Extracting Social and Community Intelligence from Digital Footprints: An Emerging Research Area

Daqing Zhang[1], Bin Guo[1], Bin Li[1], Zhiwen Yu[2]

[1] Institut TELECOM SudParis, France
{daqing.zhang, bin.guo, bin.li}@it-sudparis.eu
[2] School of Computer Science, Northwestern Polytechnical University, P. R. China
zhiwenyu@nwpu.edu.cn

Abstract. As a result of the recent explosion of sensor-equipped mobile phone market, the phenomenal growth of Internet and social network users, and the large deployment of sensor network in public facilities, private buildings and outdoor environments, the "digital footprints" left by people while interacting with cyber-physical spaces are accumulating with an unprecedented breadth, depth and scale. The technology trend towards pervasive sensing and large-scale social and community computing is making "social and community intelligence (SCI)", a new research area take shape, that aims at mining the "digital footprints" to reveal the patterns of individual, group and societal behaviours. It is believed that the SCI technology has the potential to revolutionize the field of context-aware computing. The aim of this position paper is to identify this emerging research area, present the research background and some references to the relevant research fields, define the general system framework, predict some potential application areas, and propose some initial thoughts about the future research issues and challenges in social and community intelligence.

Keywords: social and community intelligence; digital footprints; pervasive sensing

1 Introduction

With the technological advances in sensing, computing, storage, communication and Internet, a lot of research areas have emerged such as sensor network, pervasive computing, Internet of Things, social network, to name just a few. From those emerging areas, there is a clear trend to augment the physical devices/objects with sensing, computing and communication capabilities, connect them together to form a network, and make use of the collective effect of networked things. As a result of the recent explosion of sensor-equipped mobile phone market, the phenomenal growth of Internet and social network users, and the large deployment of sensor network in public facilities, private buildings and outdoor environments, the digital traces left by people while interacting with cyber-physical spaces are accumulating at an unprecedented breadth, depth and scale, and we call all those traces left by people the "digital footprints". By 2009, four billions of mobile devices [1] were carried and used by people in the world everyday, which are recording individuals' digital traces using various built-in sensors and generating huge amount of "digital footprints" [2]. According to the 'World GPS Market Forecast to 2013', a 2010 market research report by RNCOS (http://www.rncos.com/), millions of cars and taxis are being equipped with GPS each year and the mobile location technologies market is expected to grow at a CAGR (Compound Annual Growth Rate) of about 20% to cross US$ 70 Billion by 2013, being another data source about the facets of individual, family and city. In addition, Internet services like e-mails, instant messaging, etc. and social networks like Facebook, MySpace, Twitter and LinkedIn record information about people's relationship and preferences; indoor and outdoor sensor network data provide more insights about people's environmental context.

Leveraging the capacity to collect and analyze the "digital footprints" at community scale, a new research field called "**social and community intelligence (SCI)**" is emerging that aims at revealing the patterns of individual, group and societal behaviours. The scale and heterogeneity of the multimodal, mixed data sources present us an opportunity to compile the digital footprints into a comprehensive picture of individual's daily life facets, radically change the way we build computational models of human behaviour, and enable completely innovative services in areas like human health, public safety, city resource management, environment monitoring, and transportation management.

Different from other closely related research areas such as sensor-based activity recognition, the unique characteristics of this new SCI research area can be embodied in the following aspects:

1. **Infrastructure**: The scale of the SCI system goes beyond single smart space and reaches the level of a community. Real-life, real-time data sensing and inference is a key system feature. An infrastructure is required to integrate large-scale and heterogeneous devices, software, and spaces, and provide systematic support for rapid application development, deployment, and evaluation.
2. **Data**: The data sources are multi-modal and heterogeneous. The social and community intelligence can be inferred from three main data sources: the mobile/wearable sensor data about the individual and moving space, the infrastructure-bound sensor data about the environment, and the social data about the individual's preference and relationship with others from social network and Internet interaction services. While each data source can independently shows one facet of the user's daily life, the combination of the three data sources can reveal unforeseen social behaviours.
3. **Technology**: The core technologies for SCI are data mining, machine learning and AI. And the objective of data processing and inference goes from recognizing the individual's physical activity and environmental context to extracting higher-level community and social behaviours (from talking to meeting; from driving slowly to traffic jam, there exist semantic gaps between individual activities and social/community behaviours).
4. **Application**: It aims to enable innovative services in society level like community healthcare, public safety, city resource management and transportation management.

Now let's use a simple use case in a university campus to illustrate the concrete ideas about social and community intelligence:

University campus is a typical high-density populated community (as shown in Fig. 1). Students often face the problem of finding partners doing sports in a certain free time slot, searching if there are free spaces available for exercises or study, etc. When a pandemic like H1N1 occurs, how to quickly identify who has been contacted by a suspect person, when and where the contact takes place is an important issue to avoid further spread of the disease. There are also queries like when will the next bus reaches the Bus Stop near Library, how many people are waiting in the bus stop, etc. In real-world environments, it is often difficult to answer these questions merely based on today's technologies. However, all those community services in university campus can be enabled by analyzing the pervasive data streams collected from personal mobile phone sensors, GPS from buses, WiFi or Bluetooth access points inside the building, social relationship from the web, etc. In the case of pandemic, for example, the distance and contact time with the suspect, the logical places for the meeting (e.g., office, bus), the relationship with the suspect (e.g., family, friend, colleague, unknown) are all important contexts affecting the probability of disease spread.

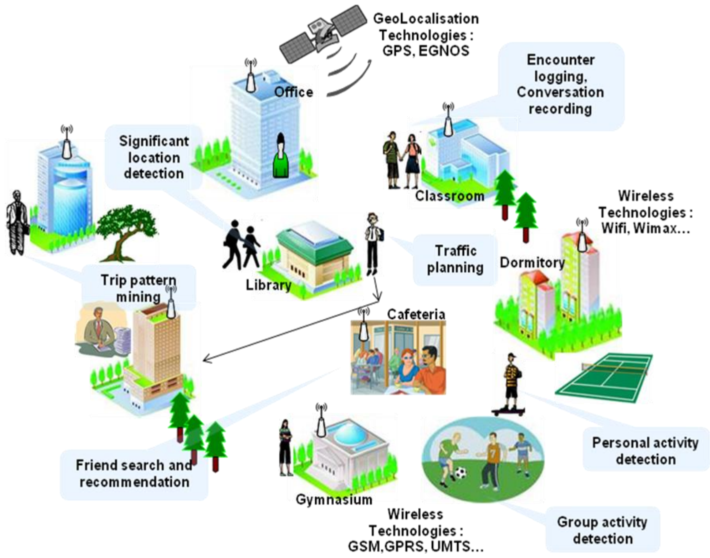

Fig. 1. Campus-scale community sensing and intelligence

The rest of this position paper is organized as follows. Section 2 presents the research background of social and community intelligence, followed by a general framework for SCI systems in Section 3. Then several potential applications of SCI and some initial thoughts about the open research issues in SCI are elaborated in

Section 4 and 5, respectively. Finally, we conclude this paper by proposing some promising research directions.

2 Research Background

Research on social and community intelligence is at its early stage. However, as the result of convergence of several research disciplines such as sensor network, ubiquitous computing, mobile computing, machine learning, data mining, and social science, SCI has its deep roots in three recent fast-growing research fields according to the origin of data sources: 1) mobile/wearable sensor-based activity recognition, 2) context inference in smart spaces, and 3) social network analysis. While each of the three abovementioned research fields is an active, multi-disciplinary area itself with rich research challenges/applications, the "convergence" of these three fields is expected to be influenced by the advances of each one, and would present new challenges and opportunities as a result of "network effects". In the rest of this section, each of these three areas will be briefly introduced with some references.

2.1 Mobile/Wearable Sensor-Based Activity Recognition

The mobile/wearable sensor-based activity recognition research leverages the prevalence of wearable sensors and mobile sensors embedded into the mobile devices that are accompanying the users most of time, it aims at collecting the sensing data in the real life and predicting the daily activities of users at real-time. RFID (radio frequency identification), GPS (global position system), accelerometer are among the most popular sensors embedded in the mobile devices. Sensor-based activity recognition can be roughly divided into two categories based on where the sensors are deployed: human body or object. Wearable sensors attached to a human body can generate various signals when the user performs activities, which is effective to detect human physical movements, such as walking, running, scrubbing, and exercising. Object-based activity recognition is based on real-world observations that activities are characterised by the objects that are manipulated during their operation. Activities involving complex physical motions and complex interactions with the environment, e.g., grooming, cooking, phoning, toileting, washing hands, and so forth, can be recognised through this approach.

The key idea behind the sensor-based activity recognition is to build or learn a mathematical model of activity based on a series of observations which are represented by the sensor readings, then by feeding the real-time sensor readings to the model, the human activities are predicted. Take a well known work of RFID-based activity recognition for example [3]. Consider a household where each object (e.g., cups, spoons, and toothbrushes) is tagged with an RFID, if the subject wears a watch-like RFID reader on her wrist and performs her daily activities, then each object she touched during her performance can be tracked in real-time. By recording the sequence of the touched objects, machine learning and inference methods can then be applied to learn a model for recognizing daily activities ranging from simple ones like brushing teeth to critical ones like taking medication and cooking a meal (safety-related), etc. Most of the early work on sensor-based activity recognition was motivated by applications in elderly care [4], healthcare [5]. Some of them have also been applied to habitat monitoring with sensor networks [6], and tracking human interaction in offices [7].

Another line of research in sensor-based activity recognition is location-based activity recognition as a result of wide deployment GPS sensors. The earliest work in this line intended to detect the trip plan derivation based on GPS traces [8]. In the following years, a large body of work was reported, including significant location identification [9], transportation mode recognition and route prediction [10].

Recently with more and more mobile phones equipped with sensors, a few researchers initiated the research in individual/group behavior mining with mobile sensing data. For instance, MetroSense [2], a people-centric paradigm for urban sensing, explores sensor-embedded mobile phones to support personal and public sensing. By taking advantage of the data collected by mobile phones, Reality Mining project initiated at MIT intends to observe and characterize the social behaviour of individual users and organizations [11]. Another interesting study based on the monitoring of 100,000 mobile phone users, conducted by Northeastern Univ. in US, discovered that human trajectory has a high degree of spatial-temporal regularity, and humans follow simple reproducible patterns regardless of the diversity of travel history of individuals [12].

2.2 Context Inference in Smart Spaces

Earlier work on context inference mainly relies on static sensing infrastructure that is already deployed in smart spaces. One early project funded by the EPSRC in the UK was concerned with measuring crowd motion and density using cameras to detect potentially dangerous situations [13]. The Active Bats system uses ultrasonic sensors and the triangulation location-sensing technique to locate indoor objects [14]. Semantic Space builds an ontology-based infrastructure for extracting and querying contexts from smart spaces [15]. Yu et al. explore a set of static cameras and wearable sensors to mine semantic information like user attitudes in a smart meeting environment [16]. Sensor Andrew [17], a campus-wide static sensor network, is designed to host a wide range of applications including campus utility monitoring, social networking, and campus security surveillance.

2.3 Social Network Analysis

Humans are social by nature. People constantly participate in social activities to interact with others and form various communities. Social activities such as making new friends, forming an interest group to exchange ideas, sharing knowledge with others are constantly taking place in human society. The analysis of the social community interactions has been studied by social scientists and physicists for couple of decades [18]. An excellent introduction to the concepts and the mathematical tools for analyzing social networks can be referred to [19]. Early efforts on social network analysis are most based on the relational data obtained by survey.

During the last two decades, we have observed an explosive growth of Internet applications such as chatting, shopping, experience sharing, photo and video sharing, etc., which are now described as social software. These applications, along with the traditional e-mail, instant messaging, have changed the way that most of us used to communicate with each other and form social communities. Corresponding to this trend, a large body of work on social network analysis and knowledge discovery springs up, including Email communication networks [20], scientific collaboration and co-authorship network [21], etc..

More recently, as the internet stepped into the era of the Web 2.0, which advocates that users interact with each other as contributors to the web sites' content, researchers turned more attention to the online social utilities, such as Facebook, Twitter, and Blogs. For example, ArterMiner [22] seeks to harvest personal profile information from a user's homepage. Amit Sheth's research group has done much work on summarization of event info like space, time and theme from social web resources for building public services [23]. Twitter, a popular micro-blogging site, has been reported to support real-time mining of natural disasters such as earthquakes [24] and the moods of citizens [25].

3 A General Architecture

A general architecture for social and community intelligence system is shown in Fig. 2, which consists of five layers: *pervasive sensing layer*, *data anonymization layer, hybrid learning layer*, *semantic inference layer*, and *application layer*.

Layer 1: The large-scale pervasive sensing layer involves the three major information sources: mobile and wearable devices, static sensing infrastructure, social web and Internet services. The three sources have different attributes and strengths:

- Mobile devices and wearable sensors are always user-centric, thus great at sensing individual activities, interpersonal interactions, and significant locations.
- Static infrastructure, on the other hand, enables the detection of indoor user activities, group activities, and space context.
- Social Web is a major source to extract user profile info, significant relationship among users in a group activity. Extracted real-time event information (e.g., from Twitter) is also useful to recognize the ongoing group activity.

Due to the diverse features, aggregation and fusion of data from those three different sources provides unique opportunities to community intelligence extraction.

Layer 2: As privacy is a major concern for both private and organizational data sharing, our proposed framework incorporates an anonymization layer before the data releasing and processing. All the data released must be sufficiently anonymized, and different anonymization algorithms can be applied for privacy protection.

Layer 3: The hybrid learning layer applies diverse machine learning and data mining techniques to converting the low-level single-modality sensing data into high-level features or micro-context, the focus is to mine the frequent data patterns to derive the individual's behavior and single space context, before extracting the complete social and community intelligence.

Level 4: The semantic inference layer is needed when different features or micro-context need to be aggregated using logic-based inferences, it is complementary with statistical learning approach and often very effective to process the explicit rules describing the logical relationship between layer 3 outputs and expected SCI, based on expert's domain knowledge.

Layer 5: The application layer includes a variety of potential services that can be enabled by the availability of SCI. An application might be installed directly on the mobile device, or run on remote servers (such as a Web application) but communicate with the mobile device via wireless gateways.

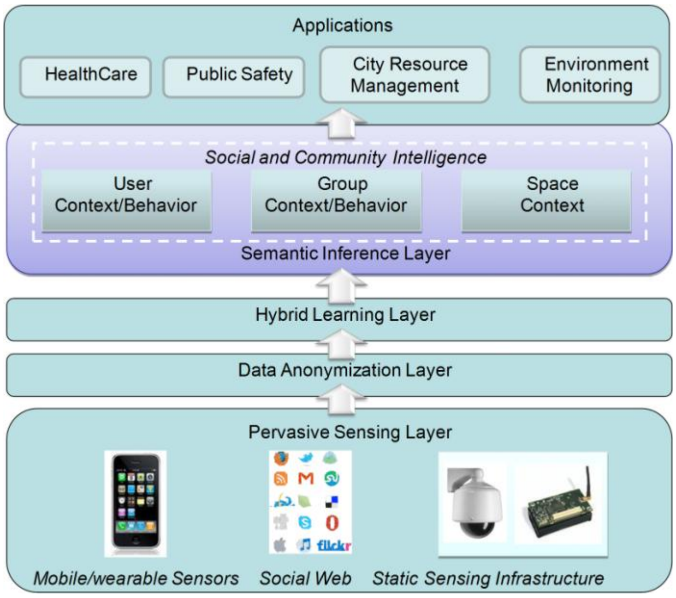

Fig. 2. A general architecture for community intelligence

4 Major Application Areas

SCI applications are mainly driven by the needs to (1) develop better social software to facilitate interaction and communication among groups of people; (2) predict the real-time change of real world to benefit human life. Here we can foresee at least the following six main SCI application areas:

4.1 Social Network Services

By logging various aspects of physical interactions among users (e.g., co-location, conversations) and mining user behavior patterns (e.g., place of interests), SCI nurtures the development of many social network services, such as friend recommendation and interpersonal interaction enhancement.

(1) Friend recommendation. By monitoring one's activities with mobile phones, including text messages, phone calls, and encounters, the FriendSensing application can recommend people to its users [26]. The Serendipity system calculates a similarity score by extracting the commonalities between two proximate users' profiles and behavioral data, and alerts the user that someone nearby might interest him/her [27].

(2) Interpersonal interaction enhancement. The CenseMe project exploits off-the-shelf smart phones to automatically infer people's presence (e.g., walking on the street, dancing at a party with friends) and then shares this presence through social network portals such as Facebook [28]. Koji et al. uses specially designed work badges to study the relationship between productivity and interpersonal interactions in a workplace. The badges contain infrared sensors, microphones, accelerometers, and location sensors to record the location and duration of conversations among workers, their physical distance apart, encounters, upper body motions, and so on [29].

4.2 Urban Sensing and City Resource Management

With wireless sensor platforms in the hands of masses, we can leverage community sensing to address urban-scale problems, such as city resource monitoring, traffic planning, and better use of public utilities.

Nericell is a system that can monitor road (e.g., potholed roads) and traffic conditions (e.g., chaotic traffic) using accelerometer, microphone, and GPS sensors in mobile phones [30]. MIT's Real Time Rome project (http://senseable.mit.edu/realtimerome) 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 combining local conditions with biker-contributed data (using mobile phones). It enables area bikers to plan routes with the least probability of traffic accidents and the best air quality. Zheng et al. extract interesting locations and travel sequences from multiple user's GPS trajectories, and enable travel recommendations for new visitors of a city [31].

4.3 Environment Monitoring

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

(1) Nature preservation. With the help of human volunteers, the Great Backyard Bird Count project reports the cumulative counts of birdwatchers from across American in its website (http://www.birdsource.org/gbbc/). The MIT Owl project (http://web.mit.edu/newsoffice/2008/tracking-0822.html) is more interesting, which aims at leveraging the network of smart phones equipped with GPS, compasses, and directional microphones, to lessen human efforts in assessing owl populations.

(2) Pollution measurement. With the aid of portable pollution sensing devices, there have also been several projects targeting environment pollution measurement. The BikeNet application measures several metrics to give a holistic picture of the cyclist experience, including the CO2 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 [32]. In the PEIR project, GPS-enabled phones are used to detect user transportation mode (e.g., driving, walking), which is then used to assess an individual's environmental impact and exposure, like carbon footprints and exposure to air pollution [33].

(3) Disaster Reporting. The real-time user contributed data is helpful for emergent or dangerous event detection and reporting. For example, Twitter has been reported to support rapid response to the social or natural disasters such as terrorism attack in Bombay [34] and earthquakes in Japan [24]. Comparing to traditional media, community sensing is more vigilant.

4.4 Human Health

SCI brings new opportunities for public health monitoring and personal well-being improvement.

(1) Public health. SCI can facilitate the anticipation and tracking of disease outbreaks across populations. For example, Epidemics of seasonal influenza are a major public health concern, causing tens of thousands of deaths worldwide each year. Its impact can be reduced by early detection of the disease activity. The Google researchers have shown that by mining indirect signals from millions of geographically localized health-related search queries, one can estimate the level of influenza-like illnesses in regions of the United States with a reporting lag of just 1 day [35]. It is faster than the estimates provided by government agencies, which publish regional data weekly based on virology and clinical statistics.

(2) Human well-being. With community sensing, we can log personal physical activity trajectory, track the food intake, sense the mental status in real-time, and record the social activities we attend each day, which can be used to improve human well-being management. For example, the Neat-o-Games system detects human movements (e.g., walking, running) by using a wearable accelerometer, and uses the computed quantity of motion to control the avatar of the user in a virtual community race game [36]. Playful Bottle is a mobile social persuasion system to motivate healthy water intake [37]. Nutrition Monitor, a mobile application, can track user food consumption and trends over time, and warn the user against unhealthy food choices [38].

4.5 Sentiment Applications

Sensing of user sentiments is important to context-aware computing, with which the applications can act accordingly. However, using physical sensors to directly sense personal sentiments is not an easy thing. Researchers have been exploring indirect ways to deal with this, one of which is to mine user-generated Web data. Some systems use a Web survey method. For example, Emotional City (http://www.emotionalcities.com/) and D-Tower (www.d-toren.nl) collect citizen moods through daily Web surveys, and display the emotions of the city through the change of light-colors of a building or a public sculpture. Others explore machine learning algorithms for sentiment mining. Bollen et al. proposes an extended Profile of Mood States (POMS) method to extract six dimensions of mood (e.g., tension, anger) from user posted tweets in Twitter [25].

4.6 Public Safety

Public safety involves the prevention of and protection from events that could endanger the safety of the general public, these events can be crimes or disasters. Public video surveillance systems have assisted a lot to city-wide event sensing and safety maintenance [39]. Recently, the Boston police department has embraced user contributed sensor data to assist in crime prevention [34].

5 Research Issues

We now turn our attention to key SCI research issues. To facilitate the development of SCI applications, one fundamental issue is gathering and management of heterogeneous data from different information sources. Other important issues are using machine learning algorithms to make sense of the "digital footprints" revealing the predefined patterns or unforeseen behaviors about individual, group and community, as well as the privacy concerns raised by sensing our daily lives.

5.1 Participatory or Opportunistic Sensing?

The first issue to be considered in sensing is what roles people should play in community sensing. For example, should they be interrupted to control the status (e.g., accept, stop) of a sensing task? There are two possible ways for sensing:

- *Participatory sensing*. It incorporates people into significant decision making process of the sensing system, deciding which application request to accept, what data to share, and to what extent privacy mechanisms should be allowed to impact data fidelity. That's to say, it allows participants to retain control over their raw data. The Personal Data Valut system is based on this idea, which seeks to provide easy-to-use toolkits to support data control [1].
- *Opportunistic sensing*. It shifts the burden of users by automatically determining when devices can be used to meet application's sensing requests. Instead of requiring human intervention to actively and consciously participate in the sensing, opportunistic sensing requests that a sensing device is automatically used whenever its state (location, user activity, and so on) matches an application's application requirements. This approach is proposed in [28].

Obviously there exists a tradeoff between participatory sensing and opportunistic sensing. Participatory sensing places demands on user involvement, which restricts the pool of willing participants, while opportunistic sensing takes on more resources for decision-making. More work needs be done to balance users' involvement and proper control while integrating proper protection mechanisms on data privacy (more discussions on privacy are given in Section 5.4).

5.2 Managing Heterogeneous and Multi-modal Data Sources

As in SCI system, the data producers can be very different in terms of modality (e.g., mobile phones, fixed cameras, Web services), their connectivity to the Internet (e.g., constant, intermittent, or affected by a firewall), their sharing willingness or privacy sensitivity, and resource capabilities for processing data locally. The information consumers are also heterogeneous in terms of running environments (applications that run locally or at community-level remotely), data needs (some might need only a high-level context information while others might need raw sensor data). The heterogeneity leads to several challenges on data management:

(1) Multi-modal. Different type of sensors have different attributes and capabilities, they might have different accuracy in sensing the physical and virtual world. Integrating information from diverse data sources adds difficulty to SCI mining. Raw data from different sensor sources need to be transformed to the same metrics and represented by a shared vocabulary/ontology to facilitate the learning and inference process [15].

(2) Temporal and Continuous. The sensing data is recorded according to the time sequence, the system should consider multiple samples in the data stream while modeling the behaviors of individual and group, rather than consider each sensor reading in an isolated way. In addition, as the real world systems are all continuous, it's important to build models catering for the discrete, sampled sensor state.

(3) Inconsistency. The same sensor may sense the same event under different conditions (for example, sensing one's voice in a quiet office or noisy restaurant). However, for the same event, user context often leads to different inference results (good or poor). Due to environmental differences, a group of co-located sensors running the same classification algorithm and sensing the same event in time and space could compute different inference results, and thus leads to the issue of system inconsistency. Miluzzo *et al.* have proposed a collaborative approach to dealing with this inconsistency problem [40] and more solutions are needed.

(4) Difficult to label all data. Asking human to label large amount of data set is often difficult since it is extremely time consuming to perform real-life experiments to collect data, it takes even more time to label all the data properly. Thus it is highly desirable to learn system models from relatively small amount of labeled data [41].

5.3 Extracting High-level SCI from Low-level Sensing Data [41]

Social and community pattern mining considers the identification of a set of characteristics or behaviors associated with a social community based on the collection of intermediate-level individual activity/space context traces. Such social communities can be flexibly formed by those people in the same organization, at same places, with same behaviors, of same interests, etc., depending on different social application requirements [42]. By pooling individual user's context traces together and mining the underlying social patterns, different social or group behaviors can be extracted [43]. The extracted social context can be a social event such as an open concert, can be a social pattern in daily activity, can be a relationship among a group of people, and can be socially significant locations.

The key of the SCI pattern mining is to identify user similarity in the aforementioned social patterns with the objectives of offering social aware services. Unsupervised learning techniques, such as clustering, latent semantic analysis, matrix factorization, can be applied to achieve social context mining based on the user behavioral similarities. The process includes mining and discovery of common social contexts such as personal characteristics, cuisine preferences, eagerness of social participation, and also discovery of undefined social patterns for interest matching and social choice ranking.

In order to infer the social events based on the user context traces, the semantic gap between the low-level individual activities/spaces (e.g., walking/street, eating/restaurant, etc) and high-level social events (e.g., meeting, party, etc) should be bridged using some machine learning and inference techniques. As highlighted previously, the analysis of the latent relations between the basic human activities and semantic social events is the research focus of this module, with the goal of learning an ontology describing the relationship between the basic human activities and semantic social events.

5.4 Privacy and Trust

Sharing and revealing personal digital data could have a number of risks on user privacy. Compared with personal data (e.g., user profile, IDs), data gathered in community can reveal much more information about individual and organization's behaviours. For example, your location might reveal your interests; the health data about an organization might suggest environmental problems for the staff. The impact is obvious: if personal data cannot be anonymized and under the control of data owners, people may be less likely to share their data [28].

Privacy protection involves many elements, including identity (who is asking for the data?), granularity (how much does the data reveal about people? does it reveal one's identity?), and time (how long will the data be retained?) [1] There are two main research areas that deal with these needs: *data anonymization* [28, 43] and *user control* [1].

(1) Data anonymization techniques. The objective of *data anonymization* is not revealing the identity of users when they contribute their data. Several methods have been proposed. For instance, MetroSense uses *k*-anonymous method when users contribute location data to a server, where a user's position is generalized to a region containing at least *k* users [28]. Another promising approach based on secure multiparty computation allows mining data from many different organizations without ever aggregating these data into a central data repository. Each organization performs part of the computation based on its privately held data, and uses cryptography to encode intermediate results that must be communicated to other organizations

performing other parts of the computation [43]. Other privacy-preserving methods are also being explored, such as sharing only statistical summaries of the individual data sets, and inserting random perturbations into individual data records before sharing them [43].

(2) Enhancing user control and decision making. User control is very important in personal data sharing as it is about what one wants to reveal and to whom one allows the system to reveal. For example, you might want to track your heart rate each day, but there is no reason to share that information with anyone but your doctor. Researchers in this field are exploiting methods that enable users to manage their data by tailoring access-control and data-management tools [1].

In addition to data privacy issue, trust of the data sources is another big thing. To mine social and community behaviors, we often need to import data from many anonymous participants. If there lacks the control to ensure the source is valid and information is accurate, this can lead to data trust issue. For example, Twitter data is sometimes unreliable due to the text's unmediated and casual nature; mobile phone users may send incorrect or even faked data to the data centre. Therefore, trust maintenance and abnormal detection methods should be developed to determine the trustworthiness and quality of collected data.

6 Conclusion

Social and Community Intelligence (SCI) represents a new interdisciplinary research and application field. With the rapid accumulation of "digital footprints" at community scale, we believe that the research scope of SCI will expand and its applications to multiply in next years to come. As we have discussed, the prevalence and development of SCI still face challenges ranging from multi-modal data gathering, heterogeneous data representation and management, to complex intelligence inference and privacy issues, which are expected to nurture a series of new research opportunities for academic researchers, industrial technologists, and business strategists as well. Even though the existing practices on social and community intelligence mainly consider single type of information sources − static sensor infrastructure, mobile and wearable sensors, or Internet and social web − we expect to see the explosion of the research on aggregated power of the three information sources as well as innovative applications enabled by SCI.

References

1. Shilton, K.: Four billion little brothers: Privacy, mobile phones, and ubiquitous data collection. Communications of the ACM 52(11), 48–53 (2009)
2. Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A.: People-Centric Urban Sensing. In: Proc of the 2nd Annual International Workshop on Wireless Internet (2006)
3. Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., Hahnel, D.: Inferring activities from interactions with objects. IEEE Pervasive Computing 3(4), 50–57 (2004)

4. Pollack, M. E.: Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment. AI Magazine 26(2), 9–24 (2005)
5. Tentori, M., Favela, J.: Activity-aware computing for healthcare. IEEE Pervasive Computing 7(2), 51–57 (2008)
6. Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A., Estrin., D.: Habitat monitoring with sensor networks. Communications of the ACM 47(6), 34–40 (2004)
7. Connolly, C. I., Burns, J. B., Bui, H. H.: Recovering social networks from massive track datasets. In: Proc. of the IEEE Workshop on Applications of Computer Vision, 1–8 (2008)
8. Wolf, J., Guensler, R., Bachman, W.: Elimination of the travel diary: An experiment to derive trip purpose from GPS travel data. In: Proc. of the 80th Annual Meeting of the Transportation Research Board (2001)
9. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing 7(5), 275–286 (2003)
10. Liao, L., Fox, D., Kautz, H.: Learning and inferring transportation routines. In Proc. of the 19th AAAI Conf. on Artificial Intelligence, 348–353 (2004)
11. Eagle, N., Pentland, A., Lazer, D.: Inferring Friendship Network Structure by using Mobile Phone Data. National Academy of Sciences (PNAS) 106(36), 15274-15278 (2009)
12. Gonzalez, M. C., Hidalgo, C. A., Barabasi, A. L.: Understanding individual human mobility patterns. Nature 453(5), 779-782 (2008)
13. Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L.: Crowd Analysis: a Survey in Machine Vision and Applications. Computer Science 19 (5-6), 345-357 (2008)
14.Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The Anatomy of a Context-aware Application. In: Proc. of MOBICOM1999 (1999)
15. Wang, W., Zhang, D., Dong, J.S., Chin, C.Y., Hettiaarchchi, S.R.: Semantic Space: An Infrastructure for Smart Spaces. IEEE Pervasive Computing, 32-39 (2004)
16. Yu, Z.W., Yu, Z.Y., Aoyama, H., Ozeki, M., Nakamura, Y.: Capture, Recognition, and Visualization of Human Semantic Interactions in Meetings. In: Proc. of IEEE PerCom 2010, Mannheim, Germany, 107-115 (2010)
17. Rowe, A., Berges, M., Bhatia, G., Goldman, E., Rajkumar, R., Soibelman, L., Garrett, J., Moura, J.: Sensor Andrew: Large-Scale Campus-Wide Sensing and Actuation. Carnegie Mellon University (2008)
18. Freeman, L. C.: The Development of Social Network Analysis: A Study in the Sociology of Science. Empirical Press (2004)
19. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994)
20. McCallum, A., Wang, X., Corrada-Emmanuel, A.: Topic and role discovery in social networks with experiments on Enron and academic email. Journal of Artificial Intelligence Research 30(1), 249–272 (2007)
21. Barabasi, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Statistical Mechanics and its Applications 311(3–4), 590–614 (2002)
22.Tang, J., Jin, R.M., Zhang, J.: A Topic Modeling Approach and its Integration into the Random Walk Framework for Academic Search. In: Proc. of 2008 IEEE International Conference on Data Mining (ICDM 2008), 1055-1060 (2008)
23. Sheth, A.: Computing for Human Experience – Semantics-Empowered Sensors, Services, and Social Computing on the Ubiquitous Web. IEEE Internet Computing 14(1), 88-97 (2010)
24. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In: Proc. of WWW 2010 Conference (2010)
25. Bollen, J., Pepe, A., Mao, H.: Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Proc. of WWW 2009 Conference (2009)
26. Quercia, D., Ellis, J., Capra, L.: Nurturing Social Networks Using Mobile Phones. IEEE Pervasive Computing (2010)

27. Eagle, N., Pentland, A.: Social serendipity: Mobilizing social software. IEEE Pervasive Computing 4(2), 28–34 (2005)
28. Campbell, A. T., et al.: The Rise of People-Centric Sensing. IEEE Internet Computing 12(4), 12-21 (2008)
29. Ara, K., et al.: Sensible Organizations: Changing Our Business and Work Style through Sensor Data. Journal of Information Processing 16, 1-12 (2008)
30. Mohan, P., Padmanabhan, V. N., Ramjee, R.: Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smart phones. In: Proc. of ACM SenSys'08 (2008)
31. Zheng, Y., Xie, X., Ma, W.Y.: GeoLife: A Collaborative Social Networking Service among User, location and trajectory. IEEE Data Engineering Bulletin 33(2), 32-40 (2010)
32. Eisenman, S. B., et al.: The bikenet mobile sensing system for cyclist experience mapping. In: ACM SenSys 07, 87–101 (2007)
33. Mun, M., et al.: PEIR: the personal environmental impact report as a platform for participatory sensing systems research. In: Proc. of MobiSys 09 (2009)
34. Sheth, A.: Citizen Sensing, Social Signals, and Enriching Human Experience. IEEE Internet Computing 13(4), 87-92 (2009)
35. Ferguson, N.M., et al.: Strategies for mitigating an influenza pandemic. Nature 442(7101), 448-452 (2006)
36. Fujiki, Y., Kazakos, K., Puri, C., Buddharaju, P., Pavlidis, I., Levine, J.: NEAT-o-Games: Blending Physical Activity and Fun in the Daily Routine. ACM Computers in Entertainment 6(2) (2008)
37. Chiu, M.C., et al.: Playful bottle: a mobile social persuasion system to motivate healthy water intake. In Proc. of UbiComp09, 185-194 (2009)
38. Dorman, K., et al.: Nutrition Monitor: A Food Purchase and Consumption Monitoring Mobile System. In: Proc. of MobiCASE09 (2009)
39. Hampapur, A., et al.: The IBM Smart Surveillance System. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Washington D.C. (2004)
40. Miluzzo, E., et al.: Darwin Phones: The Evolution of Sensing and Inference on Mobile Phones. In: Proc. of MobiSys '10, San Francisco, CA, USA (2010)
41. Yang, Q.: Activity recognition: Linking low-level sensors to high-level intelligence. In Proc. of the 21st Int'l Joint Conf. on Artificial Intelligence, 20–25 (2009)
42. Pentland, A.: Socially aware computation and communication. IEEE Computer, 38(3), 33–40 (2005)
43. Mitchell, T.M.: Mining Our Reality. Science 326 (5960), 1644 - 1645 (2009)