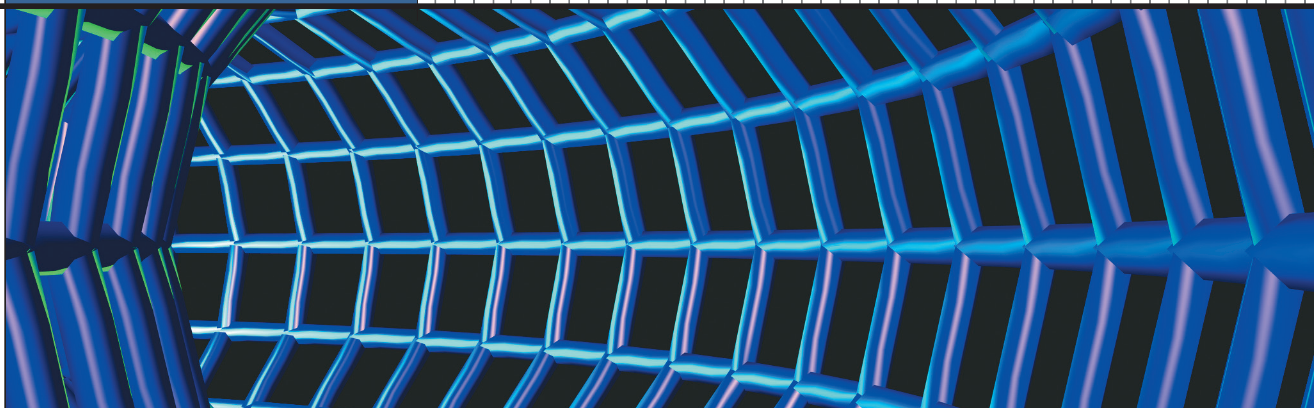

The Emergence of Social and Community Intelligence

Daqing Zhang and Bin Guo, *Institut Telecom SudParis, France*
Zhiwen Yu, *Northwestern Polytechnical University, P.R. China*

Social and community intelligence research aims to reveal individual and group behaviors, social interactions, and community dynamics by mining the digital traces that people leave while interacting with Web applications, static infrastructure, and mobile and wearable devices.

The past decade has seen a phenomenal growth of Internet and social network services, an explosion of sensor-equipped mobile phones, broader use of the Global Positioning System (GPS) in all types of public transportation, and the extensive deployment of sensor networks in facilities and outdoor environments. All these developments have led to an unprecedented accumulation of digital footprints—the digital traces that people leave while interacting with cyberphysical spaces.

Social and community intelligence (SCI) is an emerging research field that leverages the capacity to collect and analyze these footprints to reveal human behavior patterns and community dynamics. The breadth, depth, and scale of multimodal, mixed data sources provide an opportunity to compile digital footprints into a comprehensive picture of an individual's daily-life facets, transform the understanding of how people live and how organizations and societies function, and enable innovative services in human health, public safety, city resource management, environmental monitoring, and transportation management.

To understand SCI's potential, consider the activities on a typical university campus. Students often need to spontaneously locate sports partners or study space. They want instant answers to queries, such as when the next bus will reach the stop closest to the library or who is at that stop. Quick identification is a luxury in such cases, but if a pandemic like H1N1 occurs, it becomes crucial. Health organizations must quickly identify whom a suspected pandemic carrier has contacted and when and where contact has taken place. It is still difficult to answer questions about individual activities, group interaction, and society dynamics using current technology.

An SCI system can make such information available by analyzing pervasive data streams collected from personal mobile phone sensors, GPS devices on buses, WLAN or Bluetooth gateways inside a building, and Internet applications such as online social networks. In the pandemic use case, SCI data could provide distance and contact time with the suspected carrier, logical places for the encounters (office or bus), and the carrier's personal and business

0018-9162/11/$26.00 © 2011 IEEE

EVOLUTION OF SOCIAL AND COMMUNITY INTELLIGENCE RESEARCH

The understanding of human behavior, social interactions, and city dynamics has long relied on data collected through individual observations and surveys. Unfortunately, observations were usually sparse, and survey results were often incomplete and significantly delayed.

With advances in computing, storage, Internet access, wireless communication, and sensing, it is now possible to monitor and analyze human behavior, social interactions, and city dynamics on a large scale and in nearly real time. Initially, analysts used Internet content as the premier data source for understanding large-scale human interaction. Then the emergence of static sensing infrastructure made it possible to recognize human activities in a physical environment. Recently, the prevalence of sensor-enriched mobile devices has brought unprecedented opportunities to observe human behavior, social interaction, and community dynamics. The Internet and Web, static infrastructure, and wearable and mobile devices have all contributed to the evolution of SCI research.

Internet services and Web applications

The past two decades have witnessed the explosive growth of Internet services, such as e-mail, instant messaging, and Web applications, which have changed how people share and obtain information and communicate with each other. A large body of work has centered on leveraging those services, including efforts in information extraction and human interaction analysis, such as news recommendation, personal and organizational profile extraction, and e-mail network analysis. As the Internet moves into the Web 2.0 era, researchers are turning their attention to online social utilities, such as social networking sites, wikis, and blogs.

Much work has focused on social behavior study and user-generated content analysis. A group from the University of Koblenz-Landau has investigated how to mine social networks to study customer behavior.[1] Researchers from Purdue University have developed an unsupervised model to estimate relationship strength from interaction activity and user similarity on a social website.[2] Investigators from Wright State University label Web 2.0 service users as "citizen sensors" and have worked on social event detection from user-contributed contents.[3] Collaborators from the University of Arizona, Carnegie Mellon University, and the University of Southern California coined the term "social computing," defining it as social study based on the Internet and Web that aims to study and extract human social dynamics from online human interactions.[4]

Static sensing infrastructure

With the prevalence of static sensing infrastructure, such as surveillance cameras, environmental sensors, indoor positioning sensors, and radio-frequency identification (RFID), monitoring and detecting real-world events has become feasible. Early sensor applications involved mainly environmental monitoring in significant places. Surveillance cameras were the first sensing devices widely deployed in public and critical spots to detect abnormal events. Temperature, light, and humidity sensors are also widely used for environmental monitoring, for example, to detect a forest fire. With advances in sensing techniques, it is now possible to deploy massive numbers of cheap, tiny sensors, such as RFIDs and switches, to augment living and working environments—creating smart spaces. Active Bats uses ultrasonic sensors and triangulation to locate indoor objects,[5] which in turn enables location-based services like finding a lost key or other objects. Researchers from Intel Seattle are exploring techniques to recognize human activities by analyzing people's interaction with RFID-equipped everyday objects.[6]

Mobile sensing

Although static sensing infrastructure brings opportunities to infer environmental and human contexts in smart spaces,[7] it is tied to a particular physical environment. Wearable sensors, in contrast, transform people into mobile sensors for both personal and ambient environment monitoring. People can wear sensors, such as accelerometers, pedometers, heart-rate sensors, wireless webcams, and microphones, on different parts of their body to enable various human-centered services, including human behavior detection, health-status monitoring, and social-context recognition.

Although wearable sensors are portable and promising, people still do not view them as a personal companion. In contrast, smartphones—sensor-enhanced mobile phones with embedded GPS receivers, Bluetooth/WiFi, accelerometers, and cameras—always accompany users and are thus a rich information source.

The volumes of multimodal data collected from people's daily use of smartphones opens a new window to study large-scale human behavior patterns and community dynamics. For example, Real Time Rome (http://senseable.mit.edu/realtimerome), a project that the Massachusetts Institute of Technology initiated in 2006, is one of the pioneering projects that explicitly use mobile phone data to understand city dynamics, such as people's movement patterns and the spatial and social use of streets and neighborhoods. Reality mining (http://reality.media.mit.edu), on the other hand, collects and analyzes mobile phone data such as physical proximity to identify predictable patterns of social behavior, such as friendship.[8] Dartmouth's Mobile Sensing Group is looking at the use of human-centric sensing to link personal mobile sensing to mobile social networks and public environment monitoring.

References

1. S. Staab et al., "Social Networks Applied," *IEEE Intelligent Systems*, vol. 20, no. 1, 2005, pp. 80-93.
2. R. Xiang et al., "Modeling Relationship Strength in Online Social Networks," *Proc. 19th Int'l World Wide Web Conf.* (WWW 10), ACM Press, 2010, pp. 981-990.
3. A. Sheth et al., "Citizen Sensing, Social Signals, and Enriching Human Experience," *IEEE Internet Computing*, vol. 13, no. 4, 2009, pp. 87-92.
4. F. Wang et al., "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems*, vol. 20, no. 2, 2005, pp. 79-83.
5. H. Andy et al., "The Anatomy of a Context-Aware Application," *Wireless Networks*, vol. 8, no. 2-3, 2002, pp. 187-197.
6. M. Philipose et al., "Inferring ADLs from Interactions with Objects," *IEEE Pervasive Computing*, vol. 3, no. 3, 2004, pp. 50-57.
7. X. Wang et al., "Semantic Space: An Infrastructure for Smart Spaces," *IEEE Pervasive Computing*, vol. 3, no. 3, 2004, pp. 32-39.
8. N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure using Mobile Phone Data," *Proc. Nat'l Academy of Sciences* (PNAS 07), vol. 106, no. 36, Nat'l Academy of Sciences, 2007, pp. 15274-15278.

Table 1. Goals of SCI and four related research areas.

Research area	Goals
SCI	Reveal individual and group behavior, social interactions, and community dynamics, leveraging the aggregated power of three information sources: Internet and Web, static infrastructure, and mobile devices and wearable sensors
Social computing	Conduct computational social studies, analyze human interactions, and design technologies that consider social context
Reality mining	Collect and analyze mobile sensing data related to human social behavior to characterize human interaction and behavior patterns
Urban computing	Study the interaction between humans and environments using technology in public areas, such as cities, suburbs, parks, and forests
Human-centric sensing	Use mobile sensing data to derive people's daily patterns and interactions and identify characteristics of public environments

relationships—all important clues and contexts that affect the probability of disease spread.

SCI evolved from closely related research areas such as social computing, reality mining, and urban computing. However, although it embodies aspects of these areas, it has unique infrastructure, data, technology, and application needs. Unlike research areas, such as social and urban computing, SCI mines data from three sources: Internet services and Web applications, static infrastructure, and wearable sensors and mobile devices.

Many SCI applications are on the horizon, necessitating a general system framework that can both accommodate heterogeneous devices, software, and spaces and support rapid application development. We have developed such a framework on the basis of our extensive SCI application survey.

CHARACTERISTICS

SCI system scale goes beyond a single smart space to the community level. Real-life, real-time data collection and inference are key system features. SCI thus requires an infrastructure that can integrate large-scale and heterogeneous information sources and systematically support rapid application development, deployment, and evaluation.

The three main SCI data sources are multimodal and heterogeneous, including

- social network and Internet interaction services, which provide data about the individual's preferences and relationships;
- infrastructure-bound sensor data about the environment; and
- mobile and wearable sensor data about the individual and moving objects.

Although each source can independently show one facet of the user's daily life, combined sources can reveal unforeseen social and community behavior.

The core SCI technologies are data mining, machine learning, and artificial intelligence. The objective of data processing and inference ranges from recognizing the individual's physical activity and environmental context to extracting higher-level community and social behavior. Semantic gaps exist between individual activities and social and community behavior, and bridging these gaps is a key challenge for SCI research.

SCI applications aim to enable innovative services at the society level, such as community healthcare, public safety, and city resource and transportation management.

Comparison with existing research

The "Evolution of Social and Community Intelligence Research" sidebar describes how research developments have led to the birth of SCI research. Table 1 lists the goals of SCI and four closely related research areas. SCI differs from the other research areas listed in Table 1 primarily because it explores the fusion of *three* data sources, not just one source, to infer intelligence at the group and community level. Intelligence can range from human interaction to group behavior within a community to the dynamics of an entire community.

Compared to SCI, *social computing* emphasizes the analysis of human interaction and social behaviors using only Web data. It does not target the study of a large-scale physical community. Similar to social computing, *reality mining* focuses on social-interaction analysis, but relies primarily on data gathered from mobile devices. Like SCI, *urban computing* studies the relationship between individual and environment at the city scale, but SCI extends its scope from urban design to the large-scale analysis of personal, group, and community dynamics. *Human-centric sensing* is the research area closest to SCI. The two areas have similar research goals, but the underlying sensing mechanisms are different. While human-centric sensing uses only mobile phones, SCI aggregates the information from mobile phones, Internet services, and static infrastructure.

As this brief comparison shows, SCI has many aspects in common with the four research areas in Table 1, but goes beyond them in scope and data sources. Breakthroughs

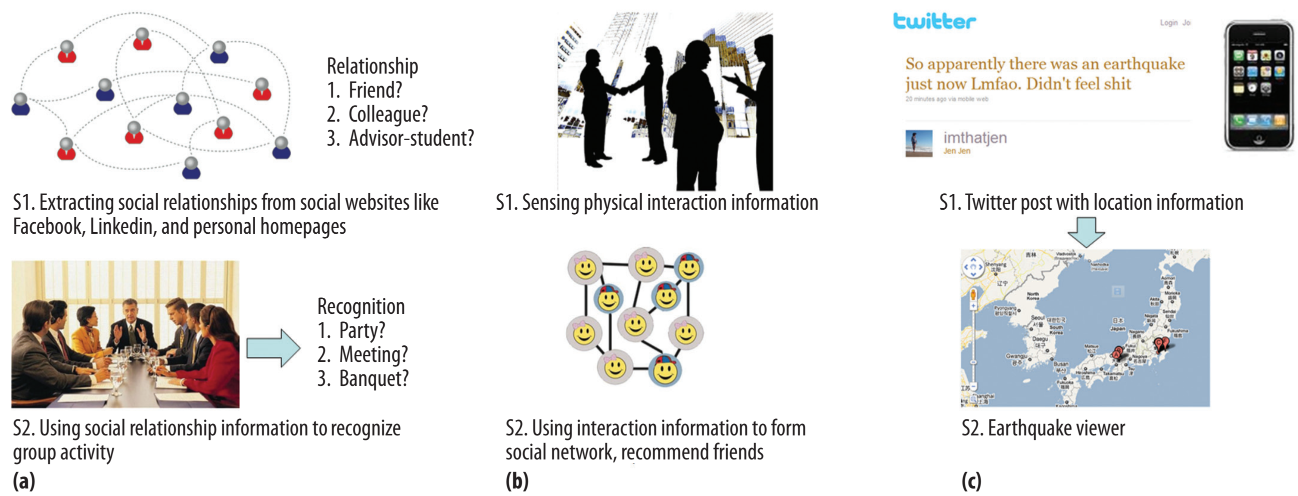

Figure 1. Examples of data-source aggregation in SCI. (a) Analysts use Web-mined knowledge to enhance sensor-based activity recognition. (b) Sensor-detected human interaction enhances an online social network. (c) Analysts merge mobile sensing and Web data to characterize a situation more accurately.

in any of the four areas will contribute to progress in SCI research.

Aggregated sources

The three SCI data sources have different attributes and strengths that affect how analysts can fuse them to extract information:

- The Internet and Web are the best sources for extracting static or slowly changing information, such as user profiles, organization structures, and user relationships in a community.
- Static infrastructure enables the detection of indoor and urban user activities, group activities, and spatial context in sensor-enriched environments.
- Mobile devices and wearable sensors are always user-centric and are thus appropriate for sensing individual activities, interpersonal interactions, significant user locations, and public environment contexts.

The three examples in Figure 1 showcase the aggregated effects among the three data sources. Many more are possible.

Figure 1a shows the power of combining Web knowledge mining and sensor-based activity recognition. Social relationships extracted from the Web can assist social-activity recognition in the physical world. For example, if the detected social gathering is in the evening and all participants are friends, the event is likely to be a party. If the gathering occurs on a weekday morning and participants are managers and subordinates, it is more likely to be a business meeting.

Figure 1b shows how sensor-detected human interaction can enhance an online social network. Online social networks still rely on user input to infer social relationships. However, because users input only partial information about themselves and their friends, the predicted social connection is often inaccurate. By tracking real-world user interactions through sensors and then mapping the detected relationship onto the online social network, analysts can significantly improve the quality of social-network services. For example, if two people are spending time together after work, they are probably close friends. If they meet only at work, they are likely to be merely colleagues. In the Serendipity project,[1] researchers used Bluetooth-enabled mobile phones to scan other devices in the user's proximity—information they then used to verify and better characterize relationships in an online social-network system.

Figure 1c shows one possible result of merging mobile sensing and Web data. Because data from a source often characterizes a specific facet, fusing distinct data sources can often draw a better picture of the entire situation. For example, by integrating the mined theme from user posts and the revealed location information from GPS-equipped mobile phones, Twitter was able to support the near real-time reporting of earthquakes in Japan.[2]

GENERAL SYSTEM FRAMEWORK

Because it is effectively a community-wide sensing system, SCI infrastructure requires a general framework that integrates large-scale and heterogeneous information sources and systematically supports rapid application development, deployment, and evaluation. On the basis of our investigations into SCI, we have developed the five-layer general framework in Figure 2.

The *pervasive sensing layer* manages the three major information sources. Because privacy is a major concern for both personal and organizational data sharing, the proposed framework also incorporates a *data anonymization layer* before data release and processing. The *hybrid learning layer* applies diverse machine-learning and data-mining techniques to convert low-level single-modality sensing data into high-level features. The goal is to mine the frequent data patterns to derive the individual's behavior and single space context before extracting the complete SCI. The *semantic inference layer* uses logic-based inferences to accommodate feature aggregation. It complements the statistical learning approach and uses explicit rules to effectively associate the hybrid learning layer with the expected SCI on the basis of expert domain knowledge. Finally, the *application layer* includes a variety of potential SCI-enabled services.

APPLICATIONS

SCI applications stem from the need to develop socially aware services that facilitate group interaction and communication, monitor the real-time change of the physical world for the public good, and track and predict specific events to benefit society. Many application areas are possible, but we have chosen six primary ones.

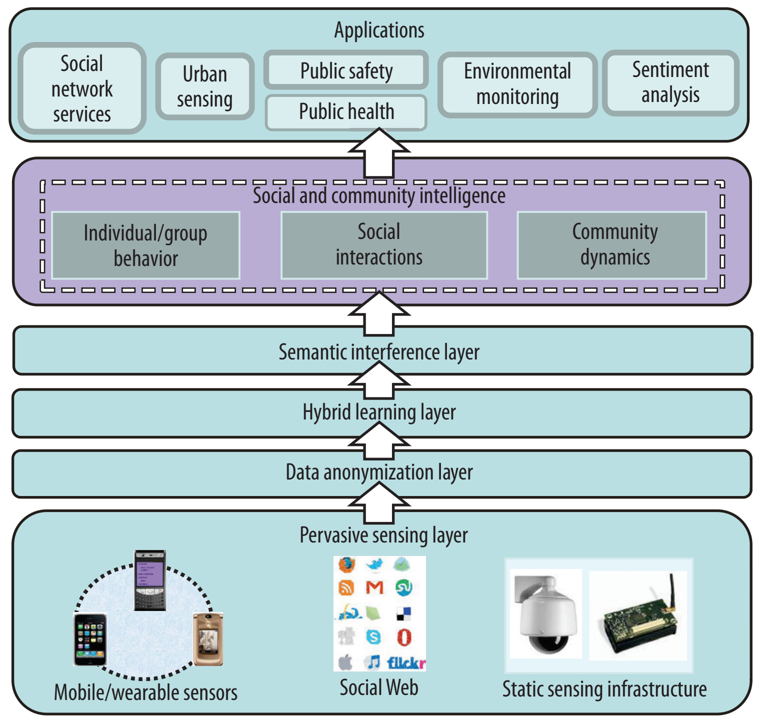

Figure 2. A general SCI system framework. The proposed framework integrates large-scale and heterogeneous information sources and systematically supports rapid application development, deployment, and evaluation.

Although most of these applications use only one or two data source types, we believe that there are ways to enhance them or even build new applications by incorporating increasingly heterogeneous data sources. For example, both mobile devices and infrastructure can help improve applications in urban planning, environmental monitoring, well-being management, and public safety.

Social network services

By recording various aspects of physical interaction and communication, such as colocation, conversations, and call logs, and by mining user behavior patterns, such as places of interest, SCI nurtures the development of many social-network services, such as friend recommendation and augmented online interaction. The FriendSensing application[3] can recommend friends by monitoring a user's encounters and mobile phone activity, such as texting and calling. The CenceMe project (www.cenceme.org) exploits off-the-shelf smartphones to automatically infer people's presence, whether they are walking on the street or dancing at a party with friends, and then shares this presence through social-network portals such as Facebook and Twitter.

In the EU FP7 Societies project (www.ict-societies.eu), we plan to support the creation and management of different social communities in pervasive computing environments. A community has several forms. It can be people colocated in a physical space, defined through an environment-sensing infrastructure. It can be a group with common interests and expertise, defined through information extraction from a homepage or social website. Or it can be a group whose members have followed a similar routine, defined by analyzing traces from wearable or mobile sensors. Social communities not only have different forms and goals but also can be highly dynamic. The more information we can obtain from different data sources about people, the better we can support and manage social communities.

Urban sensing

With wireless sensor platforms in the hands of the masses, it is possible to leverage community sensing to address urban-scale problems, such as ambient monitoring, traffic planning, and the better use of public utilities.

MIT's Real Time Rome project (see sidebar) uses aggregated data from cell phones, buses, and taxis in Rome to better understand urban dynamics in real time. The Biketastic project (http://biketastic.com) improves bike commuting in Los Angeles by collecting and mining data that bikers have contributed through their mobile phones. Bikers can then plan routes with the lowest probability of traffic accidents and the best air quality. The GeoLife project extracts information about interesting locations and travel sequences on the basis of users' GPS trajectories and provides travel recommendations to the city's first-time visitors.[4]

Environmental monitoring

The nomadic, participatory, and in situ nature of community sensing provides new opportunities for environmental monitoring and natural-resource protection.

One area is nature preservation. With the help of human volunteers, the Great Backyard Bird Count project reports the cumulative counts of birdwatchers from across America (www.birdsource.org/gbbc). The MIT Owl project (http://web.mit.edu/newsoffice/2008/tracking-0822.html) leverages a network of smartphones equipped with GPS, compasses, and directional microphones to reduce the burden of manually assessing owl populations.

Pollution measurement is another area ripe for environmental monitoring. Several projects have used portable pollution-sensing devices in various missions. The BikeNet application assesses metrics to give a holistic picture of the cyclist's experience, including the carbon dioxide level along the path. It facilitates public sensing and sharing by letting multiple users merge their individual data, for example, to create pollution and noise maps of their city.[5] The Personal Environmental Impact Report project (http://urban.cens.ucla.edu/projects/peir) uses GPS-enabled phones to detect if a user is driving, riding, or walking. The information becomes the basis for assessing an individual's environmental impact, such as the carbon footprint from the mode of transportation, and exposure to air pollution.

Public health

SCI can make it easier to anticipate and track a disease outbreak. Epidemics of seasonal influenza are a major public health concern, causing tens of thousands of deaths worldwide annually. Early detection is key to reducing this count. Google researchers have shown that, by mining indirect signals from millions of geographically localized health-related search queries, it is possible to estimate the level of influenza-like illnesses in US regions with a reporting lag of just one day.[6] This lag is much smaller than the government agency estimates of regional data, which are published weekly on the basis of virology and clinical statistics.

SCI also brings new opportunities for managing personal well-being. With community sensing, people can log their physical activities, track their food intake, sense their mental status in real time, and record their daily social interactions—all of which is information that is useful in improving their health management. The Neat-o-Games system, for example, uses a wearable accelerometer to detect if the user is walking or running and motivates users to do more exercises by showing avatars in a virtual community race game.[7]

Sentiment analysis

Sensing user sentiments is important in context-aware computing, but it is not easy to use physical sensors for this purpose. One way around the problem is to collect or mine user-generated Web data. For example, Emotional City (www.emotionalcities.com) and D-Tower (www.d-toren.nl) collect information about citizens' moods through daily Web surveys and display their emotions by changing the colors of a building or public sculpture.

Public safety

Public safety involves the prevention of and protection from events that could endanger the public, such as crimes or disasters. Public video surveillance systems have greatly enhanced citywide event sensing and safety monitoring. For example, the Boston police department has recently embraced collecting user-contributed sensor data to assist in crime prevention.

RESEARCH ISSUES

SCI applications directly motivate many research issues, which are aligned with the functional layers in our SCI system framework: sensing, data anonymization, data processing, social-context learning, and intelligence extraction.

Participatory or opportunistic sensing?

The first research issue to be considered is what roles people should play in community sensing. For example, when a mobile phone is acting as a sensing device, should the sensing system interrupt the mobile phone user to accept or stop the sensing task? There are two extreme cases for sensing.

In *participatory* sensing, people are part of the sensing system's decision-making process. They decide which application request to accept, what data to share, and to what extent they will allow privacy mechanisms to impact data fidelity. In other words, users retain control over their raw data. The Personal Data Vault system is based on this idea, which seeks to provide easy-to-use toolkits to support data control.[8]

Opportunistic sensing, in contrast, automatically determines when to use devices to meet the application's sensing requests. Instead of requiring human intervention to actively and consciously participate in the sensing, opportunistic sensing requests that a sensing device be used automatically whenever its state (location, user activity, and so on) matches an application's requirements.

Obviously, there's a tradeoff between participatory and opportunistic sensing. Participatory sensing places demands on user involvement, which restricts the pool of willing participants, and people's tolerance of interruptions limits the number of applications. Opportunistic sensing risks leaking personally sensitive information and requires more resources for decision making, such as a determination of the sampling context (indicates when sampling should be started and stopped). As such, an opportunistic system must adapt to the device's changing resource availability.

Future work should focus on how to balance users' involvement and proper control while integrating the appropriate protection mechanisms for data privacy.

Privacy, data quality, and trust

Sharing and revealing personal digital data could pose privacy risks for users. Even data gathered in a community can reveal considerable information about an individual or organization's behavior. For example, a person's location might reveal her private interests, while an organization's health data might suggest potential environmental problems for the staff. The impact is obvious: if there is no way to anonymize the data and place it under the data owner's control, people might be less likely to share their data.

Privacy. Privacy protection involves many elements, including identity (who is asking for the data?), granularity (how much does the data reveal about people or the user's identity?), and time (how long will the data be retained?). Data anonymization and user control are two research areas that address these questions.

The objective of data anonymization is to avoid revealing users' identities when they contribute their data. MetroSense uses the k-anonymity method when users contribute location data to a server. The method generalizes a user's position to a region containing at least k users, thereby hiding that user's identity.[5]

Another promising approach to secure multiparty computation allows data mining from many organizations without ever aggregating the data into a central data repository. Each organization performs part of the computation on the basis of its privately held data and uses cryptography to encode intermediate results that it must then communicate to other organizations performing other parts of the computation.[9]

Other privacy-preserving methods include sharing only statistical summaries of the individual datasets and inserting random perturbations into individual data records before sharing them.

User control is critical to personal data sharing because it ensures that users reveal only the information they want to reveal and that the system reveals only what the users want it to reveal. For example, a user might track his heart rate each day, but there is no reason to share that information with anyone but his doctor. User control research is exploiting methods that enable users to manage their data by tailoring access-control and data-management tools.[8]

Data quality. Web data quality can range from authorized to fake. The same is true of mobile phone data quality. For example, some people put their mobile phones in their pocket; others put it in a purse. If both users are walking, the data from those phones will be quite different. Rather than take the data directly from the device, it might be better to train classifiers that work in different contexts. Both data collection and context identification remain challenging issues, however.

Trust. Mining social and community behavior often requires collecting data from anonymous participants. If no mechanism ensures that the source is valid and the information is therefore accurate, the data will not be trustworthy. Twitter data is sometimes unreliable; mobile phone users can send incorrect or even faked data to the data center. Future work should look at developing trust and abnormal data-detection methods to ensure the trustworthiness and quality of collected data.

Managing large-scale heterogeneous data sources

In an SCI system, data producers can differ significantly in modality (mobile phones, fixed cameras, or Web services), Internet connectivity (constant or intermittent), sharing willingness or privacy sensitivity, and resource capabilities for processing data locally. Information consumers are also heterogeneous in terms of running environments and data needs. These myriad dimensions of heterogeneity pose hard challenges for data management.

Multimodal data. Different sensor types have different attributes and capabilities, such as varying accuracy in sensing the physical and virtual world. Integrating information from diverse data sources compounds the job of SCI mining. Raw data from different sensor sources must be transformed to the same metrics and represented by a shared ontology to facilitate the learning and inference process.

Temporal and continuous data. Because sensing data is time sequenced, when modeling individual and group behavior, the system should consider multiple data stream samples, rather than what each sensor reads in isolation. In addition, real-world systems are continuous, so it's important to build models that cater to the discrete, sampled sensor state.

Large-scale data processing. SCI applications often deal with real-time data collected from many sensing nodes, such as the computing and visualization of traffic conditions in a city. As such, they can suffer from the same modeling and computational difficulties inherent in most data-mining tasks. More work is needed on sampling optimization, problem decomposition, and the adoption of advanced computational and learning models within a particular problem domain.

Inconsistency. The same sensor might sense an event under different conditions, such as sensing a person's voice in a quiet office or noisy restaurant, which can yield conflicting inference results. Because of these environmental differences, a group of colocated sensors running the same classification algorithm and sensing the same event in time and space could compute different inference results, which leads to system inconsistency. Dartmouth's Mobile Sensing group proposed a collaborative approach to deal with this inconsistency,[10] but more solutions are needed.

Difficulty in labeling data. Labeling large amounts of data is often difficult and time-consuming. Future work should focus on learning algorithms that can derive system models from relatively small amounts of labeled data.

Extracting high-level intelligence data from low-level sensing data

SCI aims to identify a set of characteristics or behaviors associated with a social community. Social communities form flexibly from people in the same organization, at the same places, with the same behaviors and interests, and so on, depending on social application requirements.[1] By pooling individual behavior traces and mining the underlying social patterns, an SCI system can extract various social or group behaviors.[9] The extracted social context can be an event such as an open concert, a behavior pattern in daily activity, a relationship within a group, or a significant location.

The thrust of SCI pattern mining is to identify user similarity in these social patterns to facilitate offering socially aware services. Unsupervised learning techniques, such as clustering, latent semantic analysis, and matrix factorization, are possible ways to mine social context according to individual behavioral similarities. The process includes the mining and discovery of common social contexts, such as personal characteristics, cuisine preferences, and eagerness to participate socially. It also includes the discovery of undefined social patterns for interest matching and ranking social choices.

To enable systems to infer social events on the basis of user context traces, data mining and inference research should aim to bridge the semantic gap between the low-level individual activities and high-level social events.

We believe that SCI represents a new interdisciplinary research and application field and that its scope will continue to expand with innovative applications in the near future. As an emerging research area, SCI still faces challenges, but their resolution will pave the way for new research opportunities. Although existing SCI practices involve only one data source type—Web applications and Internet services, static sensor infrastructure, or mobile and wearable devices—we expect to see the rapid growth of research on using the aggregated power of three information sources as well as on enabling innovative SCI-enabled applications. C

Acknowledgment

This research is partially supported by the European Commission FP7 Societies Project.

References

1. A. Pentland, "Socially Aware Computation and Communication," *Computer*, Mar. 2005, pp. 33-40.
2. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors," *Proc. 19th Int'l World Wide Web Conf.* (WWW 10), ACM Press, 2010, pp. 851-860.
3. D. Quercia et al., "Nurturing Social Networks Using Mobile Phones," *IEEE Pervasive Computing*, July 2010, pp. 12-20.
4. Y. Zheng et al., "GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory," *IEEE Data Eng. Bull.*, vol. 33, no. 2, 2010, pp. 32-40.
5. A.T. Campbell et al., "The Rise of People-Centric Sensing," *IEEE Internet Computing*, vol. 12, no. 4, 2008, pp. 12-21.
6. N.M. Ferguson et al., "Strategies for Mitigating an Influenza Pandemic," *Nature*, vol. 442, no. 7101, 2006, pp.448-452.
7. Y. Fujiki et al., "NEAT-o-Games: Blending Physical Activity and Fun in the Daily Routine," *ACM Computers in Entertainment*, vol. 6, no. 2, 2008, pp. 1-22.
8. K. Shilton, "Four Billion Little Brothers? Privacy, Mobile Phones, and Ubiquitous Data Collection," *Comm. ACM*, vol. 52, no. 11, 2009, pp. 48-53.
9. T.M. Mitchell, "Mining Our Reality," *Science*, vol. 326, no. 5960, 2009, pp. 1644-1645.
10. E. Miluzzo et al., "Darwin Phones: The Evolution of Sensing and Inference on Mobile Phones," *Proc. Mobile Systems Conf.* (MobiSys 10), ACM Press, 2010, pp. 5-20.

***Daqing Zhang** is a professor at Institute Telecom SudParis, France. His research interests include pervasive and human-centric sensing, context-aware systems, social and community intelligence, mobile social networks, and pervasive healthcare. He received a PhD from the University of Rome and University of L'Aquila, Italy. He is a member of IEEE and the ACM. Contact him at daqing.zhang@it-sudparis.eu.*

***Bin Guo** is a postdoctoral researcher at Institute Telecom SudParis, France. His research interests include ubiquitous computing, context-aware systems, and community sensing. He received a PhD in computer science from Keio University, Japan. He is a member of IEEE. Contact him at bin.guo@it-sudparis.eu.*

***Zhiwen Yu** is a professor at Northwestern Polytechnical University, P.R. China. His research interests include pervasive computing, context-aware systems, human-computer interaction, and personalization. He received a PhD in computer science and technology from the Northwestern Polytechnical University. He is a member of IEEE and the ACM. Contact him at zhiwenyu@nwpu.edu.cn.*