliang Li, Aixin Sun, and Anwitaman Datta. 2012b. Twevent: Segment-based event detection from tweets. In *Proc. of CIKM*. ACM, 155–164.

Rui Li, Kin Hou Lei, Ravi Khadiwala, and KC-C Chang. 2012a. Tedas: A Twitter-based event detection and analysis system. In *Proc. of ICDE*. IEEE, 1273–1276.

Bruce R. Lindsay. 2011. *Social Media and Disasters: Current Uses, Future Options, and Policy Considerations*. Technical Report. US Congressional Research Service.

John Lingad, Sarvnaz Karimi, and Jie Yin. 2013. Location extraction from disaster-related microblogs. In *Proc. of WWW (Companion)*. IW3C2, 1017–1020.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.

Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. 2011. Recognizing named entities in tweets. In *Proc. of ACL*. Association for Computational Linguistics, 359–367.

Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. 2007. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition* 40, 1, 262–282.

Bernhard Lorenz, Hans Jürgen Ohlbach, and Laibing Yang. 2005. Ontology of transportation networks.

Thomas Ludwig, Tim Siebigteroth, and Volkmar Pipek. 2015. CrowdMonitor: Monitoring physical and digital activities of citizens during emergencies. In *Proc. of CHI*. 421–428.

Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. 2011. Senseplace2: Geotwitter analytics support for situational awareness. In *Proc. of VAST*. IEEE, 181–190.

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 *Proc. of CHI*. 227–236.

Michael Mathioudakis and Nick Koudas. 2010. Twittermonitor: Trend detection over the Twitter stream. In *Proc. of SIGMOD*. ACM, 1155–1158.

Andrew Kachites McCallum. 2002. *Mallet: A machine learning for language toolkit*.

Richard McCreadie, M.-Dyaa Albakour, Stuart Mackie, Nut Limosopathan, Craig Macdonald, Iadh Ounis, and B. Taner Dinçer. 2013. Experiments with terrier in contextual suggestion, temporal summarisation and web tracks. In *Proc. of TREC*.

Andrew J. McMinn, Yashar Moshfeghi, and Joemon M. Jose. 2013. Building a large-scale corpus for evaluating event detection on Twitter. In *Proc. of CIKM*. ACM, 409–418.

Patrick Meier. 2015. *Digital Humanitarians*. CRC Press.

Prem Melville, Vijil Chenthamarakshan, Richard D. Lawrence, James Powell, Moses Mugisha, Sharad Sapra, Rajesh Anandan, and Solomon Assefa. 2013. Amplifying the voice of youth in Africa via text analytics. In *Proc. of KDD*. ACM, 1204–1212. DOI:http://dx.doi.org/10.1145/2487575.2488216

Panagiotis Metaxas and Eni Mustafaraj. 2013. The rise and the fall of a citizen reporter. In *Proc. of WebSci*.

Stuart E. Middleton, Lee Middleton, and Stefano Modafferi. 2014. Real-time crisis mapping of natural disasters using social media. *Intelligent Systems, IEEE,* 9–17.

Dennis Mileti. 1999. *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Joseph Henry Press.

Jeffrey T. Mitchell, Deborah S. K. Thomas, Arleen A. Hill, and Susan L. Cutter. 2000. Catastrophe in reel life versus real life: Perpetuating disaster myth through Hollywood films. *International Journal of Mass Emergencies and Disasters* 18, 3, 383–402.

Aibek Musaev, De Wang, and Calton Pu. 2014. LITMUS: Landslide detection by integrating multiple sources. *Proc. of ISCRAM*.

Ramesh Nallapati, Ao Feng, Fuchun Peng, and James Allan. 2004. Event threading within news topics. In *Proc. of CIKM*. ACM, 446–453.

Ani Nenkova and Kathleen McKeown. 2011. *Automatic Summarization*. Vol. 5. 103–233 pages.

Natalya F. Noy. 2004. Semantic integration: A survey of ontology-based approaches. *ACM Sigmod Record* 33, 4, 65–70.

Ory Okolloh. 2009. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59, 1, 65–70.

Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In *Proc. of ICWSM*.

Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. 2015. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'15)*. ACM, New York, NY, 994–1009. DOI:http://dx.doi.org/10.1145/2675133.2675242

Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. In *Proc. of NAACL-HLT*. 380–390.

Leysia Palen and Sophia B. Liu. 2007. Citizen communications in crisis: Anticipating a future of ICT-supported public participation. In *Proc. of CHI*. ACM, 727–736.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1–2, 1–135.

Alexandre Passant and Philippe Laublet. 2008. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *Proc. of LDOW*.

William J. Petak. 1985. Emergency management: A challenge for public administration. *Public Administration Review,* 3–7.

Saša Petrović, Miles Osborne, and Victor Lavrenko. 2010. Streaming first story detection with application to Twitter. In *Proc. of ACL*. Association for Computational Linguistics, 181–189.

Swit Phuvipadawat and Tsuyoshi Murata. 2010. Breaking news detection and tracking in Twitter. In *Proc. of WI-IAT*, Vol. 3. IEEE, 120–123.

Volkmar Pipek, Sophia Liu, and Andruid Kerne (Eds.). 2014. *Special Issue: Crisis Informatics and Collaboration*. Computer Supported Cooperative Work, Vol. 23.

Daniela Pohl, Abdelhamid Bouchachia, and Hermann Hellwagner. 2012. Automatic identification of crisis-related sub-events using clustering. In *Proc. of ICMLA*, Vol. 2. IEEE, 333–338.

Hilary W. Poole, Laura Lambert, Chris Woodford, and Christos J. P. Moschovitis. 2005. *The Internet: a historical encyclopedia*. Abc-Clio Inc.

Abdulfatai Popoola, Dmytro Krasnoshtan, Attila-Peter Toth, Victor Naroditskiy, Carlos Castillo, Patrick Meier, and Iyad Rahwan. 2013. Information verification during natural disasters. In *Proc. of WWW (Companion)*. IW3C2, 1029–1032.

Robert Power, Bella Robinson, John Colton, and Mark Cameron. 2014. Emergency situation awareness: Twitter case studies. In *Proc. of ISCRAM Med*. 218–231.

Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. 2013. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* (2013).

Hemant Purohit and Amit Sheth. 2013. Twitris v3: From citizen sensing to analysis, coordination and action. In *Proc. of ICWSM*.

Yan Qu, Chen Huang, Pengyi Zhang, and Jun Zhang. 2011. Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake. In *Proc. of CSCW*.

Enrico L. Quarantelli. 2002. The Disaster Research Center (DRC) field studies of organized behavior in disasters. In *Methods of Disaster Research*, Robert A. Stallings (Ed.). Xlibris, Philadelphia, USA, 94–116.

Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, Chris Kuhlman, Achla Marathe, Liang Zhao, Ting Hua, Feng Chen, Chang-Tien Lu, Bert Huang, Aravind Srinivasan, Khoa Trinh, Lise Getoor, Graham Katz, Andy Doyle, Chris Ackermann, Ilya Zavorin, Jim Ford, Kristen Summers, Youssef Fayed, Jaime Arredondo, Dipak Gupta, and David Mares. 2014. 'Beating the News' with EMBERS: Forecasting civil unrest using open source indicators. In *Proc. of KDD*. ACM, 1799–1808.

Christian Reuter and Simon Scholl. 2014. Technical limitations for designing applications for social media. In *Mensch & Computer 2014–Workshopband*. Walter de Gruyter GmbH & Co KG, 131.

Alan Ritter, Sam Clark, Oren Etzioni, and others. 2011. Named entity recognition in tweets: An experimental study. In *Proc. of EMNLP*. Association for Computational Linguistics, 1524–1534.

Alan Ritter, Oren Etzioni, Sam Clark, and others. 2012. Open domain event extraction from Twitter. In *Proc. of KDD*. ACM, 1104–1112.

A. Robinson, Alexander Savelyev, Scott Pezanowski, and Alan M. MacEachren. 2013b. Understanding the utility of geospatial information in social media. In *Proc. ISCRAM*. 918–922.

Bella Robinson, Robert Power, and Mark Cameron. 2013a. A sensitive Twitter earthquake detector. In *Proc. of WWW (Companion)*. IW3C2, 999–1002.

Jakob Rogstadius, Maja Vukovic, Claudio Teixeira, Vassilis Kostakos, Evangelos Karapanos, and Jim Alain Laredo. 2013. CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development* 57, 5, 4–1.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proc. of WWW*. ACM, 851–860.

Aleksandra Sarcevic, Leysia Palen, Joanne White, Kate Starbird, Mossaab Bagdouri, and Kenneth Anderson. 2012. Beacons of hope in decentralized coordination: Learning from on-the-ground medical twitterers during the 2010 Haiti earthquake. In *Proc. of CSCW*. ACM, 47–56.

Hassan Sayyadi, Matthew Hurst, and Alexey Maykov. 2009. Event detection and tracking in social streams. In *Proc. of ICWSM*.

Wanita Sherchan, Surya Nepal, and Cecile Paris. 2013. A survey of trust in social networks. *ACM Comput. Surv.* 45, 4, Article 47, 33 pages. DOI:http://dx.doi.org/10.1145/2501654.2501661

Amit Sheth, Hermant Purohit, Ashutosh Jadhav, Pavan Kapanipathi, and Lu Chen. 2010. Understanding events through analysis of social media. In *Proc. of WWW*.

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,* 1228–1246.

Lidan Shou, Zhenhua Wang, Ke Chen, and Gang Chen. 2013. Sumblr: Continuous summarization of evolving tweet streams. In *Proc. of SIGIR*. ACM, 533–542.

Paul R. Smart, Alistair Russell, Nigel R. Shadbolt, Leslie A. Carr, and others. 2007. Aktivesa: A technical demonstrator system for enhanced situation awareness. *Comput. J.* 50, 6, 703–716.

Ashok Srivastava and Mehran Sahami. 2009. *Text Mining: Classification, Clustering, and Applications* (1st ed.). Chapman & Hall/CRC.

Robert A. Stallings. 2002. Methods for disaster research: Unique or not? In *Methods of Disaster Research*, Robert A. Stallings (Ed.). Xlibris, Philadelphia, PA, 21–44.

Kate Starbird. 2013. Delivering patients to Sacré Coeur: Collective intelligence in digital volunteer communities. In *Proc. of CHI*. ACM, 801–810.

Kate Starbird, Grace Muzny, and Leysia Palen. 2012. Learning from the crowd: Collaborative filtering techniques for identifying on-the-ground Twitterers during mass disruptions. *Proc. of ISCRAM*.

Kate Starbird, Leysia Palen, Amanda L. Hughes, and Sarah Vieweg. 2010. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proc. of CSCW*. ACM, 241–250.

Evan A. Sultanik and Clayton Fink. 2012. Rapid geotagging and disambiguation of social media text via an indexed gazetteer. In *Proc. of ISCRAM*, 190, 1–10.

Hristo Tanev, Maud Ehrmann, Jakub Piskorski, and Vanni Zavarella. 2012. Enhancing event descriptions through Twitter mining. In *Proc. of ICWSM*.

Jie Tang, Juanzi Li, Bangyong Liang, Xiaotong Huang, Yi Li, and Kehong Wang. 2006. Using Bayesian decision for ontology mapping. *Web Semantics: Science, Services and Agents on the World Wide Web* 4, 4, 243–262.

Robert Thomson, Naoya Ito, Hinako Suda, Fangyu Lin, Yafei Liu, Ryo Hayasaka, Ryuzo Isochi, and Zian Wang. 2012. Trusting tweets: The Fukushima disaster and information source credibility on Twitter. In *Proc. of ISCRAM*. 10.

Stefania Traverso, Valentina Cerutti, Kristin Stock, and Mike Jackson. 2014. EDIT: A methodology for the treatment of non-authoritative data in the reconstruction of disaster scenarios. In *Proc. of ISCRAM Med*. 32–45. DOI:http://dx.doi.org/10.1007/978-3-319-11818-5_4

S. Tucker, V. Lanfranchi, N. Ireson, A. Sosa, G. Burel, and F. Ciravegna. 2012. "Straight to the information I need": Assessing collational interfaces for emergency response. In *Proceedings of the 9th International ISCRAM Conference, Vancouver*.

Muhammad Moeen Uddin, Muhammad Imran, and Hassan Sajjad. 2014. Understanding types of users on Twitter. In *Proc. of SocialCom*.

UN OCHA. 2014. *Hashtag Standards For Emergencies*. Technical Report. United Nations Office for the Coordination of Humanitarian Affairs.

Jose van Dijck. 2013. *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press.

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 *Proc. of ACL*. 1619–1629.

Shari R. Veil, Tara Buehner, and Michael J. Palenchar. 2011. A work-in-process literature review: Incorporating social media in risk and crisis communication. *Journal of Contingencies and Crisis Management* 19, 2, 110–122.

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 *Proc. of ICWSM*.

Sarah Vieweg. 2012. *Situational Awareness in Mass Emergency: A Behavioral and Linguistic Analysis of Microblogged Communications*. Ph.D. Dissertation. University of Colorado at Boulder.

Sarah Vieweg, Carlos Castillo, and Muhammad Imran. 2014. Integrating social media communications into the rapid assessment of sudden onset disasters. In *Proc. of SocInfo*. 444–461.

Sarah Vieweg and Adam Hodges. 2014. Rethinking context: Leveraging human and machine computation in disaster response. *Computer* 47, 4, 22–27. DOI:http://dx.doi.org/10.1109/mc.2014.97

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 *Proc. of CHI*.

Ellen M. Voorhees and Donna K. Harman (Eds.). 2005. *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press.

Dingding Wang and Tao Li. 2010. Document update summarization using incremental hierarchical clustering. In *Proc. of CIKM*. ACM, 279–288.

Ingmar Weber and Yelena Mejova. 2013. Twitter and the Real World. CIKM'13 tutorial. Retrieved May 11, 2015 from https://sites.google.com/site/twitterandtherealworld/home.

Jianshu Weng and Bu-Sung Lee. 2011. Event detection in Twitter. In *Proc. of ICWSM*.

Tan Xu, Paul McNamee, and Douglas W. Oard. 2013. HLTCOE temporal summarization. In *Proc. of TREC*.

Christopher C. Yang, Xiaodong Shi, and Chih-Ping Wei. 2009. Discovering event evolution graphs from news corpora. *IEEE Transactions on Systems, Man and Cybernetics,* 850–863.

Yiming Yang, Tom Pierce, and Jaime Carbonell. 1998. A study of retrospective and on-line event detection. In *Proc. of SIGIR*. ACM, 28–36.

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.

Mohammed J. Zaki and Wagner Meira. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.

Qiankun Zhao, Prasenjit Mitra, and Bi Chen. 2007. Temporal and information flow based event detection from social text streams. In *Proc. of AAAI*, 7, 1501–1506.

Yiping Zhou, Lan Nie, Omid Rouhani-Kalleh, Flavian Vasile, and Scott Gaffney. 2010. Resolving surface forms to Wikipedia topics. In *Proc. of COLING*. Association for Computational Linguistics, 1335–1343.

Received July 2014; revised April 2015; accepted April 2015