Towards Large-Scale Mobile Sensing and Analytics

Qin Lv
Department of Computer Science
University of Colorado Boulder
qin.lv@colorado.edu

ABSTRACT
The rapid adoption of smartphones and their increased sensing capabilities make it possible to collect a large amount of multi-modality mobile sensing data, which tend to be diverse, dynamic, and may vary significantly across users and applications. Large-scale mobile sensing and analytics, an integrated approach which collects, analyzes, and acts upon mobile sensing data at scale, has the potential to significantly enhance people’s daily lives in a wide variety of domains. In this paper, we identify the key characteristics and design challenges for large-scale mobile sensing and analytics, which are then demonstrated through three case studies in environmental monitoring, transportation electrification, and mobile video chat.

INTRODUCTION
During the recent years, the number of users who carry mobile phones has grown significantly. The increased computing and sensing capabilities of mobile phones, along with cloud computing platforms and Web-based services, make possible various mobile computing and mobile sensing applications [3, 4]. Large-scale mobile sensing and analytics is becoming a reality. On the one hand, large amounts of multi-modality mobile sensing data (e.g., GPS location, acceleration, lighting, audio, video) can be captured by individual mobile phones. On the other hand, such rich sets of mobile sensing data, together with data available from the Web or other sources (e.g., tweets, news articles, weather information), need to be carefully analyzed and processed (on the fly) in order to support knowledge discovery, user feedback, and decision making.

Aiming to deliver both high efficiency (e.g., latency, energy efficiency) and high quality (e.g., accuracy, precision, recall), large-scale mobile sensing and analytics call for a systematic research paradigm which closely integrates the design components governing the sensing, analysis, and feedback process. In particular, the following two themes need to be considered in unison:

- Data flow: Given a specific application and mobile sensing platform, we need to determine what data can (need to) be collected and stored, what analysis will be performed on what data, how will the analysis results be presented to users and support decision making. Note that in many scenarios, the data we consider include not only sensing data collected from mobile phones, but also data generated from the Web or other stationary sources. This is a continuous, iterative, and dynamic process, and new knowledge gained from previous data should be readily integrated for further enhancements of the sensing and analytics components. Due to the limited computing capabilities of mobile sensing platforms, understanding the cost and utility of different sensing and analysis strategies is of particular importance for large-scale mobile sensing and analytics.

- System infrastructure: Besides individual mobile phones, other system components are also needed to support a large-scale mobile sensing and analytics infrastructure. These may include other sensing devices (mobile or stationary), local and/or remote servers and cloud computing facilities for data storage, analysis, and web-based services. These system components may differ significantly with regard to computation, communication, sensing capabilities, availability, latency, etc. Given the diversity and dynamics of both the system components and the data sensing/analysis needs, an intelligent and adaptive mechanism is needed to dynamically select the right sensing/analysis needs and partition them properly within the system infrastructure.

DESIGN CHALLENGES
Large-scale mobile sensing and analytics call for a systematic investigation of not only the underlying sensing and computing platforms, but more importantly, a unified mobile sensing and data analytics system/algorithm research that addresses a number of design challenges.

- Personalized mobile sensing and analytics: Mobile phones are carried by individual users, and people’s interests and daily activities vary significantly, so are their contextual environments. Although large amounts of mobile sensing data can be captured from many users, customized sensing, analysis, and feedback for specific individuals are particularly important for the adoption and success of large-scale mobile sensing and analytics applications. Personalization requires the identification...

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Mobile Sensing: Challenges, Opportunities and Future Directions. UbiComp ’11, Sep 17-Sep 21, 2011, Beijing, China.
Copyright 2011 ACM 978-1-60558-843-8/10/09...$10.00.
of a user’s specific characteristics (e.g., mobility pattern, driving behavior) and current contexts (e.g., walking, indoor, quiet). These may be achieved through a number of data sensing and analytics capabilities. However, challenges remain. For instance, limited battery energy budget and communication bandwidth constrain the total amount of data that can be gathered and processed by a mobile device. Miniature sensors have limited capabilities and are susceptible to ambient noise. The wisdom of a single device is thus bound. Delivering personalized sensing and analytics through a resource constrained mobile platform is challenging.

- **Collaborative sensing and analytics:** Due to the limited resources and sensing capabilities of individual mobile devices, collaboration among a large number of mobile devices to deliver efficient yet high-quality data sensing and knowledge discovery becomes essential. By leveraging social groups and online social communities, socially collaborative data sensing and analytics can be particularly useful. Collaboration can occur in situ (e.g., with neighboring devices) or over the Internet via data sharing and community-based forums (e.g., average fuel efficiency of other people driving a Prius). Some of the key questions are how to effectively leverage the crowd at large, efficiently collect information from them, intelligently leverage wall-powered infrastructure if available, and conduct multi-modality sensor data fusion for knowledge discovery of the crowd wisdom, thereby delivering high-quality and high-efficiency services yet allowing for service personalization.

- **Privacy and security:** Data gathered by and stored in people's mobile phones are often highly personal and sensitive, such as geo-tagged and timestamped sensor readings or audio/video recordings. Therefore, remote processing or data mining of mobile sensing data, or collaboration among a large number of mobile devices always raise serious privacy and security concerns. Besides advances in cryptography and access control, it also calls for data analysis techniques to carefully examine the gathered data (raw data and extracted features), identify shared/similar patterns such that data may be anonymized or obscured properly to maintain useful patterns while ensuring user privacy and security. Note that different users and different applications may have different implications for privacy and security. Therefore, such data analysis techniques need to be efficient, automated, and adaptive so they can be run easily on different types of mobile sensing data.

**CASE STUDIES**

The first system we consider is MAQS, a mobile sensing system for personalized indoor air quality (IAQ) monitoring [1]. IAQ is affected by different air pollutants, and can vary substantially by room, time, location, activity, etc. As people’s mobility patterns also vary, personalized IAQ monitoring and online user feedback are very useful but challenging. Through an integrated design of the IAQ sensing device, mobile sensing app, database server and web server, MAQS successfully leverages a number of mobile sensors and air pollutant sensors, as well as analysis techniques including WiFi and temporal user mobility based room classification, collaborative sensing and data sharing, ventilation rate estimation, IAQ data analysis. Users can view not only their own IAQ, but also aggregated data of others, thus enabling better understandings of IAQ and possible improvements of personal or shared environments.

Next, we consider a mobile sensing system for monitoring users’ driving behaviors and analyzing their impacts on the energy efficiency of plug-in hybrid electric vehicles (PHEVs) and the environment [2]. Through multiple mobile phone sensors and onboard diagnostic devices (OBDs), detailed user driving information can be captured for driving mode classification and correlation analysis of vehicle energy use and greenhouse gas emission. Online feedbacks can also be provided to help users identify good/bad driving behavior. Again, the system needs to be personalized as different people drive differently under different driving conditions (e.g., weather, road, traffic). Still, socially collaborative sharing, which help identify users with similar driving patterns while protecting user privacy, is desirable.

The third case study considers mobile video chat services. Recently, online video chat services that randomly pair people up with strangers have gained substantial popularity but suffer from the existence of misbehaving users and obscene content. Our SafeVchat system aims to determine whether a user is misbehaving through the fusion of multiple image-based classifiers (e.g., face, nose, eye, upper body, skin color) [5]. Sampling video frames and analyzing images can be time consuming, and different classifiers may perform differently for different users’ video chat data. When deployed in a mobile sensing and analytics platform, a careful system/algorithms design that dynamically selects and partitions the mobile video sensing and image classification/fusion functionalities are thus essential in order to achieve the best tradeoff between system efficiency and classification accuracy.

**REFERENCES**


