A Framework for Data Quality and Feedback in Participatory Sensing

Sasank Reddy*, Jeff Burke*†, Deborah Estrin*, Mark Hansen*, Mani Srivastava*
* Center for Embedded Networked Sensing,
†Center for Research in Engineering, Media and Performance
University of California, Los Angeles
{sasank, jburke, mbs}@ucla.edu,
destrin@cs.ucla.edu, cocteau@stat.ucla.edu,

Categories and Subject Descriptors

C.2.4 [**Computer Communication Networks**]: Distributed Systems

General Terms

Design, Experimentation, Human Factors

Keywords

Mobile Systems, Human Sensor Networks

1 Introduction

The rapid adoption of mobile phones by society over the last decade and the increasing ability to capture, classifying, and transmit a wide variety of data (image, audio, and location) have enabled a new sensing paradigm - where humans carrying mobile phones can act as sensor systems. Human-in-the-loop sensor systems raise many new challenges in areas of sensor data quality assessment, mobility and sampling coordination, and user interaction procedures.

The idea of using mobile phones as sensing instruments is nothing new. In fact, they have been used successfully in personal sensing applications for social dynamics and health monitoring [1, 2]. More recently there has been growing excitement in using mobile phones to analyze the physical world. [3] suggests that mobile phones can enable participatory research. [4, 5] focus on infrastructure components needed to enable this new sensing paradigm. But little attention has been paid to how having humans operate, carry, and interact with these sensors can affect participation, data quality, and spatio-temporal coverage. In turn, our research builds on the ideas of [3] and focuses on methods to model the sensing behavior of individuals, access their spatial and temporal availability, and determine appropriate feedback mechanisms to help in the data collection process. The remaining sections of the paper are arranged as follows: a.) provide an overview of our software architecture created for participatory sensing, b.) describe models for quality of sampling and ideas regarding mobility and feedback, and c.) present basic conclusions from preliminary pilot studies.

SenSys'07, November 6–9, 2007, Sydney, Australia.
ACM 1-59593-763-6/07/0011

2 System Architecture

Our system consists of three components: an application that runs on the phone to collect sensor data (Campaignr), a sensor store (SensorBase) to collect/aggregate data, and a set of services that interact with the phone and the store to enable value added services (Fig. 1).

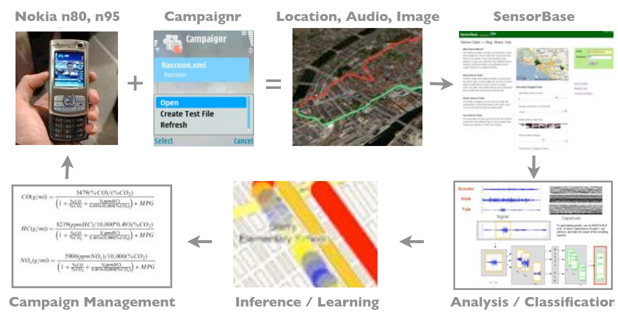

Figure 1. Components for Participatory Sensing.

2.1 Campaignr

In order to enable participatory sensing from the phone, we created a software application that facilitates in situ data collection. This application, which we refer to as Campaignr, is designed to support various data collection initiatives or "campaigns" [3]. It is easily configurable (through an XML configuration file), robust to handle various sensor modalities and network connections, and contains a intuitive interface. Campaigns can be created to use audio, image, and location modalities with continuous or triggered sampling.

2.2 SensorBase

The recorded sensor readings from Campaignr are uploaded to a data store (SensorBase). SensorBase is a sensor data archive service, based on MySQL and PHP. Users can interact with SensorBase to publish, share, and manage sensor data either through its web GUI front end or by using its HTTP Post interface. Furthermore, SensorBase provides a data sharing model that enables fine-grained data management so that sensor data can be audited before being released to a public forum.

2.3 Additional Services

Since SensorBase and Campaignr share a common interface, it is easy to build value-added services in the system to help in various data collection scenarios. For instance, we have created applications to enable users to easily tag data, to help with the visualization of data, and to add additional inferences by combining different sources of information or learning techniques.

3 Human Factor Research

Having the participant involved in the sensing process leads us to an unique research challenge: Can we model the "human factors" involved in the sensing process and use them to make the data gathering process efficient while helping users better understand their contributions. Thus, we focus on three main ideals: methods to model the quality of sampling provided by participants, ways to make inferences from the spatial and temporal mobility of users, and mechanisms to provide feedback to enable efficient participation.

3.1 Quality of Sampling

Based on extensive review of possible campaign scenarios and sampling techniques, we came up with five distinct measures of quality involved in human-in-the-loop sampling. We believe that these metrics are essential in making the participatory sensing process more efficient by enabling methods to pick, refine, and evaluate the users needed and data obtained from a campaign.

Timeliness - Represents the latency between when a phenomenon occurs and when the sample is available for a data processing unit. It is affected by the sampling, communication, and auditing phases of sensing.

Capture - Describes the quality of a particular reading in terms of the ability in determining a particular feature - probability of inference. It is an attribute affected by both the specifications of the sensors used and the capturing process by the participant.

Relevancy - Tells how well the sample describes the phenomenon that is sought for capture. Ranges from irrelevant (not related to the item of interest) to completely relevant (describes exactly the item of interest).

Coverage - Represents the spatial and temporal availability (mean and variance) associated with the coverage provided by a user. The spatial and temporal extent and resolution play a key role here as well.

Responsiveness - Describes the probability of responding to a directed, in situ sensing request.

3.2 Mobility Mining

Another important aspect of our research is to mine the spatio-temporal behaviors of participants. Mobility analysis will help campaign designers to assess which users could be available for capturing data based on spatial and temporal constraints. Also, mobility mining will help enable context-sensitive feedback for participants. Currently, two methods are being researched as methods for mobility pattern recognition. The first involves point analysis of cell id and GPS data to find significant locations and learn common tracks. The second deals with the idea of "eigenbehaviors" where location analysis is performed to find behavior features with the most variance as indicators of multi-resolution patterns [1].

3.3 Feedback Mechanisms

We plan to use feedback to communicate the state of the campaign to the user and also as a tool for passive persuasion. We are researching using prompts, positive reinforcement techniques, social validation, and adaptive interfaces as feedback mechanisms.

Just-In-Time Prompts provided using SMS and MMS will be based as spatio-temporal triggers and encourage additional sampling in areas of interest and provide feedback to enhance quality of sampling.

Positive Reinforcement will be introduced through an incentive system that rewards credits weighted on quality of sampling. Credits can be used for obtaining additional services (ability to suggest campaigns) or out of band rewards (monetary gain).

Adaptive Interfaces are essential to make feedback continually effective. By varying the time, the type, and frequency of feedback, we hope to continue the upward trajectory of quality and participation of individuals.

Social Validation refers to the concept that people determine what is correct based on what other people determine as correct. This mechanism will be especially helpful in encouraging honest use of the system by pointing out what the "mass" has concluded.

4 Initial Pilot and Future Work

We performed a few pilot campaigns in order to get some ground truth information and also test the validity of our models and metrics. The pilots typically involved groups of 5-10 individuals over the span of 1-2 weeks. The first set of pilot campaigns involved users collecting location (GPS and Cell Tower) data, and the second campaign set dealt with walk-ability of sidewalks in a community. Specifically, in the second campaign set, users were asked to take geo-tagged pictures whenever a sidewalk was encountered that had structural problems. Based on these preliminary campaigns, we found that users have distinct mobility profiles and have a general area of coverage. Furthermore, certain users were generally more careful with their sampling and thus provided data with higher quality. Also, we noticed that users have different connectivity levels in general - the mean timeliness varied from minutes for certain individuals to hours for others. Overall, these campaigns have served us in providing preliminary validation of our models and research goals. We plan to continue developing algorithms and methods to handle the "human" aspect of human-in-the-loop sampling.

5 References

[1] N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiq. Computing*, 10(4):255–268, 2006.

[2] A. Pentland. Healthwear: medical technology becomes wearable. *Computer*, 37(5):42–49, 2004.

[3] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and MB Srivastava. Participatory Sensing.

[4] A. Parker, S. Reddy, T. Schmid, K. Chang, G. Saurabh, M. Srivastava, M. Hansen, J. Burke, D. Estrin, M. Allman, et al. Network System Challenges in Selective Sharing and Verification for Personal, Social, and Urban-Scale Sensing Applications.

[5] A.T. Campbell, S.B. Eisenman, N.D. Lane, E. Miluzzo, and R.A. Peterson. People-centric urban sensing. *IEEE WICON*, 2006.