



Spatial Computing and Social Media in the Context of Disaster Management

Nabil R. Adam, *US Department of Homeland Security*

Basit Shafiq, *Rutgers University*

Robin Staffin, *US Department of Defense, Office of Research and Engineering*

The growing use of smartphones and other powerful smart devices equipped with GPS and sensors has resulted in new types of spatial computing applications and technologies. Such technologies have enabled service providers to offer location-based and context-aware services to users that are always connected and interacting with various systems and information sources. These interactions can include retrieving and uploading multimodal data, searching content from online repositories, sending messages on social media, or offloading computing jobs to the cloud. Figure 1 illustrates this emerging spatial computing environment, which relies heavily on smart devices' extensive computation power, memory, and sensing capabilities, as well as the ubiquitous connectivity, computation resources, and storage offered by the existing communication and network infrastructure.

A key characteristic of this spatial computing environment is data sources' richness and diversity. Such sources include social media feeds, blogs, maps and GIS systems, digital libraries, e-government portals, television, and newsfeeds. Moreover, most currently available smartphones and handheld-devices (such as tablets) have various sensors, including a GPS, accelerometer, gyroscope, microphone, camera, and Bluetooth. Thus, new sensing applications have arisen across a wide variety of domains such as social networks, healthcare, education, weather, transportation, disaster management, gaming, and entertainment. These applications and sensors built around smartphones and other devices create a huge volume of data with different modalities and types (see Table 1).

These diverse data sources and smart devices' sensing capabilities are being exploited to provision location-based and context-aware services to a wide user base, including citizens, educational institutions, government, and businesses. One example of a context-aware service using sensing applications is a recent IBM project on intelligent transportation. Here, researchers aimed to help commuters avoid congestion and enable transportation agencies to better understand, predict, and manage traffic flow.¹ The project collects traffic data from various sensors on roads, toll booths, intersections, and bridges and combines it with location-based data from users' smartphones to learn their mobility patterns. Based on their preferred routes, participating users automatically received traffic information and alerts on their phones, resulting in reduced traffic congestion and accidents.

Another interesting application domain is that of mobile commerce,² in which businesses can provide location-based and context-aware services to customers. For example, merchants can provide offers to customers based on customers' location (provided by location services), interests and preferences (extracted from their social media profiles or Facebook "likes"), and availability (inferred from their online activity, such as check-ins).

Here, we look at disaster management as an application domain. In particular, we illustrate how citizens' participatory sensing coupled with social media can enable effective and timely information sharing for situational awareness and informed decision making. The combination of spatial computing and social media within the context of disaster management raises some unique and interesting challenges.

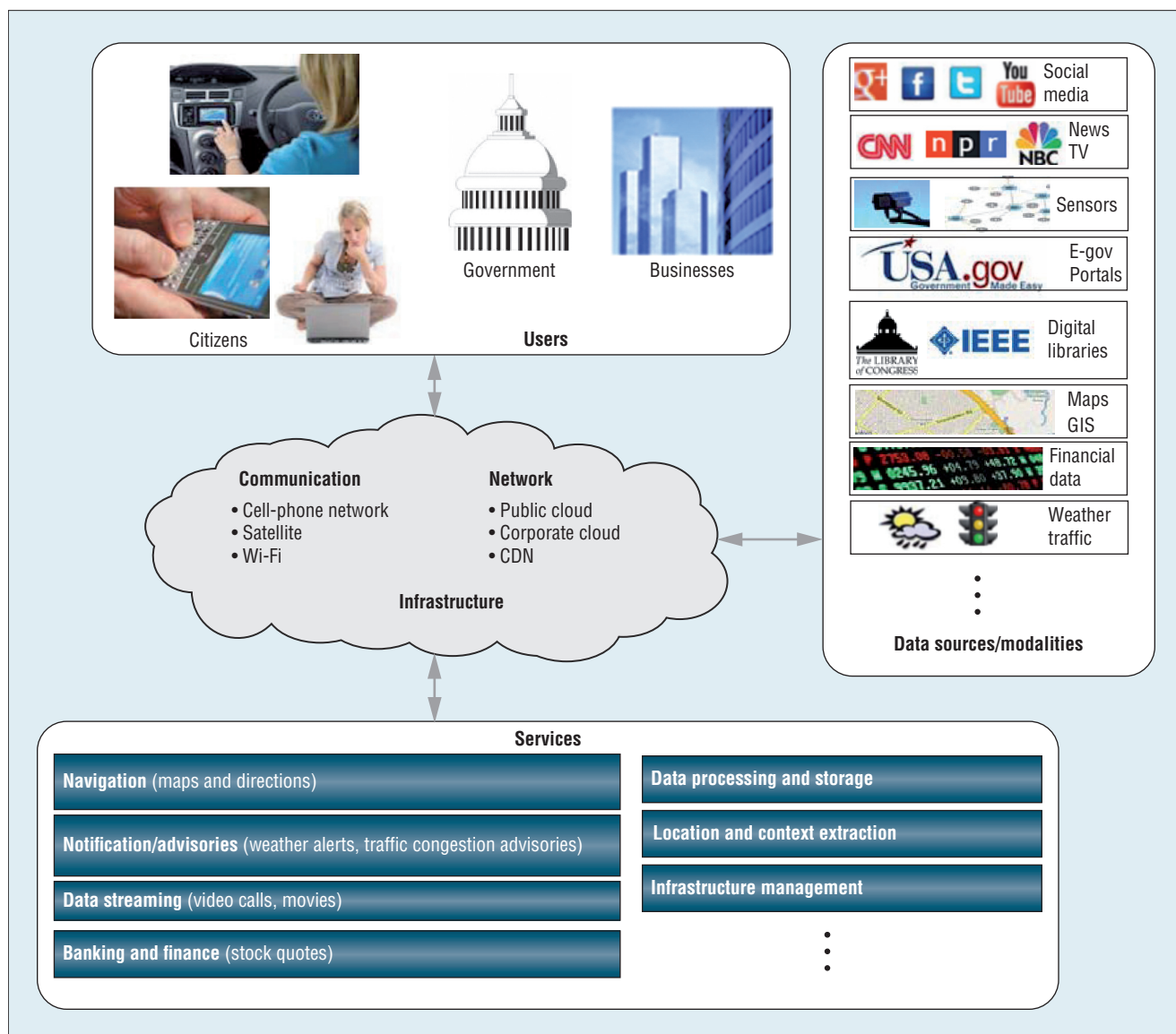


Figure 1. Spatial computing environment. This environment relies on smart devices' extensive computation power, memory, and sensing capabilities, as well as the ubiquitous connectivity, computation resources, and storage offered by the existing communication and network infrastructure.

Geoinformatics and Disaster Management

Social media, such as blogs, Twitter, and other information portals, have emerged as a dominant communication mechanism in today's society. For example, a recent survey by the American Red Cross reports that US citizens are increasingly relying on social media, mobile devices, and other on-line information portals to get information about ongoing disasters, seek assistance and safety information, and

report their safety and well-being during or after emergencies.³ The survey results indicate that close to 80 percent of survey participants expect emergency response organizations to monitor social sites during a disaster, more than 30 percent expect to receive help within one hour from posting a request to social sites, and 24 to 30 percent use social media to report their well-being to their loved ones.

Exploiting such input to gain awareness of an incident is a critical

direction for research in effective emergency management. Dynamic real-time incident information collected from onsite human responders about the extent of damage, an event's evolution, the community's needs, and the responders' ability to deal with the situation, combined with information from the larger emergency management community, could lead to more accurate and real-time situational awareness that enables informed decisions, better

Table 1. Multimodal data.

Data type	Embedded information
Voice calls	Audio sample, caller/called number, date, and time
Short Message Service (SMS)	Message transcript, caller/called number, date, and time
Multimedia Message Service (MMS)	Multimedia object (image, audio, video, and so on), geotagged location, caller/called number, date, and time
Social media feeds	Application type (Twitter, Facebook, and so on), type of event (posting or notification, for example), media object (text message, video, audio), date, and time
Geolocation data	GPS measurement of current location, accelerometer samples, gyroscope samples
Network connectivity data	Cell tower and WLAN access point observation and their location, Bluetooth observations

resource allocation, and thus better response and outcome to the overall crisis.

The US Department of Homeland Security's Science & Technology Directorate (DHS-S&T) has initiated Social Media Alert and Response to Threats to Citizens (SMART-C). This program is in line with the DHS's objective of leveraging social media effectively, securely, appropriately, and in a manner that respects individuals' privacy.⁴ SMART-C aims to develop citizens' participatory sensing capabilities for decision support throughout the disaster life cycle via a multitude of devices (such as smartphones) and modalities (MMS messages, Web portals, blogs, tweets, and so on). Specifically, the objective is to establish a bidirectional link between emergency response authorities and citizens that facilitates receiving early warning signals, detecting incidents and how they evolve, communicating alerts and advisories to citizens during and after the incident for response and recovery, and getting citizens' feedback for post-incident analysis and reconnaissance.

Figure 2 provides a detailed view of the spatial computing environment in the context of SMART-C. In particular, this figure depicts the data sources with their modalities, users, and services provided by the underlying system to enable enhanced situational awareness and informed

decision making during all phases of disaster management. These data sources, in addition to those Figure 1 shows, include resource databases and incident-management systems from both government agencies and private-sector organizations, as well as open source data and online services providing such information as weather, traffic, hospital availability, and demographics. Moreover, real-time data from "311" and "911" systems can also feed data to SMART-C. Users can be citizens who would use their smart devices to request assistance or receive incident updates and advisories from emergency management officials through social media channels, mobile apps, or the Commercial Mobile Alert System (CMAS). Government agencies and private-sector organizations involved in disaster management efforts are the key stakeholders who would use SMART-C's services and participatory sensing capabilities to achieve real-time situational awareness, request updates from citizens as the incident evolves, and provide effective response. SMART-C's open architecture and use of open standards minimize efforts needed to interface with existing incident-management systems such as E Team (www.nc4.us/ETeam.php) and WebEOC (http://esi911.com/esi/index.php?Itemid=30&id=14&option=com_content&task=view).

A key aspect of SMART-C is *service composability*, which enables rapid design and development of incident-management and response processes using underlying technical capabilities and services such as data ingestion, multimodal data enrichment and analysis, event and location extraction, and customized alert dissemination. Different government systems or third parties provide these services. Let's look at an example scenario of service composability.

Consider the public health surveillance process Figure 3 depicts. This process uses social media feeds to detect and track public health emergencies. For this surveillance process, SMART-C uses a social media subscription service to retrieve and filter those tweets or Facebook postings that mention health-related events, such as individuals getting sick. SMART-C then feeds all these messages to the event-extraction and correlation service to extract and disambiguate the different event types. This service also utilizes geospatial reasoning to automatically extract and characterize the event in both space and time. For example, assume that multiple residents post tweets about getting sick after eating at local restaurants in a given region (for example, the southern New Jersey area)—the Twitter feeds might reference different restaurants and report different symptoms (fever, stomach ache, and so on).

On the basis of these feeds, the event-extraction service identifies that several tweets in close spatial and temporal proximity are reporting a related event with possibly different symptoms. To assess the event's reliability, SMART-C must corroborate the information from tweets and Facebook postings. To do this, a separate service in the public health surveillance

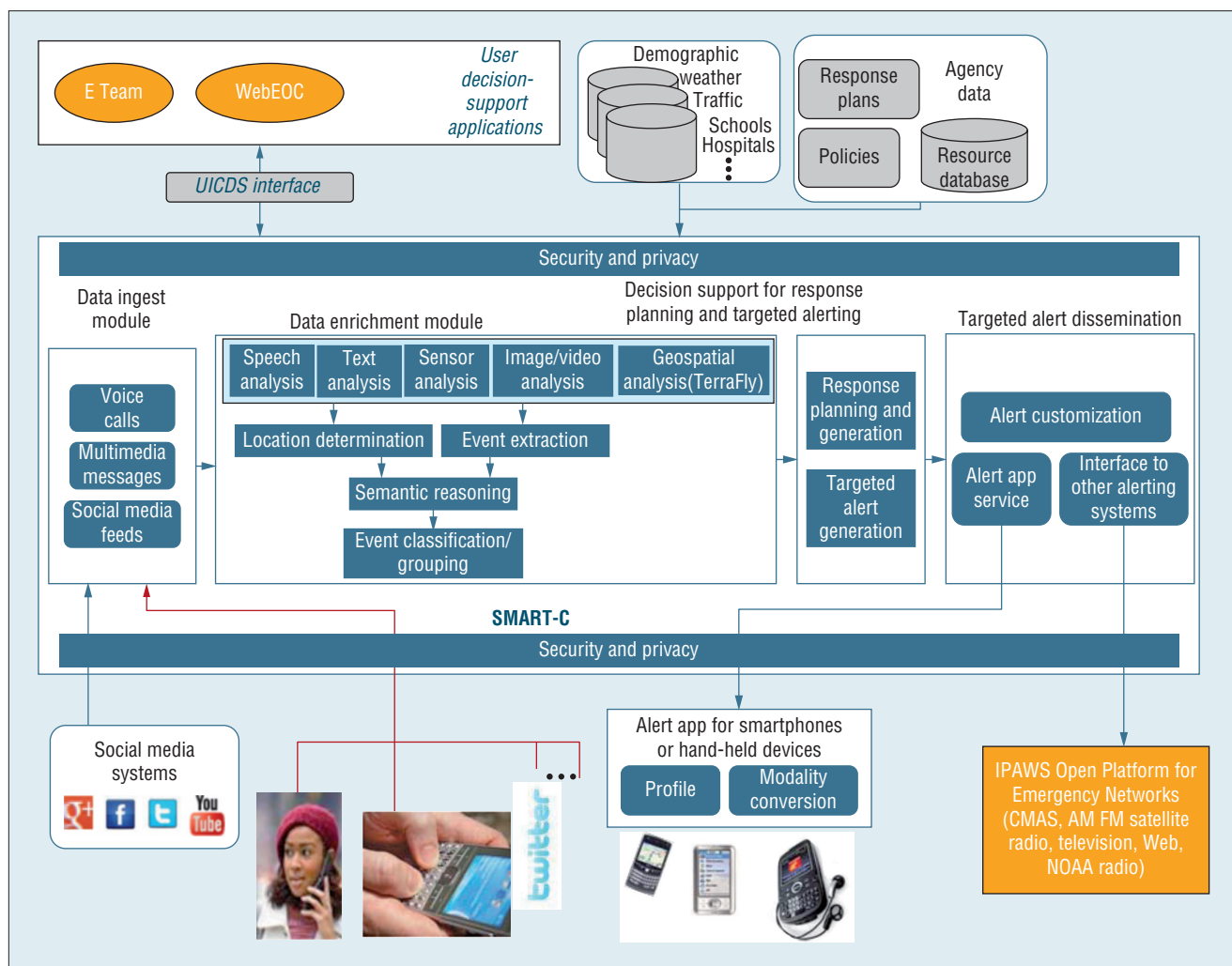


Figure 2. The Social Media Alert and Response to Threats to Citizens (SMART-C) architecture. SMART-C is a citizen-participatory sensing system that performs rigorous analytics to create a coherent and reliable picture of an event in support of customized alerting and response.

process retrieves incident reports from hospital databases within the region as well as alerts from the CDC and newsfeeds. In addition, SMART-C uses a medical ontology and reasoning service to determine the event's likely cause (for example, a salmonella outbreak) given the symptoms reported in the social media messages. Depending on the event's assessed reliability, SMART-C (through its incident-sharing service) would share incident information with local authorities' systems such as E Team and WebEOC for further investigation. Other restaurants in the region would also be

alerted. Furthermore, SMART-C would send "customized" alerts to citizens based on their location and temporal attributes—that is, alerts that those who might have visited such restaurants or purchased produce from local farms within a specific time period should seek medical help if they develop related symptoms. For this, we can use an alert-dissemination service such as CMAS or mobile apps.

Research Challenges

The combination of spatial computing and social media can present some unique challenges.

Privacy

The geospatial data retrieved from smartphones, sensors, and other devices often contains sensitive personal information. When combined with social media data, this information raises concerns about increased privacy breaches. We must address such privacy concerns in all phases of spatial computing, including data collection, storage, analysis, and dissemination. Moreover, these concerns aren't limited to data content only, but also extend to context information such as an individual's location, the time when a particular activity was observed, and activity patterns

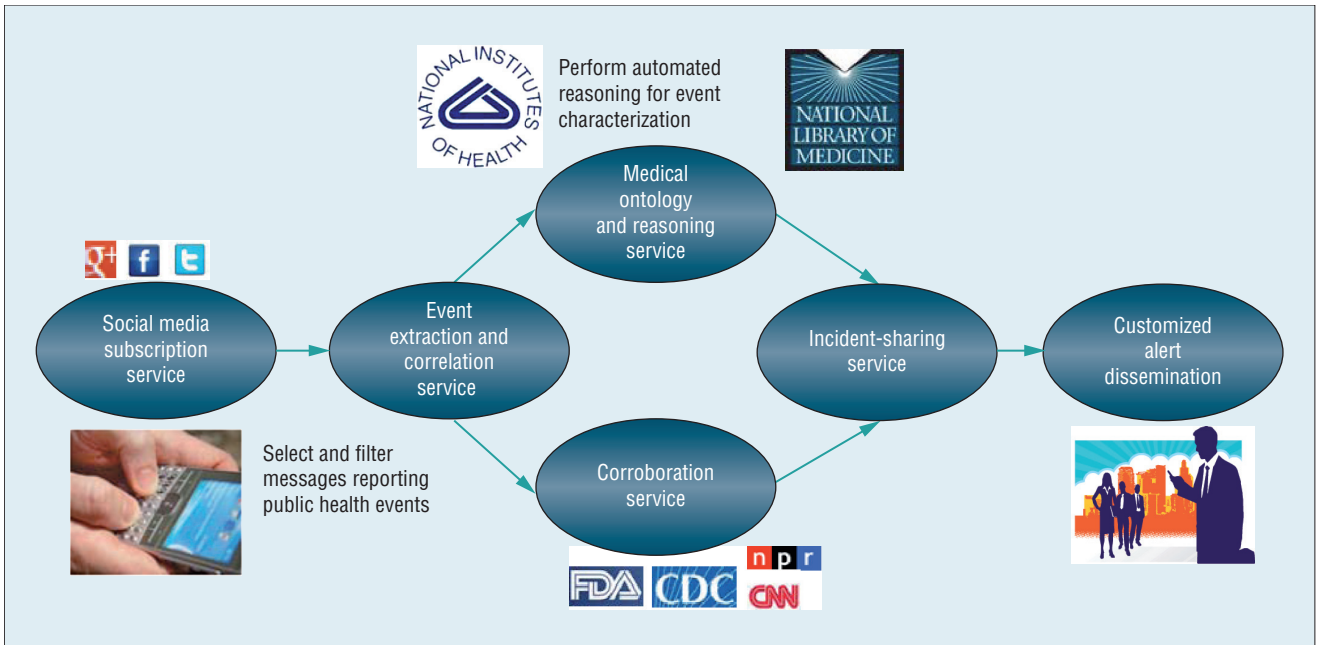


Figure 3. Public health surveillance process. This process uses social media feeds such as tweets and Facebook postings to detect and track public health emergencies.

that adversaries could use to profile a user. In a disaster management context, we can use social media and mobile apps for situational awareness and to disseminate customized alerts and advisories based on users' location, language, and special needs. The challenge is how to achieve this targeted and customized alert and response while respecting individuals' privacy.

We next look at some key privacy issues and proposed approaches for addressing them.

Data privacy. Preventing the disclosure of data content that can individually identify a person is known as *data privacy*. Such content could be tweets and user blogs, data from users' sensor-enabled smart devices (such as images, videos, driving speed, temperature, and blood pressure), or users' searches or queries. Approaches to protecting data privacy fall under several categories: anonymization,⁵ perturbation,⁶ differential privacy,⁷ auditing,⁸ and usage control.⁹

Anonymization-based approaches restrict and generalize data items into

groups such that each group has at least a minimum number k of indistinguishable data items (known as k -anonymity). Data-perturbation methods fall into two main categories: *Probability distribution* considers the dataset to be a sample from a given population that has an underlying probability distribution. In this case, the security-control method replaces the original dataset with another sample from the same distribution used to answer the queries. *Output-perturbation* approaches ensure security by perturbing the query results.

Differential privacy provides a formal and quantifiable privacy guarantee irrespective of an adversary's background knowledge and available computational power. Differential privacy is actually a condition on the data-release mechanism rather than on the dataset and requires that any computation based on the underlying database should be insensitive to the change of a single data item. This ensures that any computation on the dataset wouldn't reveal any individual item in it. These approaches have been designed primarily for

relational or graph-structured data; we can't apply them straightforwardly when dealing with unstructured multimodal data that comes from different sources such as social media, blogs, text or multimedia messages, and voice calls.

Auditing involves keeping up-to-date logs of all queries each user makes and constantly checking each new query for possible compromise. However, keeping a log of each access in a spatial computing environment with numerous data sources and users, as Figure 1 shows, is challenging. Usage control is a strategy whereby users can specify their access policies for data sharing. However, it requires infrastructure support to specify and enforce data owners' access-control and privacy policies. Given the large number of users who publish their data and query data from other users, supporting the specification and enforcement of users' (data owners') access-control policies is a major challenge.

All these approaches focus on personally identifiable information (PII) protection at the data storage and

analysis phases and not at the data collection phase. Some work addresses data privacy at the data collection phase, but is limited to specific application contexts, such as video surveillance¹⁰ and polling data.¹¹ Given the large number of data sources and data modalities in the spatial computing environment, new approaches for PII protection at the data collection phase are necessary. Moreover, such approaches must take real-time considerations into account.

Context privacy. Context privacy deals with preventing the disclosure of contextual information such as a person's location or the time when an individual performs a certain activity. Both location and temporal privacy are major concerns in the spatial computing environment because disclosing such information lets adversaries track and re-identify individuals even with de-identified data.^{12,13} Moreover, an adversary could learn the type and nature of data content being transmitted.¹²

Location and temporal privacy approaches rely on anonymization,¹⁴ perturbation, path cloaking,¹⁵ and mixing identifiers.¹⁶ In this context, anonymization techniques aim at generalizing a user's position to a region that contains k other users, thus hiding the user's identity. Perturbation approaches add random noise to users' location data, thereby preventing disclosure and prediction of the user's actual position at any given point in time. Path cloaking uses a privacy metric of time-to-confusion, which we define as the maximum time an adversary can track an anonymous user. The cloaking algorithm ensures that within the released trajectory dataset, sufficient areas of confusion exist with several mobile users, so that an adversary can't track

any particular mobile user with a high degree of certainty. The mixing identifier approach relies on changing or swapping mobile users' identifiers at intersection points so that adversaries can't track users beyond a certain duration.

The challenge with location and temporal privacy protection approaches is how to achieve the right balance between location and temporal privacy and data utility,¹⁷ as well as how to let users specify their privacy preferences at an acceptable level while receiving the desired location-based and context-aware services.

Detecting and characterizing events from unstructured multimodal data is a key challenge in the spatial computing environment.

Event Detection and Characterization

Detecting and characterizing events from unstructured multimodal data is a key challenge in the spatial computing environment given the large number of sources producing volumes of multimodal data. When viewed in isolation, data from different sources can appear irrelevant, but when analyzed collectively, it could reveal interesting events.¹⁸ For example, as illustrated in the healthcare surveillance process scenario, spatiotemporal correlation of the different tweets and Facebook postings by various users reporting their illnesses or health conditions led to the correct

characterization of the event—a salmonella outbreak.

We next examine some of the challenges related to event detection and characterization.

Integration and enrichment of multimodal data. Given the large volume of multimodal and high-velocity data, integration and enrichment of such data to extract semantically meaningful information is a challenging task, especially when real-time constraints are considered. Developing a system such as SMART-C that provides such capabilities requires exploiting recent advances in areas including speech analysis for geolocation extraction, semantic analysis of speech and text input, multimodal data alignment and event extraction, semantic event-based middleware, spatiotemporal reasoning, and fuzzy geospatial joins to disambiguate events and track event progression in space and time.⁴

Data quality and reliability. Improved data quality is essential for robust event identification and characterization. The spatial computing environment, where data are often collected and assimilated automatically, exacerbates data quality concerns (for example missing or erroneous data, uncertainty, or fidelity). Addressing these issues requires techniques that add resilience to data collection mechanisms while considering failures in sensing and communication systems, as well as the limitations of data processing systems. Another critical aspect related to data quality is data's relevance to the event or phenomenon under consideration. For example, a public health emergency event such as a salmonella outbreak can lead to a host of tweets and message postings that mention some type of health-related symptoms or illness.

However, only a few of these might be relevant to the underlying event. This raises the issue of precision and recall trade-offs—a central issue in the information retrieval area. Precision requires accurately retrieving the relevant messages and dropping the irrelevant ones. Recall, on the other hand, requires that no relevant message be missed. This trade-off becomes more challenging when the results need to be presented in real time.

Validation and corroboration. A key issue when using social media data is how reliable and accurate that data is, given that it's often collected from anonymous participants. This requires algorithms and techniques that can corroborate and correlate multimodal data from multiple sources in real time.

Smart Devices and the Cloud

Today, smart devices use the cloud, via RESTful Web services, for processing capabilities, storage, and security.^{19,20} This setting constitutes a distributed global network in which the cloud is aware of each device's state (idle/busy, battery life, and so on) and resources (such as memory and computing power) as well as the network topology in different geospatial regions. This environment presents several research challenges, some of which have been addressed in the context of traditional distributed computing and others that require new solutions. Examples of the latter include federated identity limitations on mobile platforms and discovering and composing services offered by smart devices (for example, sensing services).^{19,21}

Recently, a new generation of smart devices is emerging with extensive computing power and memory—for example, the iPad or the newly

introduced, inexpensive, seven-inch Google Nexus tablet, which has a quad-core Tegra 3 processor, 1 Gbyte RAM, 16 Gbytes internal storage, and several sensors, including a camera, microphone, accelerometer, GPS, magnetometer, and gyroscope. Such devices' powerful computing and memory extend their use beyond sensing to running computing tasks, especially when combined with the cloud. For instance, we can use these mobile devices for MapReduce jobs in which the cloud provides the middleware for scheduling, coordination, and job migration (in case the device becomes unavailable due to user activity or network unavailability). In this environment, the service discovery and composition smart devices offer, along with identity management, present significant challenges.

The use of networked mobile computing devices for a multitude of activities among the general population and government has just begun. We have only an early appreciation of its range of uses and impacts over broad, socially important application categories, in particular, human assistance, and disaster information and relief. In parallel, we're coming to understand some of the associated research issues such as privacy and event-detection and characterization. Addressing these issues becomes especially challenging given that solutions must consider the right balance between stakeholders' requirements and policies on one hand, and solutions' utility in terms of quality, timeliness, and cost on the other. We expect spatial computing, in combination with social media, to be an active area of research as well as a commercially important field over the coming years. ■

References

1. "IBM, Caltrans, and UC Berkeley Aim to Help Commuters Avoid Congested Roadways Before their Trip Begins," press release, IBM, 13 Apr. 2011; <http://www-03.ibm.com/press/us/en/pressrelease/34261.wss>.
2. M. Youssef, V. Atluri, and N. Adam, "Preserving Mobile Customer Privacy: An Access Control System for Moving Objects and Customer Profiles," *Proc. 6th Int'l Conf. Mobile Data Management (MDM 05)*, ACM, 2005, pp. 67–76.
3. "More Americans Using Social Media and Technology in Emergencies," Public Affairs Desk, Am. Red Cross, 24 Aug. 2011; <http://tinyurl.com/9bylh55>.
4. N. Adam et al., "Social Media Alert and Response to Threats to Citizens (SMART-C)," *Proc. 8th IEEE Int'l Conf. Collaborative Computing: Networking, Applications, and Work-sharing (ColCom 12)*, to be published, 2012.
5. B. Zhou, J. Pei, and W. Luk, "A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Social Network Data," *SIGKDD Explorations Newsletter*, vol. 10, no. 2, 2008.
6. C. Aggarwal and P.S. Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," *Privacy-Preserving Data Mining*, vol. 34, Springer, 2008, pp. 11–52.
7. C. Dwork, "Differential Privacy," *Automata, Languages, and Programming*, LNCS 4052, M. Bugliesi et al., eds., Springer, 2006, pp. 1–12.
8. C. Wang et al., "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," *IEEE Conf. Computer Communications (INFOCOM 10)*, IEEE, 2010, pp. 525–533.
9. X. Zhang et al., "Formal Model and Policy Specification of Usage Control," *ACM Trans. Information and Systems Security*, vol. 8, no. 4, 2005, pp. 351–387.

IEEE Computer Society Publications Office

10662 Los Vaqueros Circle, PO Box 3014
Los Alamitos, CA 90720-1314

Lead Editor
Brian Kirk
bkirk@computer.org

Content Editor
Jake Widman

Manager, Editorial Services
Jenny Stout

Publications Coordinator
isystems@computer.org

Director, Products & Services
Evan Butterfield

Senior Manager, Editorial Services
Lars Jentsch


Digital Library Marketing Manager
Georgann Carter

Senior Business Development Manager
Sandra Brown

Senior Advertising Coordinator
Marian Anderson
manderson@computer.org

Submissions: For detailed instructions and formatting, see the author guidelines at www.computer.org/intelligent/author.htm or log onto *IEEE Intelligent Systems'* author center at Manuscript Central (www.computer.org/mc/intelligent/author.htm). Visit www.computer.org/intelligent for editorial guidelines.

Editorial: Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *IEEE Intelligent Systems* does not necessarily constitute endorsement by the IEEE or the IEEE Computer Society. All submissions are subject to editing for style, clarity, and length.

10. J. Wickramasuriya et al., "Privacy Protecting Data Collection in Media Spaces," *Proc. 12th Ann. ACM Int'l Conf. Multimedia (MULTIMEDIA 04)*, ACM, 2004, pp. 48–55.
 11. P. Golle, F. McSherry, and I. Mironov, "Data Collection with Self-Enforcing Privacy," *Proc. 13th ACM Conf. Computer and Communications Security (CCS 06)*, ACM, 2006, pp. 69–78.
 12. N. Li et al., "Privacy Preservation in Wireless Sensor Networks: A State-of-the-Art Survey," *Ad Hoc Networks*, vol. 7, no. 8, Nov. 2009, pp. 1501–1514.
 13. R. Chen et al., "Differentially Private Transit Data Publication: A Case Study on the Montreal Transportation System," *Proc. 18th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 12)*, ACM, 2012, pp. 213–221.
 14. H. Shin, J. Vaidya, and V. Atluri, "A Profile Anonymization Model for Location-Based Services," *J. Computer Security*, vol. 19, no. 5, 2011, pp. 795–833.
 15. B. Hoh et al., "Preserving Privacy in GPS Traces via Uncertainty-Aware Path Cloaking," *Proc. 14th ACM Conf. Computer and Communications Security (CCS 07)*, ACM, 2007, pp. 161–171.
 16. M. Jadhwal, I. Bilogrevic, and J.-P. Hubaux, "Optimizing Mixing in Pervasive Networks: A Graph-Theoretic Perspective," *Proc. 16th European Symp. Research in Computer Security (ESORICS 11)*, LNCS 6879, Springer, 2011, pp. 548–567.
 17. "From GPS and Virtual Globes to Spatial Computing—2020: The Next Transformative Technology," US Nat'l Science Foundation/Computing Community Consortium workshop proposal, Sept. 2012.
 18. N. Adam et al., "Approach for Discovering and Handling Crisis in a Service-Oriented Environment," *Proc. Intelligence and Security Informatics (ISI 2007)*, IEEE, 2007, pp. 16–24.
 19. J. Christensen, "Using RESTful Web Services and Cloud Computing to Create Next-Generation Mobile Applications," *Companion to the 24th Ann. ACM SIGPLAN Conf. Object-Oriented Programming, Systems, Languages, and Applications*, ACM, 2009, pp. 627–634.
 20. H. Artail, K. Fawaz, and A. Ghandour, "A Proxy-Based Architecture for Dynamic Discovery and Invocation of Web Services from Mobile Devices," *IEEE Trans. Services Computing*, vol. 5, no. 1, 2012, pp. 99–115.
 21. J. García-Macías et al., "Browsing the Internet of Things with Sentient Visors," *Computer*, vol. 44, no. 5, 2011, pp. 46–52.
-
- Nabil R. Adam** is a Professor II (Distinguished Professor) of Computers and Information Systems at Rutgers University, and the founding director of the Rutgers University Center for Information Management, Integration, and Connectivity (CIMIC). He's currently on loan as a fellow to the US Department of Homeland Security, Science & Technology Directorate, where he serves as a senior program manager and a branch chief. Contact him at adam@adam.rutgers.edu.
-
- Basit Shafiq** is a research assistant professor in the Center for Information Management, Integration, and Connectivity (CIMIC) at Rutgers University. Contact him at basit@cimic.rutgers.edu.
-
- Robin Staffin** is the director for basic research in the US Department of Defense, Office of Research and Engineering. Contact him at robin.staffin@osd.mil.
-
-  *Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.*