The Study of Node Reliability in a Participatory Sensor Network

By

Raushan Ara Dilruba

Student ID: 100705032P

MASTER OF ENGINEERING

Department of Computer Science and Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

December, 2013

The project entitled **The Study of Node Reliability in Participatory Sensor Network** submitted by Raushan Ara Dilruba, Student No: 100705032P, Session: October 2007 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING on ------------------

<u>BOARD OF EXAMINERS</u>

1.

(Dr. Mahmuda Naznin) **Chairman**
Associate Professor (Supervisor)
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh.

2.

(Dr. A. S. M. Latiful Hoque) **Member**
Professor and Head
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh.

3.

(Dr. Md. Shohrab Hossain) **Member**
Assistant Professor
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh.

4.

(Dr. Md. Yusuf Sarwar Uddin) **Member**
Assistant Professor
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh.

Declaration

It is hereby declared that this project or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Signature of the candidate

(Raushan Ara Dilruba)

Dedication

To my beloved parents, husband and child Zayra Afsheen Oriel

ACKNOWLEDGEMENT

We thank the Almighty for the successful completion of our research. I express my heartily gratitude to the respected supervisor Dr. Mahmuda Naznin, Associate Professor, Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET). We would like to acknowledge with great thanks to all friends who help us with giving various information and supports. I also thank my board members for their valuable comments and suggestions.

ABSTRACT

Participatory sensor network is a network where participants or nodes use mobile phones or social network and feed data to detect an event. Data is gathered and analyzed to an event. As data gathering is open to many participants, one of the major challenges is to identify if the reported observations are true or false. It becomes more challenging when node or participant's reliability is unknown or even the probability of event to be true is unknown. In our research, we study this challenge and observe that applying evolutionary method, event detection becomes more reliable. We call this approach Population Based Reliability Estimation (PBRE). In this research project we do simulation study and our experimental results show that PBRE performs better than other reliability estimation method in our defined environment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

In this chapter, we provide an introduction to the participatory sensor research field. Currently there are more than 3 billion smart phone users in the world, and this number is increasing at an impressive rate. This makes cell phones an excellent platform for sensing the environment at unprecedented spatio-temporal granularity. The new generation smart phones or devices have multiple embedded sensors (e.g., accelerometer, gyroscope, light, video, microphone, etc.) and can easily communicate with external static sensors via any of the built-in interfaces including bluetooth, infrared, or WiFi [32] and are increasingly capable of capturing, classifying and transmitting image, acoustic, location aware data. Smart device users can participate in an event detection or as location-aware data collection sources [3]. Through the use of sensors (e.g., cameras, motion sensors, and GPS) built into smart phones and web services to aggregate and interpret the assembled information, a new collective capacity has emerged in which people participate in sensing and analyzing aspects of their lives that were previously invisible. This concept is known as $ParticipatorySensing$. Let us consider some illustrative examples:

1. A group of citizens, organized through a social network, could use their mobile phones to take geo-tagged images as they move about town. Images of community assets are automatically uploaded and displayed on an interactive map could be used to promote neighborhood identity and local services. Images of safety hazards can help prioritize maintenance services.

2. Runners can use similar techniques to document and select scenic and shaded running routes away from roadways and can combine their shared data with personal fitness journaling.

3. Using the smart devices and web capabilities, each member of a family can monitor his or her travel and activity patterns for a few days to allow reflection and discussion of daily routines. In concert with engaging visual displays of similar information about their peer groups and neighborhoods, family-wellness coaches could provide highly personalized guidance [17].

ParticipatorySensing is a process of data collection and interpretation. Participatory sensing emphasizes the involvement of citizens and community groups in the process of sensing and documenting where they live, work, and play. It can range from personal observations to the combination of data from hundreds, or even thousands, of individuals that reveals patterns across an entire city. Most important, participatory sensing begins and ends with the people, individually or community wise. The type of information collected, how it is organized, and how it is ultimately used, may be determined in a traditional manner by a centrally organized body, or in a distributed manner by the collection of participants themselves. [17].

A *Participatory Sensor Network* consists of nodes which recruits participant to actively collect data for a common project goal within its framework. The nodes or participants use their personal mobile phones to sense various activities of their surrounding environment and submit sensed data through mobile network or social networking sites.

Finding reliable sources in a participatory sensor network is challenging due to the large proliferation of sensing and communication capabilities of the participant nodes and the availability of ubiquitous, real-time data sharing opportunities among nodes via mobile devices and social networking sites [30]. One conventional way to collect the reliable data on human interactions is to analyze the self-reported surveys. But this is a time consuming procedure. Data collection can be done using mobile devices like smart-phones, wearable sensing devices or through social networks [4, 11, 25, 28]. In a participatory network where the users are considered as participatory sensors, an event can be reported or detected [10, 23, 24]. One of the major challenges in this participatory sensing is ascertaining the truthfulness of the data and the reliability of the sources because data collection is open to a very large population. The reliability of the participants (or sources) denotes the probability that the participant reports correct observations. Reliability may be impaired because of the lack of human attention to the task, or because of the bad intention to deceive. Without knowing the reliability of sources, it is difficult to measure whether the reported observations are true or not. [34, 36, 37]

We study participatory sensing challenges in our research.

1.2 Participatory Sensor Network Architecture

In this section we discuss the architecture of a Participatory Sensor Network. Despite the diversity in why and how individuals engage in participatory sensing, the basic sensing process can be broken down into the following steps, each of which is facilitated by the combination following technologies: coordination, capture, transfer, storage, access, analysis, feedback, and visualization [17].

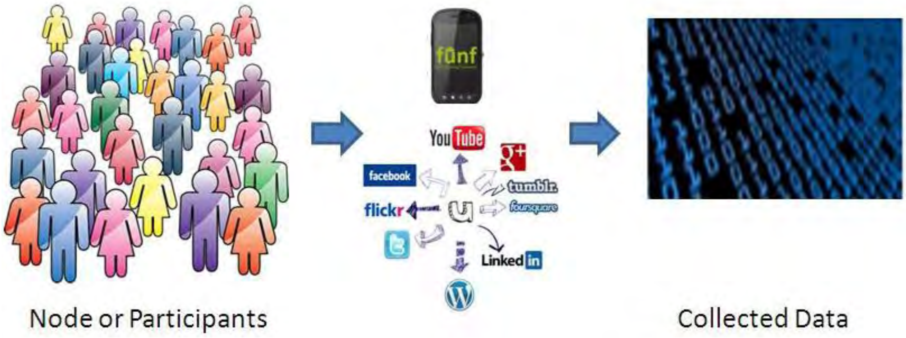

Figure 1.1: A Participatory Sensor Network

1.2.1 Coordination

Coordination involves recruiting and communicating with participants to explain the sensing effort and provide necessary guidance. Such communication is assisted by existing social networks, which can be accessed via computers, mobile phones, or face-to-face gatherings.

1.2.2 Capture

Capture is the acquisition of data on a smart phone or other smart devices. In addition to standard capabilities of smart phones, specialized software applications can be downloaded in project-specific configurations or be programmed directly by participants.

1.2.3 Transfer

Transfer takes place using mobile phones and wireless networks. Mobile phone software can make data uploading transparent to the participant and tolerant of inevitable network interruptions. Depending on the approach and the purpose of a project, either the participants or the organizers can bear the cost of data transfer.

1.2.4 Storage

Storage occurs on servers distributed across the internet: privately owned servers, commercially managed but privately accessed storage services such as Google, and sharing-oriented services such as Facebook.

1.2.5 Access

Access is managed according to policies guided by project organizers and participants. Participatory Sensing data often include particularly sensitive information such as images of oneś family and friends, and the participantś location collected over time. While many privacy mechanisms can already be put in place, which is a crucial issue that requires continual attention and improvement to reduce the risks associated with misuse.

1.2.6 Analysis

Analysis includes a wide variety of data-processing methods, from aggregation of contributed data for display to the participant, to higher-level analysis of data to classify a participantś activities, to image processing that automatically eliminates blurry or poorly exposed images. Analysis also includes the calculation of group statistics and the integration of contributed data into statistical and spatial models that can be used to determine event location and time.

1.2.7 Feedback

Feedback may be required during event detection. Systems can use contributed data and mobile phone messaging to deliver such triggers in context-dependent ways. For example, when a person travels to a location of interest, the systems can trigger a message for the person to respond to or for the phone to answer automatically (to record sounds when the participant stops walking).

1.2.8 Visualization

Visualization goes hand-in-hand with analysis and is the step in which data are displayed in a legible format. The effectiveness of any project depends on how well its results are understood by the target audience. Excellent methods for mapping, graphing, and animation make this a rich area to develop in the context of Participatory Sensing.

1.3 Applications of Participatory Sensor Network

1. **Health and Fitness**

   Individuals can self-monitor to observe and adjust their medication, physical activity, nutrition, and interactions. Communities and health professionals can also use participatory approaches to better understand the development and effective treatment of disease. [14]

   - eCAALYX: Chronic disease management; an Android smart phone application that receives input from a BAN (a patient-wearable smart garment with wireless health sensors) and the Global Positioning System (GPS) location sensor in the smart phone and communicates over the Internet with a remote server accessible by health care professionals who are in charge of the remote monitoring and management of the older patient with multiple chronic conditions.
   - BeWell: A system that uses sensors embedded within a smart phone (gyroscope, accelerometer, microphone, camera, and digital compass) to enable a new class of personal wellbeing applications to monitor activities such as sleep, social interactions, and physical activity which in turn impact physical and mental health of an individual.[19].
   - StressSense: A system that recognizes stress from human voice using smart phone and can robustly recognize stress among multiple individuals in diverse acoustic environments.[22].
   - Ambulation: a mobility monitoring system that employs mobile phones to automatically detect a userś mobility mode using GPS data. The gathered information is critical for patients suffering from mobility-affecting chronic diseases such as MS, Parkinson, and Muscular Dystrophy. For energy efficiency, the system uses accelerometer as the means for detecting motion and triggering GPS [29].
   - AndWellness: it is a personal data collection system that uses mobile phones to collect and analyze data from on board sensors and triggered user experience samples . They have conducted a two week in-lab deployment of this system using the same campaign settings from a planned future deployment. They plan to deploy these systems for the following two applications: (1) to measure the behaviors and emotions of young breast cancer survivors and (2) to assess at risk HIV+ participants [18].

2. **Urban Planning** By lowering the complexity of creating trustworthy adhoc observing applications at the metropolitan scale, participatory sensing enables a very exciting application space for urban planning.

    - The Grand Avenue Project : Los Angeles is preparing for a two billion dollar redevelopment of a portion of its downtown, . The Norman Lear Center has invited citizen submission of design ideas for the projectś park component, receiving and publishing hundreds of such submissions. Participatory sensing tools will enable these and other organizations to initiate data collection that similarly connect people (and their data) to the planning of their own environments [3].
    - GIS-based noise planning tool: [26] Describes the tool created for the city of Belo Horizonte in Brazil, noting that noise is a major source of nuisance and, for many, an important quality of life metric. They model it in a GIS system but do not address how real world data might be gathered.A participatory sensing approach suggests that a simple service running on citizensḿobile devices, gathering and publishing basic statistics on ambient sound at regular intervals, with appropriate context checks, might be able to gather such data.Citizens could join a data-collection campaign to document noise levels in a community. They would configure simple selective sharing options to choose when and where samples are taken to calculate average sound amplitude, as well as the spatial and temporal resolution acceptable for network context tagging. A collaboratively generated city-scale analysis of noise levels at different times a day becomes feasible. When combined with participatory GIS techniques , incredible potential exists for developing important, accessible planning tools for communities of all sizes.

3. **Cultural Identity and Creative Expression**
   In 1996, Caroline Wang supplied women in a rural Chinese village with 35mm film cameras to document ẃhat is worth remembering and what needs to be changed .́ They documented the results of a lack of adequate day care for children and midwifery training for women. As a result of presenting this work in a gallery seen by political leaders, local health policies for themselves and their children improved. [33] This became the Photovoice movement. Another set of motivating applications, new for sensor networks, seeks to combine the ethos of Wang's efforts with the increased ubiquity of image capture possible with network-connected, imager-equipped, always-on mobile device. The decision to create a campaign to gather imagery might come from initiators within a community, and the Partisan architec-

ture enables the network to lend some credibility to notions of when and where media is gathered. These features could allow the scale of participation to increase without losing the sense that the location was actually ‘known’ by the gatherer.

4. **Personal Reflections on Environmental Impact and Exposure**

   - **Impacts of Climate Change** Scientists have unveiled new evidence that a changing climate is affecting our ecosystems. The latest data, however, come from an unusual source: hundreds of botanists using their mobile phones to photograph and send pictures of plants to researchers for analysis. Their ongoing study in phenology examines the link between increasing temperatures attributed to global warming and the timing of specific events in the lives of some critical plant species [17].
   - **Pollution Sources by Community Groups** Using data-gathering software run on residentsḿobile phones, the community organization initiated data collection, recruited and coordinated participants, and analyzed the resulting data to make the case that diesel truck traffic on neighborhood streets created unexpected h́ot spotsóf traffic near homes and schools [17].

5. **Transportation and Civil Infrastructure Monitoring**

   - BikeNet: a mobile sensing system for mapping the cyclist experience. The system collects and stores data about the cycling performance metrics, including current speed, average speed, and distance traveled, and calories burned over the long term. The gathered data is archived and analyzed for understanding long-term performance trends. For example, a cyclist can monitor his/her performance improvement or his/her exposure to health risks like automobile exhaust. The system also provides information to cyclists about the healthiness of a given route in terms of pollution levels, allergen levels, noise levels, and terrain roughness [13].
   - Biketastic: it is a platform that enriches this experimentation and route sharing process by letting bikers to document and share routes, ride statistics, sensed information to infer route roughness and noisiness, and media that documents ride experience . The application running on a smart phone records high-frequency GPSdata (latitude, longitude, and speed) every 1 second. The microphone and the accelerometer embedded on the phone are sampled to infer route noise

level and roughness. This will allow bikers to know the areas that have excessive noise levels, which could be indicators of large vehicles or heavy traffic. The onboard accelerometer is sampled to measure acceleration variance of the axis corresponding to the direction pointing towards earth, which gives an indication of divots and bumps. Authors evaluated the system based on feedback from expert bicyclists provided during a two-week trial period [27].

1.4 Challenges of Participatory Sensor Network

Participatory sensing applications provide numerous research challenges from the perspective of analysis. [1, 30] listed some of these challenges below:

1. **Privacy**: Since the collected data typically contains sensitive personal data (eg. location data), it is extremely important to use privacy sensitive techniques in order to perform the analysis.

2. **Low Battery Life**: Sensors, whether wearable or embedded in mobile devices, are typically operated with the use of batteries, which have limited battery life. Certain kinds of sensor data collection can drain the battery life more quickly than others (eg. GPS vs. cell tower/WiFi location tracking in a mobile phone). Therefore, it is critical to design the applications with a careful understanding of the underlying trade-offs, so that the battery life is maximized without significantly compromising the goals of the application.

3. **Data Volume**: The volume of data collected can be very large. For example, in a mobile application, one may track the location information of millions of users simultaneously. Therefore, it is useful to be able to design techniques which can compress and efficiently process the large amounts of collected data.

4. **Trustworthiness or Participant's Reliability**: Since the data are often collected through sensors which are error prone, or may be input by individuals without any verification, this leads to numerous challenges about the trustworthiness of the data collected. Furthermore, the goals of privacy and trust tend to be at odds with one another, because most privacy-preservation schemes reduce the fidelity of the data, whereas trust is based on high fidelity of the data.

5. **Real-time Processing**: Many of the applications require dynamic and real time responses. For example, applications which trigger alerts are typically

time sensitive and the responses may be real-time. The real-time aspects of such applications may create significant challenges, considering the large number of sensors which are tracked at a given time.

6. **Participant Recruitment**: Developers of a sensor data collection campaign face the challenge of identifying the appropriate set of individuals who would collect the data, for example, using their mobile phones. In most cases, the participation by individuals is voluntary, although there may be applications where an organization may have its employees be the data contributors as part of their jobs. The problem lies in identifying what subset of individuals who are interested and meet the basic requirements of being data contributors (i.e., have the right type of sensors and reside in the geographical area where the data collection is to be done in a window of time) is actually selected to contribute.

7. **Data Quality and Participant Reputation**: There may be significant differences between the data collection performance of different participants, and the performance of a participant may vary over time across different campaigns or during the course of the same campaign.

8. **Sparse Sampling and Generalization**: Consider applications that attempt to learn from collected observations and generalize by building models of system behavior where some components, interactions, processes, or constraints are not well-understood.

Chapter 2

Background Study

2.1 Overview

In this chapter, we discuss the motivation of this research work and provide the relevant research work. The major tasks of a participatory sensor network are data collection, processing and result publishing. We know that the data collection is often open to a large population. Therefore, the main challenge is to know which of the reported observations are true and which are not. This challenge is also known as *trust* or *reliability* of a participant. In this chapter, Section 2.1 discusses the motivation of this research and in Section 2.2, we provides some relevant research results.

2.2 Motivation

The openness of participatory sensing systems provides a tremendous amount of power consumption in collecting information from a wide variety of sources, and distilling this information for data mining purposes. However, it is this very openness in data collection, which also leads to numerous questions about the quality, credibility, integrity, and trustworthiness of the collected information [5, 9, 15, 16]. Furthermore, the goals of privacy and trust would seem to be at odds with one another, because all privacy-preservation mechanisms reduce the fidelity of the data for the end-user, whereas the end-user trust is dependent on high fidelity of the data. Numerous questions may arise in this respect:

- How do we know that the information available to the end user is correct, truthful and trustworthy?
- When multiple sources provide conflicting information, how do we know which one is trustworthy?

- Have errors been generated in the process of data collection, because of inaccuracy or hardware errors?

The errors which arise during hardware collection are inherent to the device used, and their effect can be ameliorated to some extent by careful design of the underlying application. For example, the LiveCompare [6] application, which is used for comparison shopping of grocery products, works by allowing individuals to transmit photographs taken in stores of grocery products, and then presents similar pictures of products taken in nearby stores. The approach allows the transmitting of product photos taken by individual users of competing products, but does not automatically try to extract the pricing information from the price tags in the photograph. This is because the extraction process is known to be error-prone, and this design helps avoid the inaccuracy of reporting the pricing of competing products. It also avoids manual user input about the product which reduces error and maximizes trustworthiness.
A more critical question about trustworthiness arises when the data is collected through the actions of end users. In such cases, the user responses may have an inherent level of errors which may need to be evaluated for their trustworthiness. The issue of truthfulness and trust arises more generally in any kind of application, where the ability to contribute information is open. Such openness creates challenging trade-off which increases information availability at the expense of trust. Aside from the social and participatory sensing platforms, any web enabled platforms which allow the free contribution of information may face such challenges.

2.3 Related Work

Here, we discuss some research work on the reliability estimation of the nodes in a participatory sensing network.

- Certificate Based Reliability or Trust
  For the case of specific kinds of data such as location data, a variety of methods can be used in order to verify the truthfulness of the location of a mobile device [21]. The key idea is that time-stamped location certificates signed by wireless infrastructure are issued to co-located mobile devices. A user can collect certificates and later provide them to a remote party as verifiable proof of his or her location at a specific time.The major drawback of this approach is that the applicability of these infrastructure based approaches for mobile sensing is limited as cooperating infrastructure may not be present in remote or hostile environments of particular interest to

some applications. Furthermore, such an approach can be used only for particular kinds of data such as location data.

- Platform Attestation
  In the context of participatory sensing, where raw sensor data is collected and transmitted, a basic approach for ensuring the integrity of the content has been proposed in [9], which guards whether the data produced by a sensor has been maliciously altered by the users. Thus, this approach relies on the approach of platform attestation which vouches that the software running on the peripheral has not been modified in an unintended manner. This kind of approach is more useful for sensors in which the end data is produced by the device itself, and an automated software can be used for detection of malicious modification. In essence, the approach allows the trusted sensing peripherals to sign their raw readings, which allows the remote entity to verify that the data was indeed produced by the device itself and not modified by the user. Trusted Platform Module (TPM) hardware [15], commonly provided in commodity PCS, can be leveraged to help provide this assurance. To address the problem of protecting the privacy of data contributors, techniques such as requiring explicit authorization for applications to access local resources and formulating and enforcing access control policies can be used. A TPM is a relatively inexpensive hardware component used to facilitate building trusted software systems. It is possible to leverage the TPM functionality of attesting to the integrity of software running on a device to a remote verifier. The TPM can attest to the software platform running on the machine by providing a signed quote of its PCR(s) in response to a challenge from a remote verifier. In many cases, user actions may change the data (such as the cropping of an image), but this may not actually affect the trust of the underlying data. The work in [16] proposes YouProve, which is a partnership between a mobile device's trusted hardware and software that allows un-trusted client applications to directly control the fidelity of data they upload and services to verify that the meaning of source data is preserved. The approach relies on trusted analysis of derived data, which generates statements comparing the content of a derived data item to its source. For example, the work in [16]tests the effectiveness of the method on a variety of modifications on audio and photo data, and shows that it is possible to verify which modifications may change the meaning of the underlying content.

- Heuristic Based Approach
  In this context, the problem of trustworthiness has been studied for re-

solving multiple, conflicting information provision on the web. The earliest work in this regard was proposed in [38], where the problem of studying conflicting information from different providers was studied [38]. Subsequently, the problem of studying trustworthiness in more general dynamic contexts was studied in [7, 8].

- Likelihood Reliability
  A number of recent methods [20, 34, 35, 37] address this issue, in which a consistency model is constructed in order to measure the trust in user responses in a participatory sensing environment. The key idea is that untrustworthy responses from users are more likely to be different from one another, whereas truthful methods are more likely to be consistent with one another. This broad principle is used in order to model the likelihood of participant reliability in social sensing with the use of a Bayesian approach [34, 35]. A system called Apollo [20] has been proposed in this context in order to distill the likely truth from noisy social streams.

- Fuzzy Logic
  In [2] authors present an application agnostic framework to evaluate trust in social participatory sensing systems. the system independently assesses the quality of the data and the trustworthiness of the participants and combines these metrics using fuzzy logic to arrive at a comprehensive trust rating for each contribution. These trust ratings are then used to calculate and update the reputation score of participants. By adopting a fuzzy approach, this system is able to concretely quantify uncertain and imprecise information, such as trust, which is normally expressed by linguistic terms rather than numerical values.

- Streaming Approach
  In this paper [36] authors present a streaming approach to solve the truth estimation problem in crowdsourcing applications. They consider a category of crowdsourcing applications where a group of individuals volunteer (or are recruited to) share certain observations or measurements about the physical world. Ascertaining the correctness of reported observations is a key challenge in such applications, referred to as the truth estimation problem. This problem is made difficult by the fact that the reliability of individual sources is usually unknown a priori, since any concerned citizen may, in principle, participate. Moreover, the timescales of crowdsourcing campaigns of interest can be as small as a few hours or days, which does not offer enough history for a reputation system to converge. Fact-finding algorithms are used to solve this problem by iteratively assessing the credibility

of sources and their claims in the absence of reputation scores. Such algorithms, however, operate on the entire dataset of reported observations in a batch fashion, which makes them less suited to applications where new observations arrive continually. Authors describe a streaming fact-finder that recursively updates previous estimates based on new data. The recursive algorithm solves an expectation maximization (EM) problem to determine the odds of correctness of different observations.

- Semi-supervised Methods
  Accessing online information from various data sources has become a necessary part of our everyday life. Unfortunately such information is not always trustworthy, as different sources are of very different qualities and often provide inaccurate and conflicting information. Existing approaches attack this problem using unsupervised learning methods, and try to infer the confidence of the data value and trustworthiness of each source from each other by assuming values provided by more sources are more accurate. However, because false values can be widespread through copying among different sources and out-of-date data often overwhelm up-to-date data, such bootstrapping methods are often ineffective.In this paper [39] authors propose a semi-supervised approach that finds true values with the help of ground truth data. Such ground truth data, even in very small amount, can greatly help to identify trustworthy data sources. Unlike existing studies that only provide iterative algorithms, the optimal solution to the problem can be derived and an iterative algorithm that converges to it can be provided.

Chapter 3

Problem Domain

3.1 Overview

In this chapter, we define the problem and discuss some preliminaries relevant to our research problem.

3.2 Preliminaries

Let us consider a participatory sensing network model where a group of M participants, $S_1 \ldots$ S_M, make individual observations about a set of N events $C_1 \ldots$ C_N. Therefore, total reported observations are $M \times N$. It is very challenging to find if the reported observations are true or false.

It will be more challenging if source reliability and the probability of event to be true are unknown. We define *source reliability* a as the probability that the participant reports correct observation and z as the probability that the event to be true. Figure 3.1 illustrates a system model. Our goal is to estimate a for z. Let us assume that a and z both are unknown.

3.3 Population Based Reliability Estimation (PBRE)

Population Based Reliability Estimation (PBRE) uses a set of reliability instead of single reliability. In our approach, we call this set of reliability as P and use Genetic Algorithm to estimate the reliability of participant. Population-based methods keep around a sample of candidate solutions rather than a single candidate solution. Each of the solutions is involved in tweaking and quality assessment. Most of the population-based methods steal concepts from biology, genetics or evolution. An algorithm chosen from this collection is known as an Evolutionary Algorithm(EA). Common EAs include the Genetic Algorithm (GA)

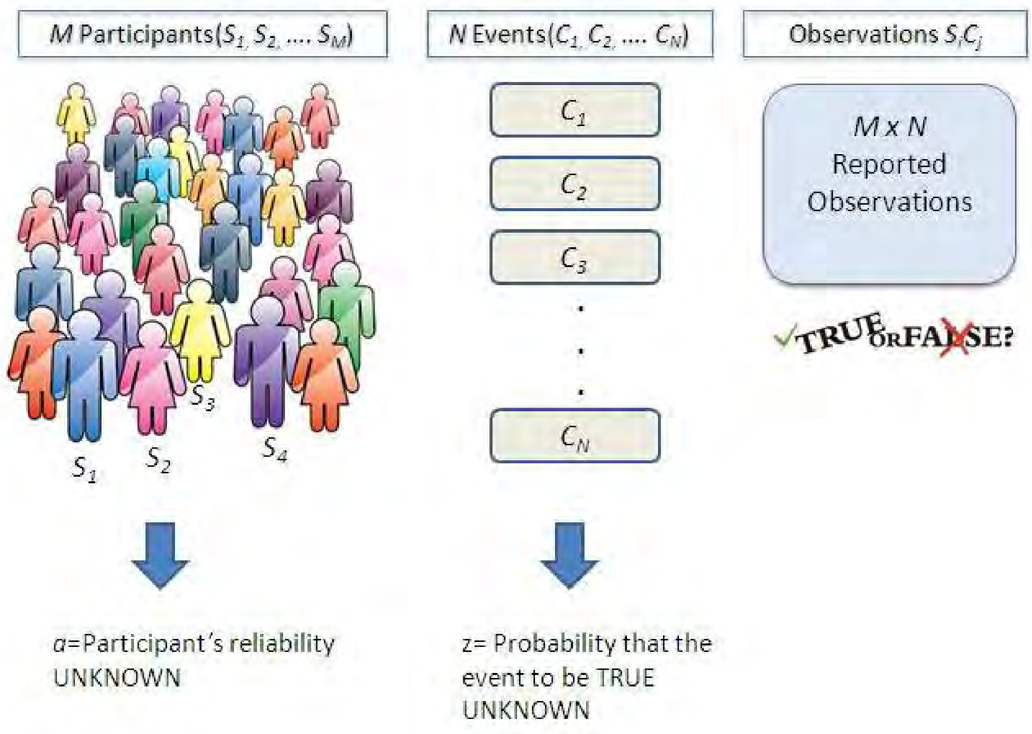

Figure 3.1: A System Model

and Evolutionary Strategies(ES). Table 3.1 describes some common terms in EA.

Terms	**Description**
individual	a candidate solution
child and parent	a *child* is the tweaked copy of a candidate solution(its parent)
population	set of candidate solutions
fitness	quality
fitness landscape	quality function
fitness assessment or evaluation	computing fitness of an individual
selection	picking individuals based on their fitness
recombination or crossover	A special Tweak which takes two parents, swaps sections of them, and (usually) produces two children.
breeding	producing one or more children from a population of parents through an iterated process of selection and tweaking(typically mutation or recombination)

Table 3.1: Terminology Used in Evolutionary Algorithm

3.3.1 Genetic Algorithm

Genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic (also sometimes called a metaheuristic) is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as *inheritance*, *mutation*, *selection*, and *crossover*.

3.3.2 Steps of Basic Genetic Algorithm

There are two steps in basic Genetic Algorithm which are listed below:

- Constructs an initial population,
- Then iterates through three procedures:
    - Fitness Assessment: First, it assesses the fitness of all the individuals in the population.
    - Breeding: Second, it uses this fitness information to breed a new population of children. It begins with an empty population of children. Then produces two children as follows:
        - select two parents from the original population
        - copy them
        - cross them over with one another
        - mutate the results
    - Joining: Third, it joins the parents and children in some fashion to form a new next-generation population,

3.4 Our Approach

3.4.1 Population Based System Model

Here, we consider a participatory sensing application model where a group of M participants, S_1 to S_M, make individual observations about a set of N events C_1 to C_N. Probability that participant S_i reports a true event when the event is actually true is a_i and the probability that participant S_i reports a true event when the event is actually false is b_i. θ is the set of a_i and b_i e.i. $\{a_i, b_i\}$.

Here, $1< = i< = M$ and $1< = j< = N$. We call the set of θ as P which is a set of reliability. z_j is the probability that the event C_j is indeed true. Figure 3.2 illustrates the model.

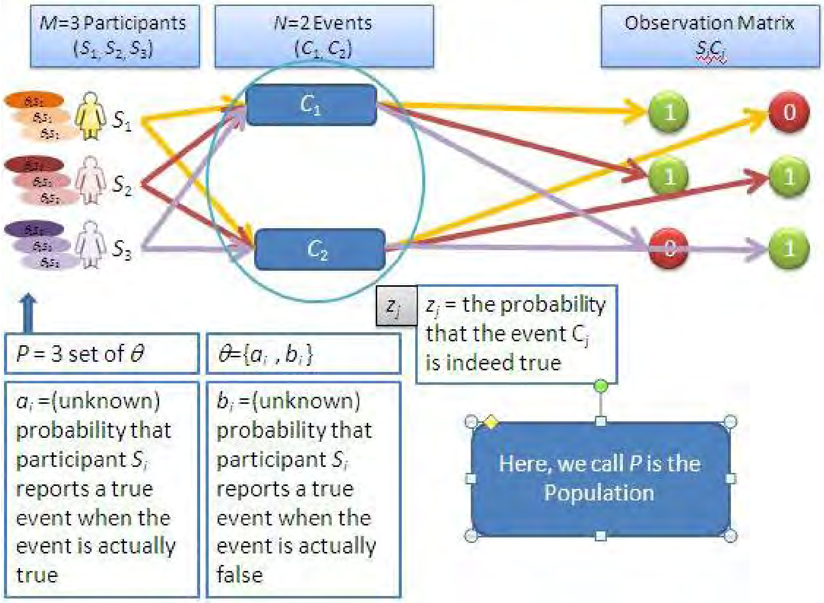
M=3 Participants (S1, S2, S3)
N=2 Events (C1, C2)
Observation Matrix SiCj
S1
S2
S3
C1
C2
1
0
1
1
0
1
zj
zj = the probability that the event Cj is indeed true
P = 3 set of θ
θ={ai , bi}
ai =(unknown) probability that participant Si reports a true event when the event is actually true
bi =(unknown) probability that participant Si reports a true event when the event is actually false
Here, we call P is the Population

Figure 3.2: Population Based System Model

3.4.2 Detail Methodology

1. **Step1**: We initialize and build popoulation as following:

   - Firstly, we initialize M, N
   - Secondly, we initialize SC matrix = [0,1]. Here, 0 = When participant reports an event as false, and 1 = When participant reports an event as true.
   - Thirdly, we initialize d=[random value from 0 to 1] = Overall bias on event to be true.
   - Finally, P= The set of θ ={a, b}= [any value between 0 to 1].

2. **Step2**: We calculate z_j as following:
   p(z_j |X_j , θ) is the conditional probability z_j to be true given the observation matrix X_j related to the j^{th} and current estimate of θ .
   Here, we define d as overall bias on event to be true. The value of d is any real value between 0 to 1. If the value is greater than 0.5, we consider the event to be true and if the value is below or equal to 0.5, we consider the event to be false. For example, if d=0.7, it implies that the probability of event to be true is true. Using this bias factor, we try to converge z_j towards d.

3. **Step3**: Fitness Function
   Then we assess fitness of the P, set of reliability. We compare P with the best reliability. The ***target reliability*** or $target\_a_i$ is computed as below:

$target\_a_i = \sum_{i=1}^{M}(\sum_{j=1}^{N} = \frac{SC(i,j)}{N})$

For example, in ideal case, the probability of all events to be true, z_j =1. Let us consider, there are 2 events and 3 participants which is illustrated in Figure 3.3. Participant S_1 reports event C_1 as true and event S_2 as false. Therefore, $target\_a_1 = \frac{SC(1,1)+SC(1,2)}{2} = \frac{1+0}{2} = 0.5$

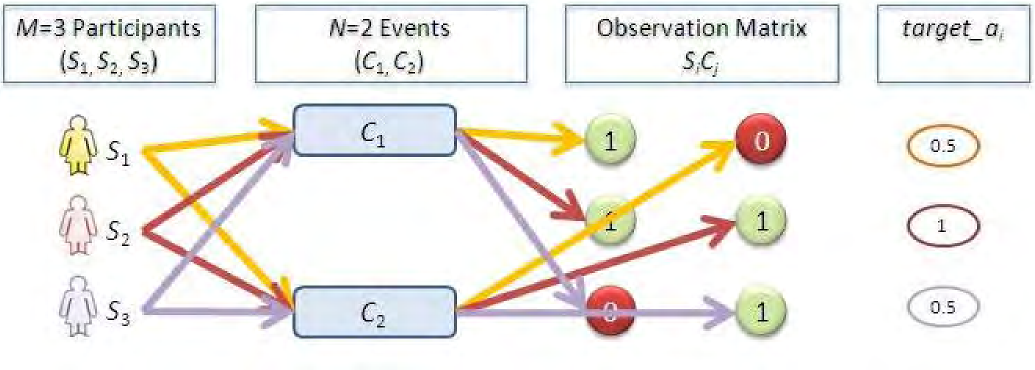

Figure 3.3: Calculating *best reliability* or $target\_a_i$

Now, the objective is to select fittest a_i from P that helps to converge z_j. We take the fittest value from the initial set of values of a_i using fitness function. The similar calculation can be done for finding fit_b_i. We call this fittest value as ***fit reliability*** or fit_a_i or fit_b_i. Now, we take two types of fitness functions Fit_Parent and Replace_Parent:

(a) Type 1 : **Fit_Parent**

The idea of fitness function Fit_Parent is to select fit_a_i from set of a_i of S_i. Here, fit_a_i is that one which is closest to $target\_a_i$.

Figure 3.5 is an illustrative example of Fit_Parent computation.

Here, we initialize three sets of a_1 for participant S_1 e.i. 0.3, 0.1 and 0.8. From previous Figure 3.3, we see that the $target\_a_1$ is 0.5. Therefore, the closest a_1 e.i. fit_a_1 is 0.3. Similarly, we calculate for participant S_2 and S_3 which are a_2=0.8 and a_3=0.6 respectively.

(b) Type 2 : **Replace_Parent**

In Replace_Parent, instead of selecting one fit_a_i from every participant S_i's P, we select the full set of a_i which is closest to set of $target\_a_i$.

Now, we give an illustrative example of Replace_Parent in Figure 3.5.

Here, we initialize three sets of a_i for each participant S_1 e.i. (a_{11}, a_{12}, a_{13}) = (0.3, 0.1, 0.8), for S_2 it is (a_{21}, a_{22}, a_{23})=(0.8, 0.4, 0.5) and for S_3 it is (a_{31}, a_{32}, a_{33})= (0.8, 0.5,0.9). Now, we make another set taking the

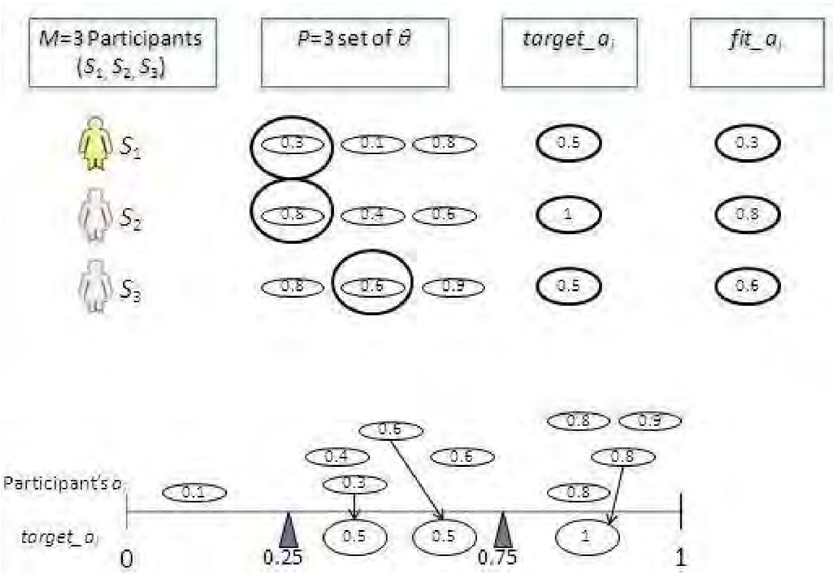

Figure 3.4: Fit_Parent Computation

first a_i from each S_i e.i. (a_{11}, a_{21}, a_{31})=(0.3, 0.8, 0.8) and similarly (a_{12}, a_{22}, a_{32})=(0.1, 0.4, 0.6) and (a_{13}, a_{23}, a_{33})=(0.8, 0.6,0.9). Our $target\_a_i$ = (0.5, 1, 0.5). Therefore, we find that there are 2 fit_a_is in the first set, similarly 1 and 0 fit_a_i for the second and the third set. Finally, we take the first set as the set of fit_a_i.

4. **Step 4**: *Breeding*

   Now, the objective is to generate new child θ from parent θ . We choose recombination technique [12] as breeding technique. This new values are called child values $anew_i$ and $bnew_i$, where,

   $anew_i = \alpha a_i + (1 - \alpha) b_i$

   $bnew_i = \beta b_i + (1 - \beta) a_i$

   Where, α = random value between 0 to 1 and
   β = random value between 0 to 1

5. **Step 5**: *Joining*

   We form the next generation parent by using new children.

   Joining Formula

   $a_i = anew_i$

   $b_i = bnew_i$

6. **Step 6**: *Error percentage of participant's reliability*

   We calculate *Error percentage of participant's reliability* by below formula,

   $$\frac{\textit{Total number of converged reliability}}{\textit{Total number of reliability}}$$

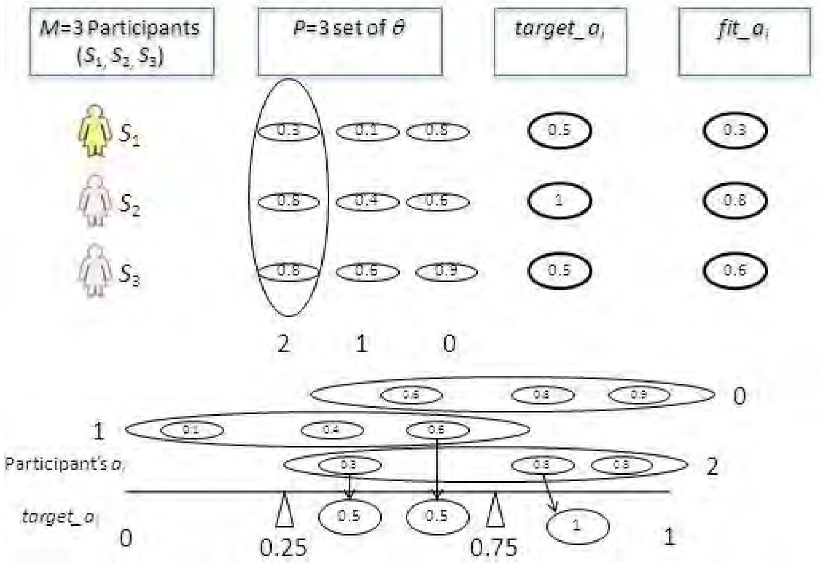

Figure 3.5: Replace_Parent Computation

3.5 Algorithm

In this section, we provide the detail steps of our computation.

3.5.1 Population Based Reliability Estimation(PBRE)

```
procedure PBRE
Initialize $M$, $N$, $P$, $d$
Initialize observation matrix $SC$ with random values either 0 or 1
Initialize $\theta$ = {a,b} with random values between 0 and 1
Initialize $fit\_a(i)$ as NULL
Initialize $zcount$=0 // $z_j$ convergence metric
Calculate $target\_a(i) = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{SC(i,j)}{N}$
while $z_j$ does not converge do
    for i=1:M do
        $a(i, P+1) = fit\_a(i)$ //add $fit\_a(i)$ in the population of a(i)
    end for
    for $j$=1:N do
        for $K$=1:P+1 do
            $z(j, K)$
            if $z(j, K) = d$ do
                $z_j$ convergence counter
            end if
            end for
        end for
                // Fitness Function
        assesFitness-Fit_Parent() or assesFitness-Replace_Parent()
        breed() //breeding Function
        join() // Joining Function
    reliabilityEstimation()
end while
```

3.5.2 Computing probability that the event z_j is true or false

```
procedure $z(j, K)$
begin
// Calculate $at$=conditional probability that participant's observation is true given $\theta$ and event $z$=1.
$at(j,K) = \sum_{i=1}^{M} a(i,K)^{SC(i,j)}(1 - a(i,K))^{(1-SC(i,j))}$
// Calculate $bt$=conditional probability that participant's observation is true given $\theta$ and event $z$=0.
$bt(j,K) = \sum_{i=1}^{M} b(i,K)^{SC(i,j)}(1 - b(i,K))^{(1-SC(i,j))}$
$z(j,K) = \frac{at1(j,K) \times d}{at1(j,K) \times d + bt1(j,K) \times (1-d)}$
end procedure
```

3.5.3 Fitness Function

```
procedure assesFitness-Fit_Parent()
begin
//Select closest $a$ to $target\_a$ as $fit\_a$
for $i$ = 1:M do
    for $K$ = 1:P + 1 do
        if (0<= $target\_a(i)$ <= 0.25 AND 0<= $a(i,K)$ <=0.25) OR (0.25<$target\_a(i)$ <= 0.75 AND 0.25 <$a(i,K)$ <=0.75) OR (0.75<$target\_a(i)$ <=1 AND 0.25 <$a(i,K)$ <=1)then
            $fit\_a(i) = a(i)$
        end if
    end for
end for
```

```
return $fit\_a$
end procedure

procedure assesFitness-Replace_Parent()
begin
//Select closest set of $a$ to set of $target\_a$ as $fit\_a$
for $i$ = 1:M do
    for $K$ = 1:P + 1 do
        if (0<= $target\_a(i)$ <= 0.25 AND 0<= $a(i, K)$ <=0.25) OR (0.25< $target\_a(i)$ <= 0.75 AND 0.25 < $a(i, K)$ <=0.75) OR (0.75< $target\_a(i)$ <=1 AND 0.25 < $a(i, K)$ <=1) then
            $count(K) + +$
        end if
    end for
end for
$best$=0
for $K = 1 : P + 1$ do
    if $count(K)$>$best$ then
        $best = count(K)$
        $L = K$
    end if
end for
for $i = 1 : M$ do
    $fit\_a(i) = a(i, L)$
end for
return $fit\_a$
end procedure
```

3.5.4 Breeding Function

```
procedure breed()
begin
//breed using recombine - multiply
for i=1:M do
    for i=1:P do
        $t(i, K) = \alpha \times a(i, K) + (1 - \alpha) \times b(i, K) // newchild1$
      //$\alpha$ = random number between 0 and 1
        $s(i, K) = \beta \times b(i, K) + (1 - \beta) \times a(i, K) // newchild2$
      //$\beta$ = random number between 0 and 1
    end for
end for
return $t, s$
end procedure
```

3.5.5 Joining Function

```
procedure join()
begin
//Replace new children with parents
for $i = 1 : M$ do
    for $K = 1 : P$ do
        $a(i, K) = t(i, K)$
        $b(i, K) = s(i, K)$
    end for
end for
return $a, b$
end procedure
```

3.5.6 Reliability Estimation

```
reliabilityEstimation()
begin
for $i = 1 : M$ do
    if(0< =target_a(i)< = 0.25AND0< =fit_a(i, K)< = 0.25)OR(0.25<target_a(i)< = 0.75AND0.25<fit_a(i, K)< = 0.75)OR(0.75<target_a(i)< = 1AND0.25<fit_a(i, K)< = 1) do
        $truecount++$ // count of correct reliability estimation
    end if
end for
$error = (1 - \frac{truecount}{M}) \times 100$ // percentage of error reliability estimation
end procedure
```

3.6 Summary

In this section, we summarize the whole PBRE Algorithm.

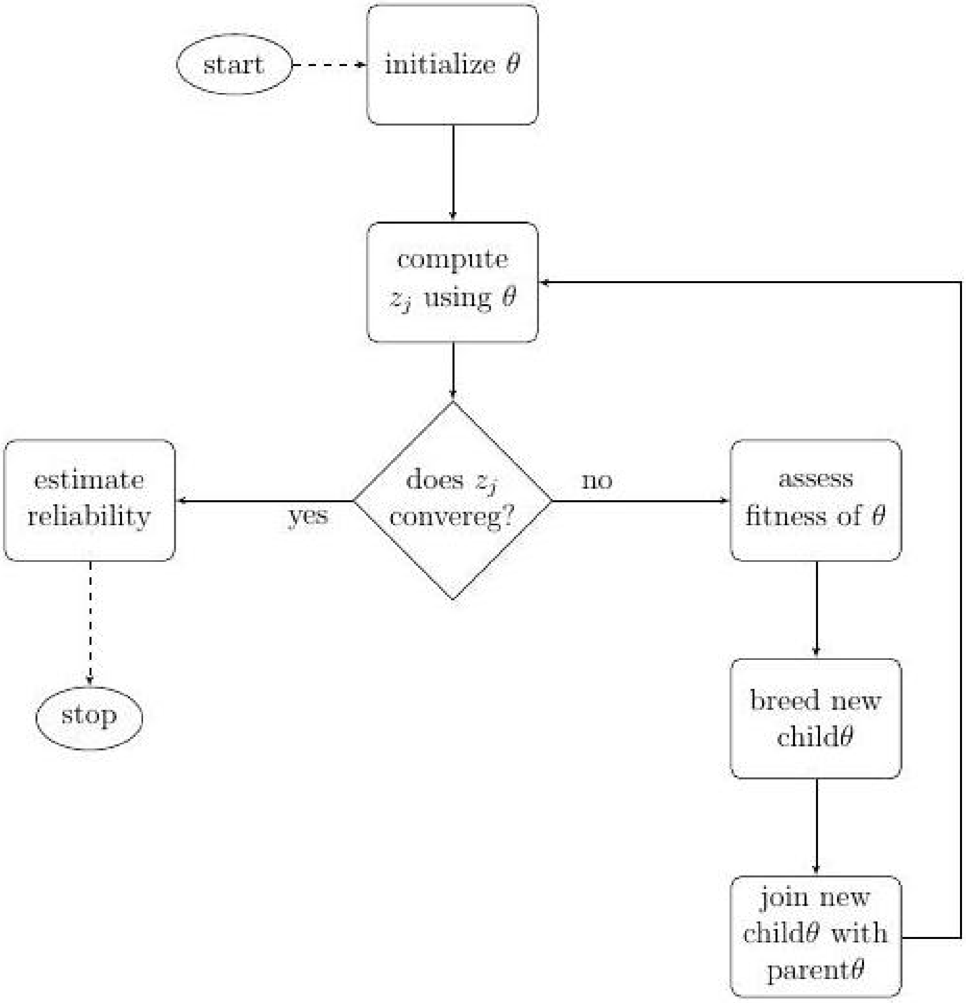

Figure 3.6: Flow Chart

Chapter 4

Experimental Results

4.1 Overview

In this chapter, we have provided our experimental results. We have also compared our results with another relevant algorithm Expectation Maximization [37].

4.2 Testbed Description

The simulation of **PBRE** runs on 1.58 *GHz Intel Core 2 Duo Processor* with *2 GB* memory. The operating system used to run the simulation is *Windows XP Professional Version 2002*. We have simulated **PBRE** using Visual Basic for Applications (VBA) which is closely related to Visual Basic and uses the Visual Basic Runtime Library.

4.2.1 Simulation Metric

In this section, we have presented the simulation metrics as performance measures to show the effectiveness of our **PBRE**. The simulation metrics are as follows:

1. The *error percentage of participant's reliability* denotes the estimation of reliability of a participant to a converged event z.

2. The *convergence rate* denotes how quickly participant can provide the correct event. It is computed by the participantś reliability divided by the total iteration needed to converge.

4.2.2 Simulation Settings

To set up a simulation environment, at first we need to specify some parameters for participatory sensing. In the Table 4.1 we have described simulation settings.

Parameters	**Value**
Participant number, M	30-900
Event number, N	2-10
Set of reliability per participant or population number , P	2-15
Observation Matrix, SC	0,1
Probability that a participant S_i reports a true event when the event is actually true, a	any value from 0 to 1
Probability that a participant S_i reports a true event when the event is actually false, b	any value from 0 to 1
Probability that the event C_j is indeed true ,d	0.7
α	any real value from 0 to 1
β	any real value from 0 to 1

Table 4.1: Simulation Settings

4.3 Experimental Results

In this section, we have carried out experiments using simulation to evaluate the performance of the proposed PBRE scheme in terms of estimation accuracy of the probability that a participant is right or a measured variable is true compared to another existing reference method Expectation Maximization. We have taken the average of 10 experiments involving the same sources and variables. We have shown that the new algorithm performs better.

4.3.1 Expectation Maximization

It is a general algorithm for finding the maximum likelihood estimates of parameters in a statistic model, where the data are "incomplete" or the likelihood function involves latent variables. Intuitively, what EM does is iteratively "completes" the data by "guessing" the values of hidden variables then re-estimates the parameters by using the guessed values as true values. Let us consider, an observed data set =X, one should judiciously choose the set of latent or missing values Z, and a vector of unknown parameters θ, then formulate a likelihood function L(θ;X;Z) = p(X;Z | θ), such that the maximum likelihood estimate (MLE)

of the unknown parameters θ is decided by:

L(θ;X) = p(X | θ)

Once the formulation is complete, the EM algorithm finds the maximum likelihood estimate by iteratively performing the following steps: E-step: Compute the expected log likelihood function where the expectation is taken with respect to the computed conditional distribution of the latent variables given the current settings and observed data.

$Q(\theta \mid \theta^{t})$

M-step: Find the parameters that maximize the Q function in the E-step to be used as the estimate of for the next iteration.

$\theta^{(t+1)} = \text{argmax } Q(\theta \mid \theta^{t})$

4.3.2 Error percentage of participant's reliability

_A_: For Variable Number of Participants

We have compared the estimation accuracy of PBRE (Fit_Parent and Replace_Parent) and Expectation Maximization(EM) scheme by varying the number of participants in the system.

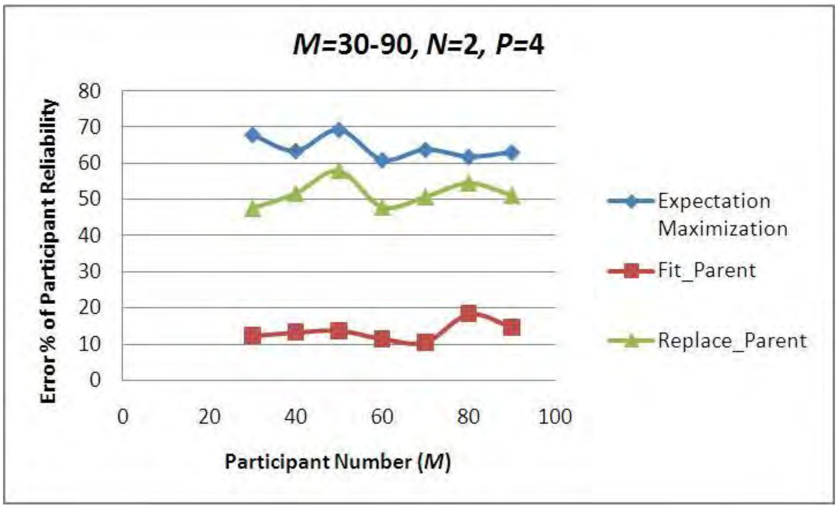

Figure 4.1: Error estimation for M=30-50, N=2, P=2

In Figure 4.1, we have varied participants from 30 to 90. We have taken 2 events and 2 sets of reliability per person. We have observed that, PBRE has a lower estimation error in participant reliability compared to EM scheme. Between two schemes of PBRE, Fit_Parent and Replace_Parent, Fit_Parent has much lower estimation error. This is because Fit_Parent takes only the fit values whereas Replace_Parent takes the fit set of values.

We have run experiments for the increased number of participants from 300 to 900. We have increased the number in the set of reliability per person to 15.

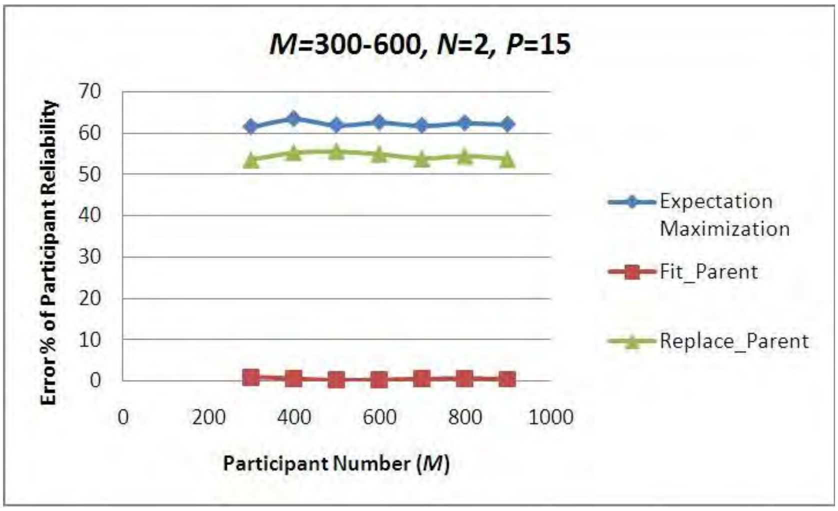

Figure 4.2: Error estimation for M=300-900, N=2, P=15

Event number is same as before e.i. 2. Now, in Figure 4.2, we have observed that the error percentage decreases for Fit_Parent to 1% compared to 10 to 15% in Figure 4.1 for participants with 4 set of reliability per person. The reason behind this decline is the increased number in the set of reliability.

B: For Variable Number of Events
Now, we have compared the results by varying the number of events from 2 to 10.

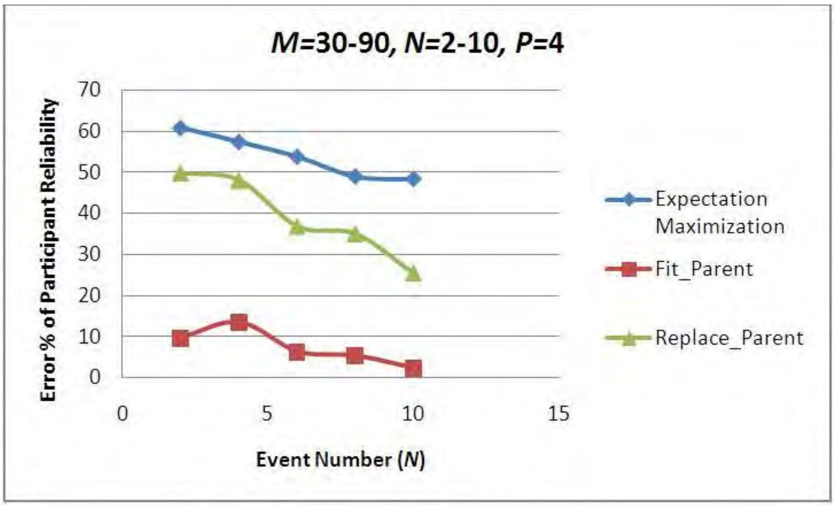

Figure 4.3: Error estimation for M=50, N=2-10, P=4

In Figure 4.3, we have run experiments for 50 participants, 2-10 events and 4 set of reliability per person. Here also, PBRE shows better results than EM. Because, when the event number increases, $target\_a_i$ decreases (line 6, procedure PBRE). Therefore, there are more matches of a_i as fit_a_i to $target\_a_i$.

We have examined the results for the increased number of participants 600 and

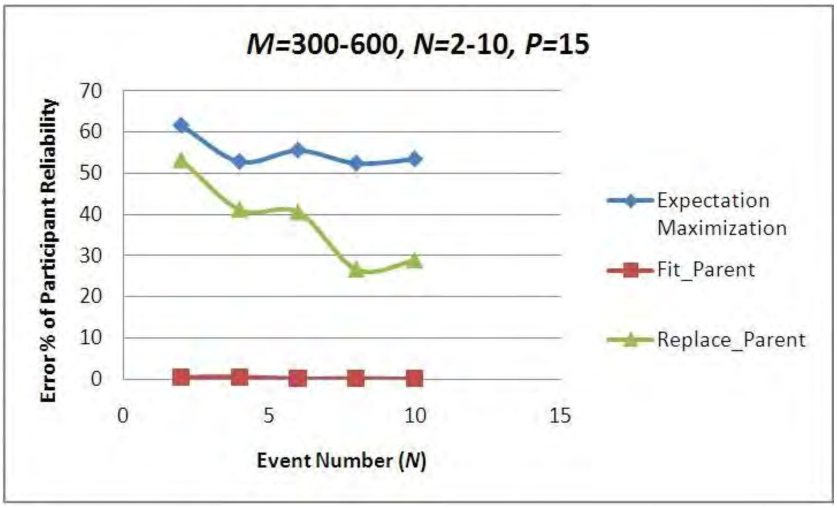

Figure 4.4: Error estimation for M=600, N=2-10, P=15

the increased number of set of reliability 15 in Figure 4.4. Here, we have analyzed that the error percentage decreases for Fit_Parent (0-1%) compared to Figure 4.3(2-15%) whereas the error percentage for Replace_Parent and EM for Figure 4.3 and 4.4 remain the same. Here, we have found no impact on the increased number in participants or the increased number in the set of reliability. This is because, Fit_Parent takes only the fit values whereas Replace_Parent takes the fit set of values.

C: For Variable Number of the Set of Reliability

Now, we have compared the estimation accuracy of Fit_Parent and Replace_Parent scheme by varying the number of set of reliability per person. The number of set of reliability per person varies from 2 to 15.

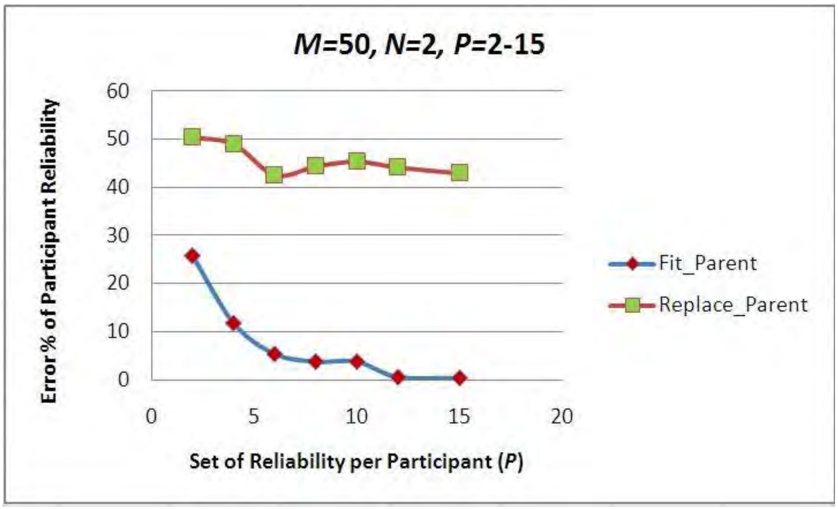

Figure 4.5: Error estimation for M=50, N=2, P=2-15

In Figure 4.5 for 50 participants and 2 events, we have observed that, for increased number of set of reliability, error estimation for both Fit_Parent and Replace_Parent decreases. However, in case of Fit_Parent it drops from 25% to 5% whereas for Replace_Parent it is around 45%. This is because Fit_Parent takes only the fit values whereas Replace_Parent takes the fit set of values.

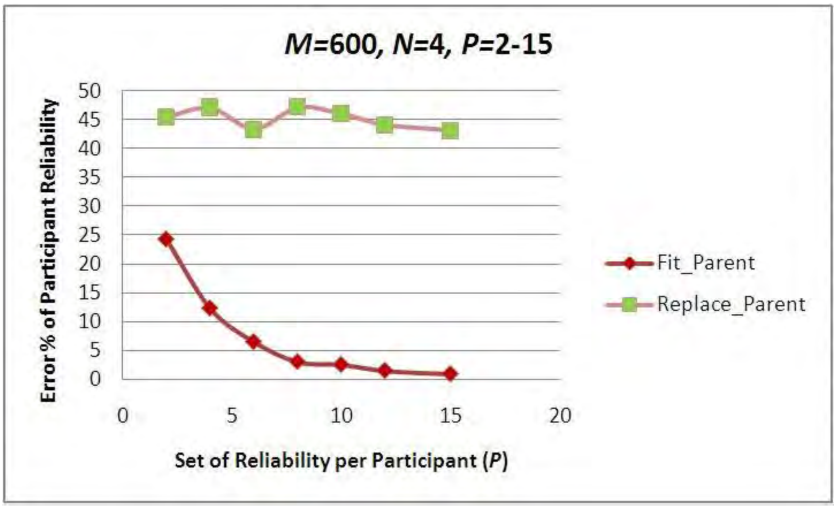

Figure 4.6: Error estimation for M=600, N=4, P=2-15

We have raised the number of participants to 600 and event to 4 in Figure 4.6. Here, we have analyzed that the error estimation for Replace_Parent keeps around 45% whereas in Figure 4.5, it drops from 50% to 40%. This is because, the event number has been increased and when the event number increases, $target\_a_i$ decreases (line 6, procedure PBRE). Therefore, there are more matches of a_i as fit_a_i to $target\_a_i$. But for Fit_Parent strategy, it shows the minimal impact.

4.3.3 Convergence Rate

_A_: For Variable Number of Participants

We have compared the convergence vs. estimation accuracy of PBRE and EM scheme by varying the number of participants from 30 to 80. Event number is fixed at 2 and the set of reliability per person is 4.

In Figure 4.7, we have observed that the convergence rate for PBRE is lower than the EM. This is because, though PBRE has the lower error percentage of reliability than EM, it iterates more than EM to converge. Here, the convergence rate for Fit_Parent, Replace_Parent and EM are 0-2, 3.5-4.5 and 8-10 respectively.

_B_: For Variable Number of Events

Now, we have examined results by varying number of events from 2 to 10. Participant number is fixed at 50 and set of reliability per person is 4.

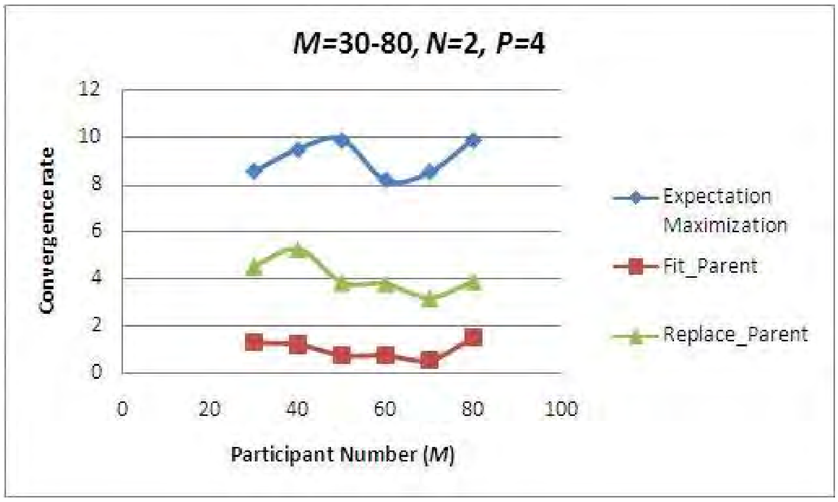

Figure 4.7: Convergence rate for M=30-80, N=2, P=4

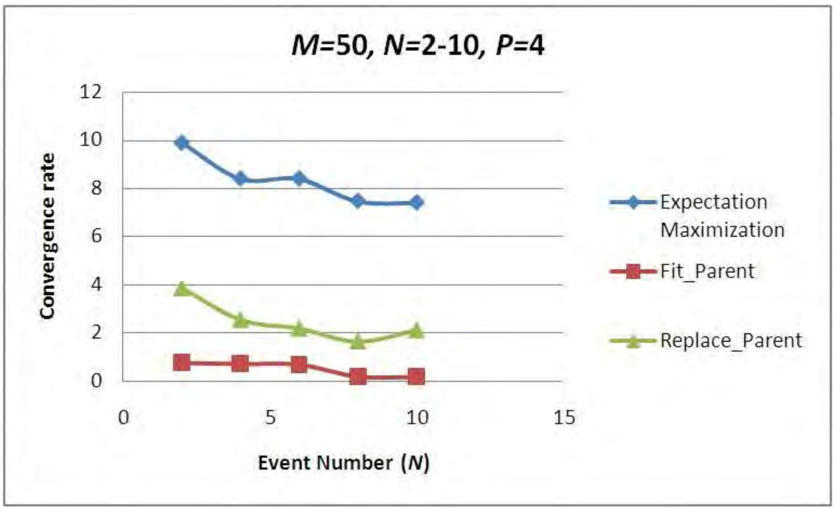

Figure 4.8: Convergence rate for M=50, N=2-10, P=4

In Figure 4.8, we have found that the convergence rate for PBRE is lower than the EM. This is because, though PBRE has the lower error percentage of reliability than EM, it iterates more than EM to converge. Here, convergence rate for Fit_Parent, Replace_Parent and EM are 0-1, 2-4 and 7.5-10 respectively. We have also observed that the rate in Figure 4.8 is lower than the rate in Figure 4.7 for increased number of events. Because, when the event number increases, $target\_a_i$ decreases (line 6, procedure PBRE). Therefore, there are more matches of a_i as fit_a_i to $target\_a_i$ and when there is a set of reliability(PBRE) instead of one (EM), to find fit_a_i from set of a_i is less error-prone.

C: For Variable Number of the Set of Reliability

We have compared results by varying the number in the set of reliability per person from 4 to 12 keeping the fixed number of participants at 50 and the fixed

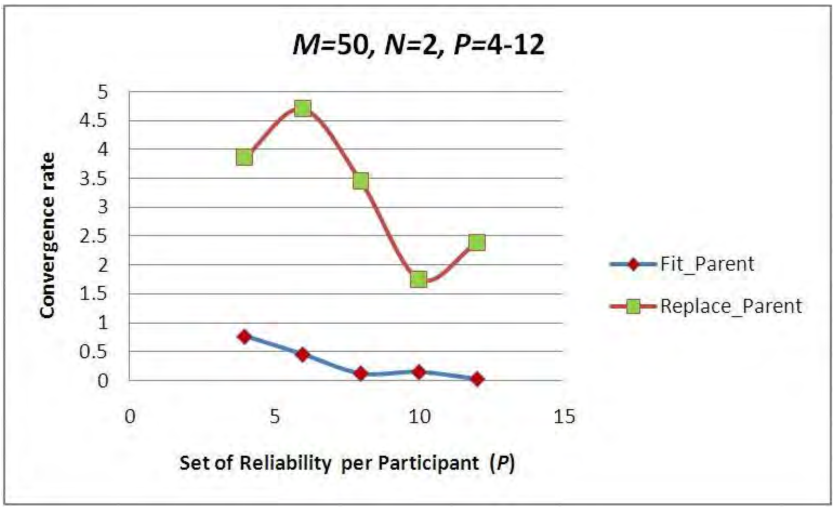

Figure 4.9: Convergence rate for M=50, N=2, P=4-12

number of events at 2.

In Figure 4.9, we have found that the convergence rate for Fit_Parent is lower than the Replace_Parent. Convergence rate for Fit_Parent and Replace_Parent are 0-1 and 4.5-1.5. This is because Fit_Parent takes only the fit values whereas Replace_Parent takes the fit set of values.

Chapter 5

Conclusion

5.1 Overview

In this chapter, we draw the conclusion of our project followed by some future research directions.

5.2 Summary

In this research, we study reliability of event detection in a participatory sensor network. We propose Population Based Reliability Estimation(PBRE) which starts with assuming a set of reliability. Then, we compute conditional probability of event to be true with these set of reliability until it converges. We have used the concept from genetics or evolution. An algorithm chosen from this collection is known as an evolutionary algorithm. More specifically, we have used genetic algorithm to estimate the reliability by iterating through fitness assessment, breeding and joining. Two types of fitness assessment are defined e.i. Fit_Parent and Replace_Parent. We vary the number of participants, number of events and number of set of reliability per person. The metrics for performance measurement is error percentage of participant's reliability and the convergence rate. We have compared the results with another Expectation Maximization in our defined environment and find that our approach provides better results.
We also observe that, for different experiments, *Fit_Parent* is less error prone because it uses the fit reliability whereas *Replace_Parent* uses the whole set of fit reliability.

5.3 Future Work

Reliability estimation in participatory sensor network is a challenging research field. We would like to extend our work by experimenting with real-life data. We also want to do experiments by varying overall bias d and the probability that a person reports a true event when the event is actually false, b.

Bibliography

[1] Aggarwal, C. C., and Abdelzaher, T. F. Social sensing. In *Managing and Mining Sensor Data*, C. C. Aggarwal, Ed. Springer, 2013, pp. 237–297.

[2] Amintoosi, H., Allahbakhsh, M., Kanhere, S., and Torshiz, M. N. Trust assessment in social participatory networks, 2013.

[3] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., and Srivastava, M. B. Participatory sensing. In *Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications* (2006), pp. 117–134.

[4] Choudhury, T., Philipose, M., Wyatt, D., and Lester, J. Towards activity databases: Using sensors and statistical models to summarize people's lives. *IEEE Data Eng. Bull. 29*, 1 (2006), 49–58.

[5] Cox, L. P. Truth in crowdsourcing. *IEEE Security and Privacy 9*, 5 (2011), 74–76.

[6] Deng, L., and Cox, L. P. Livecompare: Grocery bargain hunting through participatory sensing. In *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications* (New York, NY, USA, 2009), HotMobile '09, ACM, pp. 4:1–4:6.

[7] Dong, X. L., Berti-Equille, L., and Srivastava, D. Integrating conflicting data: The role of source dependence. *Proc. VLDB Endow. 2*, 1 (Aug. 2009), 550–561.

[8] Dong, X. L., Berti-Equille, L., and Srivastava, D. Truth discovery and copying detection in a dynamic world. *Proc. VLDB Endow. 2*, 1 (Aug. 2009), 562–573.

[9] Dua, A., Bulusu, N., Feng, W.-C., and Hu, W. Towards trustworthy participatory sensing. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security* (Berkeley, CA, USA, 2009), HotSec'09, USENIX Association, pp. 8–8.

[10] Eagle, N., and Pentland, A. Reality mining: Sensing complex social systems, 2005.

[11] Eagle, N., Pentland, A. S., and Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences 106*, 36 (Sept. 2009), 15274–15278.

[12] Eiben, A. E., and Smith, J. E. *Introduction to Evolutionary Computing.* SpringerVerlag, 2003.

[13] Eisenman, S. B., Miluzzo, E., Lane, N. D., Peterson, R. A., Ahn, G.-S., and Campbell, A. T. The bikenet mobile sensing system for cyclist experience mapping. In *Proceedings of the 5th international conference on Embedded networked sensor systems* (New York, NY, USA, 2007), SenSys '07, ACM, pp. 87–101.

[14] Estrin, D. L. Participatory sensing: applications and architecture. In *MobiSys* (2010), ACM, pp. 3–4.

[15] Gilbert, P., Cox, L. P., Jung, J., and Wetherall, D. Toward trustworthy mobile sensing. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems &#38; Applications* (New York, NY, USA, 2010), HotMobile '10, ACM, pp. 31–36.

[16] Gilbert, P., Jung, J., Lee, K., Qin, H., Sharkey, D., Sheth, A., and Cox, L. P. YouProve: authenticity and fidelity in mobile sensing. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems* (New York, NY, USA, Nov. 2011), SenSys '11, ACM, pp. 176–189.

[17] Goldman, J., Shilton, K., Burke, J., Estrin, D., Hansen, M., Ramanathan, N., Reddy, S., Samanta, V., Srivastava, M., and West, R. Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world. *Foresight & Governance Project, White Paper* (2009).

[18] Hicks, J., Ramanathan, N., Kim, D., Monibi, M., Selsky, J., Hansen, M., and Estrin, D. Andwellness: an open mobile system for activity and experience sampling. In *Wireless Health 2010* (New York, NY, USA, 2010), WH '10, ACM, pp. 34–43.

[19] Lane, N., Choudhury, T., and Campbell, A. Bewell: a smartphone application to monitor, model and promote wellbeing. In *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare* (2011).

[20] Le, H. K., Pasternack, J., Ahmadi, H., Gupta, M., Sun, Y., Abdelzaher, T. F., Han, J., Roth, D., Szymanski, B. K., and Adali, S. Apollo: Towards factfinding in participatory sensing. In *IPSN* (2011), X. D. Koutsoukos, K. Langendoen, G. J. Pottie, and V. Raghunathan, Eds., IEEE, pp. 129–130.

[21] Lenders, V., Koukoumidis, E., Zhang, P., and Martonosi, M. Location-based trust for mobile user-generated content: Applications, challenges and implementations. In *Proceedings of the 9th Workshop on Mobile Computing Systems and Applications* (New York, NY, USA, 2008), HotMobile '08, ACM, pp. 60–64.

[22] Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., and Choudhury, T. Stresssense: detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (New York, NY, USA, 2012), UbiComp '12, ACM, pp. 351–360.

[23] Madan, A., Cebrián, M., Moturu, S. T., Farrahi, K., and Pentland, A. Sensing the "health state" of a community. *IEEE Pervasive Computing 11*, 4 (2012), 36–45.

[24] Madan, A., Moturu, S. T., Lazer, D., and Pentland, A. S. Social sensing: obesity, unhealthy eating and exercise in face-to-face networks. In *Wireless Health 2010* (2010), WH '10, ACM, pp. 104–110.

[25] Olguín, D. O., Gloor, P. A., and Pentland, A. Wearable sensors for pervasive healthcare management. In *PervasiveHealth* (2009), IEEE, pp. 1–4.

[26] Oliveira, M. P. G., Medeiros, E. B., and Davis, C. A. Planning the acoustic urban environment: A gis-centered approach. In *ACM-GIS* (1999), C. B. Medeiros, Ed., ACM, pp. 128–133.

[27] Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D., and Srivastava, M. B. Biketastic: sensing and mapping for better biking. In *CHI* (2010), ACM, pp. 1817–1820.

[28] Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., Mavridis, N., Tellex, S., Salata, A., Guinness, J., Levit, M., and Gorniak, P. The Human Speechome Project Symbol Grounding and Beyond. vol. 4211 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2006, ch. 15, pp. 192–196.

[29] Ryder, J., Longstaff, B., Reddy, S., and Estrin, D. Ambulation: A tool for monitoring mobility patterns over time using mobile phones. In *CSE (4)* (2009), IEEE Computer Society, pp. 927–931.

[30] Srivastava, M., Abdelzaher, T., and Szymanski, B. Human-centric sensing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 370*, 1958 (Jan. 2012), 176–197.

[31] Tang, L. A., Yu, X., Kim, S., Han, J., Hung, C.-C., and Peng, W.-C. Tru-alarm: Trustworthiness analysis of sensor networks in cyber-physical systems. In *ICDM* (2010), G. I. Webb, B. L. 0001, C. Zhang, D. Gunopulos, and X. Wu, Eds., IEEE Computer Society, pp. 1079–1084.

[32] Tilak, S. Real-world deployments of participatory sensing applications: Current trends and future directions. *ISRN Sensor Networks 2013* (mar 2013).

[33] Wang, C., Burris, M. A., and Ping, X. Y. Chinese village women as visual anthropologists: A participatory approach to reaching policymakers. *Social Science and Medicine 42*, 10 (1996), 1391–1400.

[34] Wang, D., Abdelzaher, T., Ahmadi, H., Pasternack, J., Roth, D., Gupta, M., Han, J., Fatemieh, O., Le, H., and Aggarwal, C. On bayesian interpretation of fact-finding in information networks. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on* (2011), pp. 1–8.

[35] Wang, D., Abdelzaher, T., Kaplan, L., and Aggarwal, C. On quantifying the accuracy of maximum likelihood estimation of participant reliability in social sensing, 2011.

[36] Wang, D., Abdelzaher, T., Kaplan, L., and Aggarwal, C. Recursive fact-finding: a streaming approach to truth estimation in crowdsourcing applications. In *Proceedings of the 33rd International Conference on Distributed Computing Systems (ICDCS)* (Jul 2013).

[37] Wang, D., Kaplan, L., Le, H., and Abdelzaher, T. On truth discovery in social sensing: a maximum likelihood estimation approach. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks* (New York, NY, USA, 2012), IPSN '12, ACM, pp. 233–244.

[38] Yin, X., Han, J., and Yu, P. S. Truth discovery with multiple conflicting information providers on the web. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2007), KDD '07, ACM, pp. 1048–1052.

[39] Yin, X., and Tan, W. Semi-supervised truth discovery. In *Proceedings of the 20th International Conference on World Wide Web* (New York, NY, USA, 2011), WWW '11, ACM, pp. 217–226.