Proc. 13th International Conference on Information Processing in Sensor Networks,
ACM/IEEE IPSN, Berlin, Germany, April 15-17, 2014, pp. 35-46.

# Using Humans as Sensors: An Estimation-theoretic Perspective

Dong Wang[1], Md Tanvir Amin[1], Shen Li[1], Tarek Abdelzaher[1], Lance Kaplan[2],
Siyu Gu[1], Chenji Pan[1], Hengchang Liu[6], Charu C. Aggarwal[3], Raghu Ganti[3],
Xinlei Wang[4], Prasant Mohapatra[4], Boleslaw Szymanski[5], Hieu Le[7]
[1]Department of Computer Science, University of Illinois at Urbana Champaign, Urbana, IL 61801
[2]Networked Sensing and Fusion Branch, US Army Research Labs, Adelphi, MD 20783
[3]IBM Research, Yorktown Heights, NY 10598
[4]Department of Computer Science, University of California, Davis, CA 95616
[5]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180
[6]Department of Computer Science, University of Science and Technology of China, Hefei, Anhui 230027
[7]Caterva, Inc. Champaign, IL 61820

***Abstract*—The explosive growth in social network content suggests that the largest "sensor network" yet might be *human*. Extending the participatory sensing model, this paper explores the prospect of utilizing social networks as sensor networks, which gives rise to an interesting *reliable sensing* problem. In this problem, individuals are represented by sensors (data sources) who occasionally make observations about the physical world. These observations may be true or false, and hence are viewed as binary claims. The reliable sensing problem is to determine the correctness of reported observations. From a networked sensing standpoint, what makes this sensing problem formulation different is that, in the case of human participants, not only is the reliability of sources usually unknown but also the original data provenance may be uncertain. Individuals may report observations made by others as their own. The contribution of this paper lies in developing a model that considers the impact of such information sharing on the analytical foundations of reliable sensing, and embed it into a tool called *Apollo* that uses Twitter as a "sensor network" for observing events in the physical world. Evaluation, using Twitter-based case-studies, shows good correspondence between observations deemed correct by Apollo and ground truth.**

**Categories and Subject Descriptors**
H.4 [Information Systems Applications:] Miscellaneous

***Keywords*—humans as sensors, social sensing, data reliability, uncertain data provenance, maximum likelihood estimation, expectation maximization**

## I. Introduction

The advent of online social networks, such as Twitter, where humans volunteer free information at scale about the physical world, begs the question of whether or not they can be leveraged as a category of sensor networks. Indeed, for example, in the aftermath of the Boston Marathon bombing in the US, in April 2013, individuals as well as authorities turned to Twitter for real-time information on the evolving event. Twitter also carried some of the first descriptions of damage from the Japan Tsunami, in March 2011. The Japan government later issued a study encouraging the use of social networks to assist with disaster recovery. An interesting research question is whether estimation-theoretic approaches can be used to reconstruct the accurate state of the physical environment from social observations?

The reader must be cautioned that social networks carry a lot of extraneous information as well, such as subjective opinions (e.g., "It is an inspiring day!") and personal emotions (e.g., "I am depressed"). These are not the topic of this paper as they fall outside the scope of sensing applications that obsereve the external physical world.

We are interested in participatory sensing of external physical state. In the aftermath of important events, many microblog entries offer physical descriptions of the event (e.g., "Shooting erupts on Libery Square!"). Such reporting is an act of sensing of the physical environment that is external to the (human) sensor. In this case, the physical environment has a unique state, leading to a unique ground truth, according to which these descriptions are either true or false (e.g., either there was shooting on Liberty Square or not). It is this category of claims about the physical environment that the current paper is concerned with.

The paper formulates a *reliable sensing problem* focusing on three related research questions from a networked sensing perspective; namely, (i) how can one model networked human sources (who report observations about the physical world) as participatory sensors, (ii) given this model, how can one filter out "bad data", reported by such sources, and (iii) since the human sensor model is necessarily a simplified abstraction, how good is the filtering algorithm at distinguishing good data from bad when tested on real human observations in the field? We address these questions by suggesting a simple abstraction that models human participants as *sources of unknown reliability generating binary measurements of uncertain provenance*. We show that a rigorous estimation-theoretic problem can be formulated based on the above model to optimize filtering of correct observations in a maximum likelihood sense. We then empirically demonstrate that, despite its simplicity, our human sensing model is indeed useful at enabling the reconstruction of ground truth from noisy human observations in practice.

For practical validation on real-world examples, we integrated our results into a tool, called Apollo, that triages

real-time Twitter feeds. Apollo computes the probability of correctness of individual tweets based on the aforementioned estimation-theoretic optimization problem formulation, taking uncertain provenance into account. As an example use-case, we used Apollo to collect Twitter feeds about gas availability from New York City during and shortly after Hurricane Sandy (in October 2012), when gas availability was severely disrupted due to hurricane damage. Tens of thousands of tweets were collected by Apollo about (rumored) availability of gas at different locations. Apollo determined the likelihood of correctness of individual tweets, taking into account unknown source reliability and uncertain provenance as described later in the paper. The top ranked tweets (by likelihood of correctness) were then manually compared to ground truth, made available after the fact from sources such as credit card gas purchase transaction data. The evaluation shows great correspondence between Apollo estimation results and ground truth, suggesting that over 90% of "top tweets", believed by Apollo, are actually true, compared to less than 50% of actual true observations in the original data. We also show a significant improvement over the state of the art [45], [52]. We report similar results from other use-cases as well.

The rest of this paper is organized as follows: We first review related work in Section II to put the contribution in context. In Section III, we present a model of humans as sensors. Section IV introduces the Apollo tool, and outlines our problem and solution approach. The proposed maximum likelihood estimation algorithm is detailed in Section V. Evaluation results are presented in Section VI. The limitation and future work are discussed in Section VII. Finally, we conclude the paper in Section VIII.

## II. Related Work

Our paper builds on recent work on assessing correctness of observations from sources of *unknown reliability* [45], [47]–[49], [52] and deriving analytic performance bounds for the resulting maximum likelihood estimator [50], [51]. Specifically, we extend the previous work by addressing the challenge of *uncertain provenance* as well, which is a main distinguishing factor between networked physical sensors and networked humans. In prior sensing literature on sources of unknown reliability, uncertain provenance was either ignored altogether [47], [48], [52], or addressed via admission control that selects only independent sources [45]. We show, in our evaluation, that such limitations lead to an inferior assessment of observation correctness in the case of humans as sensors. This is because humans can "gossip" about their observations, creating non-independent errors, which throw off algorithms that assume error independence.

The paper should not be confused with work from sociology and statistics on opinion polling, opinion sampling, influence analysis and surveys. Opinion polling and sampling are usually carefully designed and engineered by the experts to create appropriate questionnaires and select representative participants [25], [41]. These are often controlled experiments, and the provenance of the information is also controllable [57]. Moreover, data cleaning is domain specific and semantic knowledge is required [15]. In contrast, in the reliable sensing problem studied in this paper, the data collection is open to all. We assume *no control* over both the participants (data sources) and the measurements in their reports. The *reliability of sources* and their *data provenance* is usually unknown to the applications. The approach proposed in this paper is designed to be *general* and not require domain specific knowledge to clean the data.

Influence analysis and surveys are often subjective [6]. They tend to survey personal facts, or individual emotions and sentiments [13]. This is as opposed to assessing physical state that is external to the human (sensor). For example, a survey question may ask "Was the customer service representative knowledgeable?" or it may ask "Do you support government's decision to increase tax?". Survey participants answer the questions with their own ideas independently, and the responses are often private [44]. Source dependency is not the main issue in these studies [2]. In contrast, in this paper, it is not our goal to determine what individuals feel, think, or support, or to extract who is influential, popular, or trending. Instead of assessing humans' own beliefs, opinions, popularity, or influence, we focus on applications concerned with the *observation and state estimation of an external environment*. That external state has a *unique ground truth* that is independent of human beliefs. Humans act merely as sensors of that state. There is therefore an objective and unambiguous notion of *sensing error*, leading to a clear optimization problem whose goal is to reconstruct ground truth with minimum error from reported human observations.

Remote sensing generally refers to detecting, monitoring and identifying objects on Earth using aerial sensor technology [26]. Remote sensing applications include environment monitoring, natural resource management, national security, and reports of nature disasters [1], [19]. In these applications, specific physical sensors are chosen and large scale data acquisition and processing infrastructure is built [46]. In contrast, this paper takes the first step to model humans as sensors. Compared to physical sensors, humans are able to observe a much broader spectrum of physical and social events at much lower costs (e.g., disaster tracking using online social media). However, humans are not as reliable as well tested infrastructure sensors and humans can propagate observations through the social network. This paper addressed these unique challenges brought by taking humans as sensors to report the status of the physical world.

The work extends the ideas of participatory [4], opportunistic [24] and human-centric [20] sensing, where humans play a key role in the sensing process. These ideas were investigated at length in projects such as MetroSense [5] and Urban Sensing [8]. Examples of some early applications include CarTel [21], BikeNet [12], and CabSense [37]. The suggestion that this people-centric genre of sensing should also cover *humans as the sensors themselves* (as opposed to being sensor carriers and operators) did not come until more recently [42].

A recent survey of human-centric sensing [42] covers many sensing challenges in human context such as accommodating energy constraints of mobile sensing devices [28], protecting the privacy of participants [3], and promoting social interactions in different environments [35]. It also suggests that humans are the most versatile sensors, listing new challenges that stem from the fact that observations may propagate among

such "sensors", leading to correlated noise and bias; a problem we refer to in this paper as *uncertain provenance*.

Sakaki et al. modeled twitter users as social sensors to report earthquake in Japan [36]. Zhao et al. reported their experience of using Twitter to monitor the US National Football League (NFL) games in real-time [56]. Our paper is inspired by these results. We propose a more general model for humans as sensors that includes the uncertain provenance aspect and accounts for the resuling source non-independence in the theoretical problem formulation.

When considering uncertain provenance, we should note that much work has addressed the challenge of ascertaining the authenticity of data and source devices. For example, the Trusted Platform Module (TPM), commonly used in commodity PCs, provides a certain level of assurance that the source device and software application are who they claim to be [27]. Similarly, YouProve [17] ensures authenticity of data (such as images taken by a phone), even after certain meaning-preserving modifications, such as cropping, have been performed. Such solutions, however, do not help with our uncertain provenance problem when humans are the sensors because authenticating Bob or his email client as the source of a reported observation about the physical world does not tell me whether Bob made the observation himself, or heard it from Sally. The real nature of our problem, therefore, is that information may propagate among sources along social network routes before it is reported to our "base station".

Techniques for deriving accurate conclusions from sources whose reliability is unknown are traced back to data mining and machine learning literature on *fact-finding*. One of the early papers on the topic was Hubs and Authorities [23] that allows one to iteratively compute both source trustworthiness and claim credibility, hence overcoming the unknown source reliability problem. Other instances of this iterative model include TruthFinder [54], and the Investment and Average·Log algorithms [32]. More general frameworks have been proposed to enhance the above basic model: Pasternack et al. [33] incorporate prior knowledge of the claims into fact-finding to leverage what the user already knows. Gupta et al. [18] accounted for a source's expertise across different topics. Galland et al. [14] considered the notion of hardness of facts by rewarding sources that correctly assert highly disputed claims. While such prior work was essentially heuristic in nature, an optimal solution to (a simplified version of) the problem was recently proposed [52], and compared to four of the best fact-finders in literature in the context of a social sensing model, demonstrating improved performance. In turn, we outperform this winning approach from [52] by taking uncertain provenance into account in a rigorous maximum-likelihood problem formulation. Our algorithm requires knowledge of source dependencies. The source dependency detection problem was discussed in past literature and several solutions were proposed [11], [34].

The uncertain provenance problem addressed in this paper is not new to social networks work, which addressed the related challenge of rumor detection. Nel et al. [29] propose a method to detect rumors by using the information publishing behavior of the sources and clustering sources with similar behavior. Shah and Zaman [39] propose "rumor centrality" as a maximum likelihood estimator to detect the source of rumors. Jin et al. [22] applied epidemiological models to study information cascades in twitter resulting from both news and rumors. Castillo et al. [7] develop a method that uses source, content, and propagation patterns to classify rumors from non-rumors. Seo et al. [38] use a number of specialized "monitor-nodes" and the social network to identify possible rumors and the rumor source. The work on rumor-detection is largely complementary to ours. Our contribution lies in incorporating the social (and information dissemination) network topology into a framework for evaluating the likelihood of correctness of claims.

Finally, security is an important problem that we defer to future work. In this work, we do not cover mitigation measures against potential security attacks such as collusion or sybil attacks, limiting ourselves instead to an empirical evaluation using actual Twitter traces as a representation of the real world. We believe security is important and should be addressed in future publications. For example, recently proposed algorithms for collusion and sybil attack detection in social networks [40], [55] may be incorporated as additional filters that identify and drop "bad sources" from consideration before employing techniques described in this paper on the remaining data.

## III. A Binary Model of Human Sensing

We model humans as sources of (i) *unknown reliability*, generating (ii) *binary observations* of (iii) *uncertain provenance*. Of the these three features, the first is perhaps the most intuitive. Unlike physical sensors whose characteristics, calibration, and failure modes are known, we do not, in general, know the reliability of human observers and hence cannot assume it in our problem formulation. In the following subsections, we explain the remaining two model features; namely, binary observations and uncertain provenance.

### A. A Binary Sensor Model

From a sensor modeling perspective, an obvious difference between physical sensors and human observations is one of functional specialization versus breadth. Humans, are much broader in what they can observe, albeit less accurate. Table I gives examples of actual observations made on Twitter.

| |
|---|
| Crash blocking lanes on I-5S @ McBean Pkwy in Santa Clarita |
| BREAKING NEWS: Shots fired in Watertown; source says Boston Marathon terror bomb suspect has been pinned down |
| The police chief of Afghanistan's southern Kandahar province has died in a suicide attack on his headquarters. |
| Yonkers mayor has lifted his gas rationing order. Fill it up! |

TABLE I. Examples of Twitter Observations

Such observations can be thought of as measurements of different *binary variables*. They are binary because the observation reported can either be true or false. In a system featuring a collaboration of sensors and humans, it is therefore meaningful to collect from humans these binary states, whereas collect from sensors exact continuous values of related variables of interest. This has been the practice in participatory sensing, where participants were asked to make *binary* observations, such as "there is garbage here", where as sensors, such as GPS, would provide the corresponding continuous variable (e.g., location).

With the above in mind, in this paper, we focus on a binary observation model, common to geotagging applications. Generalizing from participatory sensing, each human reports an arbitrary number of observations, we call *claims* that can be individually either true or false. Different individuals have different reliability, expressed as the probability of producing true claims. In this model, the physical world is just a collection of mention-worthy facts. For example, "Main Street is flooded", "The BP gas station on University Ave. is out of gas", or "Police are shooting people on Market Square". Human observers report some of the facts they observe (e.g., on Twitter). The problem of reliable sensing is to infer which of the reported human observations match ground truth in the physical world.[1]

### B. Uncertain Provenance

A feature that lends novelty to our sensor model, is the notion of *uncertain data provenance*. Namely, it is not unusual for a person to report observations they received from others as if they were his/her own. Such rumor spreading behavior has no analogy in correctly functioning physical sensors. We call this problem one of uncertain data provenance because when Bob tweets that "Main Street is flooded", even if we authenticate Bob as the actual source of the tweet, we do not know if Bob truely observed that first-hand or heard it from Sally. From a sensing perspective, this means that errors in "measurements" across "sensors" may be non-independent, as one erroneous observation may be propagated by other sources without being verified.

### C. A Word on Simplicity

To conclude our introduction of the model, it is worth noting that the exercise this paper undertakes is to evaluate the efficacy of the simplest viable abstraction of humans as sensors first. The reader will legitimately find several key ways our simplified model can be extended. One can think of this paper as offering a performance baseline against which such future potential enhancements can be benchmarked. Clearly, the performance of the baseline sheds light on the utility of such enhancements. To emphasize its simplicity, we call our baseline model the *binary model of human sensing* and show in our evaluation that the resulting ground truth reconstruction algorithm does very well.

## IV. A Solution Architecture

To enable reconstruction of ground truth information from data reported by human sources, we need to (i) collect data from the "sensor network", (ii) structure the data for analysis, (iii) understand how sources are related, and (iv) use this collective information to estimate the probability of correctness of individual observations. These steps are described in the following subsections, respectively. We focus on Twitter as the underlying "sensor network".

[1]One should mention that the reliable sensing problem, in the case of binary variables, is in fact harder than its counterpart in the case of continuous measurements. When sensors report 10-bit numbers on a scale from 0 to 1023, all bits are related as part of the same number. One can thus exploit properties of numbers such as ordering to eliminate outliers, average the results, or compute medians. If the 10 bits, however, are independent binary variables, they are not related and there is less that one can exploit to remove noise and bad data.

### A. Data Collection

We perform data collection using Apollo. In principle, Apollo can collect data from any participatory sensing front end, such as a smart phone application. In this paper, we report on collecting data from Twitter. Tweets are collected through a long-standing query via the exported Twitter API to match given query terms (keywords) and an indicated geographic region on a map. These can either be *and*ed or *or*ed. In essence, Apollo acts as the "base station" for a participatory sensing network, where the query defines the scope of information collected from paricipants.

### B. Computing the Source-claim Graph

Next, we need to determine the internal consistency in reported observations. For this reason, observations are clustered based on a *distance function*. This function, $\texttt{distance}\ (t_1, t_2)$, takes two reported observations, $t_1$ and $t_2$, as input and returns a measure of similarity between them, represented by a logical distance. The more dissimilar the observations, the larger the distance. In the case of data collection from Twitter, we regard individual tweets as individual observations, and borrow from natural language processing literature a simple cosine similarity function [43] that returns a measure of similarity based on the number of matching tokens in the two inputs. The distance function nicely separates natural language processing concerns from sensing concerns, and is not the contribution of this paper.

As distances are computed, the set of input observations is transformed to a graph where vertices are individual observations and links represent similarity among them. We then cluster the graph, causing similar observations to be clustered together. We call each such cluster a *claim*. Hence, the claim represents a piece of information that several sources reported. We can now construct a source-claim graph, $SC$, in which each source, $S_i$, is connected to all claims they made (i.e., clusters they contributed to), and each claim, $C_j$, is connected to all sources who espoused it (i.e., all sources of tweets in the corresponding cluster). We say that $S_iC_j = 1$ if source $S_i$ makes claim $C_j$. Each claim can either be true or false. The claim is true if it is consistent with ground truth in the physical world. Otherwise, it is false. The source-claim graph constitutes an input to our analysis algorithm.

### C. Adding the Social Dissemination Graph

Next, we need to account for uncertain provenance. Sources may have reported either their own observations or observations they heard from others. We assume the existence of a latent social information dissemination graph, $SD$, that estimates how information might propagate from one person to another. A recent Sigmetrics paper [30] describes an algorithm to infer the latent contagion network underlying epidemic cascades, given the time when each node got infected. For our experiments, we construct the epidemic cascade (EC) social graph using the iterative greedy strategy described in their paper, where each distinct observation is modeled as a cascade and the time of contagion of a source describes when the source mentioned this observation. We call the resulting graph, the EC network. Specific to Twitter, we also try three other ways to estimate potential information dissemination among

sources. The first is to construct this graph based on the follower-followee relationship. A directed link $(S_i, S_k)$ exists in the social graph from source $S_i$ to source $S_k$ if $S_k$ is a follower of $S_i$. We call this graph the FF network. The second option is to construct the social network from the retweeting behavior of twitter users. In this case, a directed link $(S_i, S_k)$ exists in the social graph if source $S_k$ retweets some tweets from source $S_i$. We call this graph the RT network. The third option combines the above two, forming a network where a directed link $(S_i, S_k)$ exists when either $S_k$ follows $S_i$ or $S_k$ retweets what $S_i$ said. We call the third type of social network the RT+FF network.

### D. Solving the Estimation Problem

With inputs computed, the next stage is to perform the analysis that estimates correctness of claims. For each claim, $C_j$, Apollo determines if it is true or false. Apollo uses a sliding window approach for analyzing received tweets. Let the total number of claims computed from tweets received in the last window be $N$. A trivial solution would be to count the number of sources, $S_i$, that made each claim. In other words, for each $C_j$, where $1 \leq j \leq N$, count all $S_i$, where $S_iC_j = 1$. The idea being that claims with more support are more believable. This solution is called *voting*, in an analogy with counting the number of votes. Unfortunately, it is suboptimal for two reasons. First, different sources have different degrees of reliability. Hence, their "votes" do not have the same weight. Second, sources may not be independent. When a source simply repeats what they heard from others, their "vote" does not add to the credibility of the claim.

Since the only information we have (other than the reported observations themselves) is the source claim graph, $SC$, and the social dissemination graph, $SD$, computed from the two steps above, the question becomes: Given graphs $SC$ and $SD$ what is the likelihood that claim $C_j$ is true, for each $j$? Formally, we compute:

$$\forall j, 1 \leq j \leq N : P(C_j = 1|SC, SD) \tag{1}$$

where $P(C_j = 1|SC, SD)$ is the conditional probability that $C_j$ is true given $SC$ and $SD$. With the aforementioned probability computed, Apollo forwards to the user those tweets that meet a specified (user configurable) probability of correctness. This feed is the solution to the reliable sensing problem.

## V. Expectation Maximization

It remains to show how to cast the problem of computing the probability of correctness of claims as a maximum likelihood estimation problem when sources have *unknown reliability* and data has *uncertain provenance*. Let $m$ be the total number of sources in our system from which we have data. Let us describe each source (i.e., "sensor"), $S_i$, $1 \leq i \leq m$, by two parameters; the odds of true positives, $a_i = P(S_iC_j = 1|C_j = 1)$ and the odds of false positives, $b_i = P(S_iC_j = 1|C_j = 0)$, neither of which are known in advance. Let us also denote by $d$ the unknown expected ratio of correct claims in the system, $d = P(C_j = 1)$. Let us now define the vector $\theta$ to be the vector of the above unknowns:

$$\theta = [a_1...a_m b_1...b_m d] \tag{2}$$

A maximum likelihood estimator finds the values of the unknowns that maximize the probability of observations, $SC$, given the social network $SD$. Hence, we would like to find $\theta$ that maximizes $P(SC|SD, \theta)$. The probability $P(SC|SD, \theta)$ depends on which claims are true and which are false. Let us therefore introduce the vector $Z$ where element $z_j = 1$ if $C_j$ is true and zero otherwise. Using the total probability theorem, we can now rewrite the expression we want to maximize, namely $P(SC|SD, \theta)$, as follows:

$$P(SC|SD, \theta) = \sum_z P(SC, z|SD, \theta) \tag{3}$$

We solve this problem using the expectation maximization (EM) algorithm [9], [10]. We note that authors in [9] used the EM algorithm in a crowdsourcing application to estimate the error rate of data sources. They assume the sources independently report their data and the data provenance is known to the application. However, such assumptions no longer hold in our applications where social networks are modeled as sensor networks and the information propagation between sources is common. In this paper, we explicitly model the source dependency and uncertain data provenance in our maximum likelihood estimator and present an enhanced EM algorithm in this section to address these challenges. The proposed EM scheme starts with some initial guess for $\theta$, say $\theta_0$ and iteratively updates it using the formula:

$$\theta_{n+1} = argmax_\theta\{E_{z|SC,\theta_n}\{ln\, P(SC, z|SD, \theta)\}\} \tag{4}$$

The above breaks down into three quantities that need to be derived:

- The log likelihood function, $ln\, P(SC, z|SD, \theta)$
- The expectation step, $Q_\theta = E_{z|SC,\theta_n}\{ln\, P(SC, z|SD, \theta)\}$
- The maximization step, $\theta_{n+1} = argmax_\theta\{Q_\theta\}$

Note that, the latter two steps are computed iteratively until the algorithm converges. The above functions are derived below.

### A. Deriving the Likelihood

The key contribution of this paper lies in incorporating the role of uncertain provenance into the maximum likelihood estimation algorithm. To compute the log likelihood, we first compute the function $P(SC, z|SD, \theta)$. Let us divide the source claim graph $SC$ into subsets, $SC_j$, one per claim $C_j$. The subset describes which sources espoused the claim and which did not. Since claims are independent, we can re-write:

$$P(SC, z|SD, \theta) = \prod_{j=1}^{N} P(SC_j, z_j|SD, \theta) \tag{5}$$

which can in turn be re-written as:

$$P(SC, z|\theta) = \prod_{j=1}^{N} P(SC_j|SD, \theta, z_j)P(z_j) \tag{6}$$

where $P(SC_j|SD, \theta, z_j)$ is the joint probability of all observations involving claim $C_j$. Unfortunately, in general, the sources that make these observations may not be independent since they may be connected in the social network leading to a possibility that one repeated the observation of another. Let $p_{ik} = P(S_iC_j|S_kC_j)$ be the probability that source $S_i$ makes claim $C_j$ given that his parent $S_k$ (in the social dissemination network) makes that claim. We call $p_{ik}$ a *repeat ratio* and can approximately compute it from graph $SC$, for pairs of nodes connected in graph $SD$, as follows:

$$p_{ik} = \frac{\text{number of times } S_i \text{ and } S_k \text{ make same claim}}{\text{number of claims } S_k \text{ makes}} \tag{7}$$

Hence, the joint probability that a parent $S_p$ and its children $S_i$ make the same claim is given by $P(S_pC_j) \prod_i P(S_iC_j|S_pC_j)$ which is $P(S_pC_j) \prod_i p_{ip}$. This probability accounts for the odds of one source repeating claims by another. For illustration, let us now consider the special case of social network topology $SD$, where the network is given by a forest of two-level trees [2]. Hence, when considering claim $C_j$, sources can be divided into a set $M_j$ of independent subgraphs, where a link exists in subgraph $g \in M_j$ between a parent and child only if they are connected in the social network and the *parent claimed* $C_j$. The link implies source dependency as far as the claim in question is concerned. The intuition is that if the parent does not make the claim, then the children act as if they are independent sources. If the parent makes the claim, then each child repeats it with a given repeat probability. The assumed repeat probability determines the degree to which the algorithm accounts for redundant claims from dependent sources. The higher it is, the less credence is given to the dependent source. Two scenarios are illustrated by the two simple examples in Figure 1, showing the situation where source $S_1$, who has children $S_2$, $S_3$, and $S_4$, makes claim $C_1$ and when it does not make it, respectively. Note the differences in the computed probabilities of its children making claim $C_1$. In general, let $S_g$ denote the parent of subgraph $g$ and $c_g$ denote the set of its children, if any. Equation 6 can then be rewritten as follows:

$$P(SC, z|SD, \theta) = \prod_{j=1}^{N} P(z_j) \times \{ \prod_{g \in M_j} P(S_gC_j|\theta, z_j) \prod_{i \in c_g} P(S_iC_j|S_gC_j) \} \tag{8}$$

where

$$P(z_j) = \begin{cases} d & z_j = 1 \\ (1-d) & z_j = 0 \end{cases}$$

$$P(S_gC_j|\theta, z_j) = \begin{cases} a_g & z_j = 1, S_gC_j = 1 \\ (1-a_g) & z_j = 1, S_gC_j = 0 \\ b_g & z_j = 0, S_gC_j = 1 \\ (1-b_g) & z_j = 0, S_gC_j = 0 \end{cases}$$

$$P(S_iC_j|S_gC_j) = \begin{cases} p_{ig} & S_gC_j = 1, S_iC_j = 1 \\ 1-p_{ig} & S_gC_j = 1, S_iC_j = 0 \end{cases} \tag{9}$$

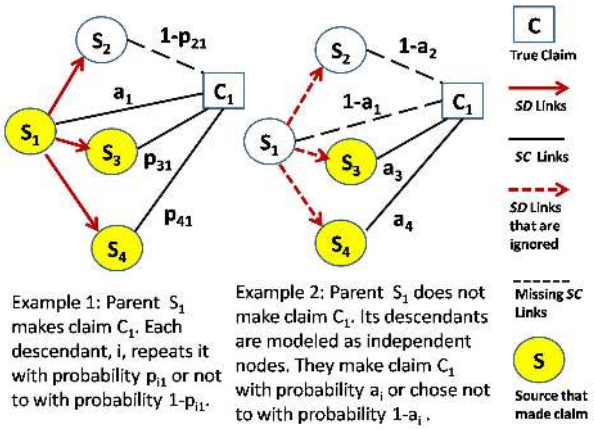


Fig. 1. Simple Illustrative Examples for Proof

### B. Deriving the E-step and M-step

Given the above formulation, substitute the likelihood function defined in Equation (8) into the definition of Q function of Expectation Maximization. The Expectation step (E-step) becomes:

$$\begin{aligned} Q\left(\theta|\theta^{(n)}\right) = \sum_{j=1}^{N} \Bigg\{ & Z(n,j) \times \Big[ \Big\{ \sum_{g \in M_j} \Big( \log P(S_gC_j|\theta, z_j) \\ & + \sum_{i \in c_g} \log P(S_iC_j|S_gC_j) \Big) \Big\} + \log d \Big] \\ & + (1 - Z(n,j)) \times \Big[ \Big\{ \sum_{g \in M_j} \Big( \log P(S_gC_j|\theta, z_j) \\ & + \sum_{i \in c_g} \log P(S_iC_j|S_gC_j) \Big) \Big\} + \log(1-d) \Big] \Bigg\} \end{aligned} \tag{10}$$

where $Z(n,j)$ is the conditional probability of claim $C_j$ to be true given the observed source claim subgraph $SC_j$ and current estimation on $\theta$. It is given by:

[2]The derivation can be easily extended to the network of multi-level tree and DAG

$$Z(n,j) = p(z_j = 1|SC_j, \theta^{(n)}) = \frac{p(z_j = 1; SC_j, \theta^{(n)})}{p(SC_j, \theta^{(n)})}$$
$$= \frac{p(SC_j, \theta^{(n)}|z_j = 1)p(z_j = 1)}{p(SC_j, \theta^{(n)}|z_j = 1)p(z_j = 1) + p(SC_j, \theta^{(n)}|z_j = 0)p(z_j = 0)}$$

where

$$p(SC_j, \theta^{(n)}|z_j = 1 \text{ or } 0)$$
$$= \prod_{g \in M_j} P(S_g C_j|\theta^{(n)}, z_j) \prod_{i \in c_g} P(S_i C_j|S_g C_j) \quad (11)$$

where $P(S_g C_j|\theta^{(n)}, z_j)$, $P(S_i C_j|S_g C_j)$ and $P(z_j)$ are defined in Equation 9.

We choose $\theta^*$ (i.e., $a_1^*, ..., a_m^*, b_1^*, ..., b_m^*, d^*$) that maximizes the $Q\left(\theta|\theta^{(n)}\right)$ function in each iteration to be the $\theta^{(n+1)}$ of the next iteration. To get $\theta^*$ that maximizes $Q\left(\theta|\theta^{(n)}\right)$, we set the derivatives $\frac{\partial Q}{\partial a_g} = 0$, $\frac{\partial Q}{\partial a_i} = 0$, $\frac{\partial Q}{\partial b_g} = 0$, $\frac{\partial Q}{\partial b_i} = 0$, $\frac{\partial Q}{\partial d} = 0$ which yields:

$$\begin{aligned}
a_g^{(n+1)} &= a_g^* = \frac{\sum_{j \in SJ_g} Z(n,j)}{\sum_{j=1}^{N} Z(n,j)} \\
a_i^{(n+1)} &= a_i^* = \frac{\sum_{j \in \overline{SJ}_g \cap SJ_i} Z(n,j)}{\sum_{j \in \overline{SJ}_g} Z(n,j)} \qquad \text{for } i \in c_g \\
b_g^{(n+1)} &= b_g^* = \frac{\sum_{j \in SJ_g} (1 - Z(n,j))}{\sum_{j=1}^{N} (1 - Z(n,j))} \\
b_i^{(n+1)} &= b_i^* = \frac{\sum_{j \in \overline{SJ}_g \cap SJ_i (1 - Z(n,j))} }{\sum_{j \in \overline{SJ}_g} (1 - Z(n,j))} \qquad \text{for } i \in c_g \\
d^{(n+1)} &= d^* = \frac{\sum_{j=1}^{N} Z(n,j)}{N}
\end{aligned} \quad (12)$$

where $N$ is the total number of claims in the source claim graph $SC$. $Z(n,j)$ is defined in Equation (11). $SJ_g$ denotes the set of claims the group parent $S_g$ makes in $SC$, and $\overline{SJ}_g$ denotes the set of claims $S_g$ does not make. Similar definitions apply to the children sources in the group (i.e., $SJ_i$ and $\overline{SJ}_i$). One should note that the computation of *repeat ratios* (i.e., $p_{ig}$ ) falls out of the estimation step in the EM algorithm and the result is not dependent on previous values of $\theta$ during the iteration.

Given the above, The E-step and M-step of EM optimization reduce to simply calculating Equation (11) and Equation (12) iteratively until they converge. The convergence analysis has been done for EM scheme and it is beyond the scope of this paper [53]. In practice, we can run the algorithm until the difference of estimation parameter between consecutive iterations becomes insignificant. Since the claim is binary, we can classify the correctness of claims based on the converged value of $Z(n,j)$. Specially, $C_j$ is true if $Z(n,j) \geq 0.5$ and false otherwise. This completes the mathematical development. We summarize the resulting algorithm in the subsection below.

**Algorithm 1** Expectation Maximization Algorithm

```
Initialize θ with random values between 0 and 1
Estimate the dependent ratio (i.e., p_ig) from source dissemination graph SD based on Equation (7)
while θ^(n) does not converge do
    for j = 1 : N do
        compute Z(n,j) based on Equation (11)
    end for
    θ^(n+1) = θ^(n)
    for i = 1 : M do
        compute a_1^(n+1), ..., a_m^(n+1), b_1^(n+1), ..., b_m^(n+1), d^(n+1) based on Equation (12)
        update a_1^(n), ..., a_m^(n), b_1^(n), ..., b_m^(n), d^(n) with a_1^(n+1), ..., a_m^(n+1), b_1^(n+1), ..., b_m^(n+1), d^(n+1) in θ^(n+1)
    end for
    n = n + 1
end while
Let Z_j^c = converged value of Z(n,j)
Let a_i^c = converged value of a_i^n; b_i^c = converged value of b_i^n; d^c = converged value of d^(n)
for j = 1 : N do
    if Z_j^c ≥ 0.5 then
        C_j^* is true
    else
        C_j^* is false
    end if
end for
Return the claim classification results.
```

### C. The Final Algorithm

In summary of the EM scheme derived above, the input is the source claim graph $SC$ from social sensing data and the source dissemination graph $SD$ estimated from social network, and the output is the maximum likelihood estimation of source reliability and claim correctness. In particular, given the source claim graph $SC$, our algorithm begins by initializing the parameter $\theta$ with random values between 0 and 1[3]. We also estimate the dependent ratio of each non-independent source (i.e., $p_{ig}$) from the source disseminate graph $SD$. The algorithm then iterates between the E-step and M-step until $\theta$ converges. Specifically, we compute the conditional probability of a claim to be true (i.e., $Z(n,j)$) from Equation (11) and the estimation parameter (i.e., $\theta^{(n+1)}$ ) from Equation (12). Finally, we can decide whether each claim $C_j$ is true or not based on the converged value of $Z(n,j)$ (i.e., $Z_j^c$). The pseudocode is shown in Algorithm 1.

## VI. Evaluation

In this section, we evaluate Apollo using three real world case studies based on Twitter. Evaluation results show the viability of predominantly correct ground truth reconstruction from social sensing data. In our evaluation, we compare the new maximum likelihood estimation algorithm, *Apollo-social*, to three baselines from current literature. The first baseline is *voting*, where data credibility is estimated by the

[3]In practice, if the a rough estimate of the average reliability of sources is known *a priori*, EM will converge faster

number of times the same tweet is collected from the human network. The larger the repetition, the more credibility is attributed to the content. Considering possible retweets on Twittter, we have two versions of the voting scheme: one that counts both retweets and original tweets as full votes (called regular Voting) and one that only counts the original tweets (called Voting-NoRT). The second baseline is the EM-based data cleaning algorithm proposed for participatory sensing applications in IPSN 2012 [52]. We henceforth call it *regular EM*. The algorithm differs from ours in that it assumes that all sources constitute independent observers, and was shown to outperform four current information ranking schemes. The last baseline is the social data cleaning scheme suggested in [45], which extends regular EM with *admission control*. The admission controller is designed to improve source independence by simply removing dependent sources using some heuristic approaches from social networks. We use the winning admission control scheme in [45], called Beta-1.

To compare these algorithms, we implemented them inside Apollo. Apollo was used to capture tweets from many events of interest such as hurricanes, riots, civil unrest, and other natural and man-made disasters. In particular, Apollo has a data collection component that allows users to specify a few key words and a geo-graphic location to collect tweets that contain the specified key words and originate from the specified location. The collected tweets were logged. For the purposes of evaluation, in this paper, we select three such traces of different sizes. The first was collected by Apollo during and shortly after hurricane Sandy, from around New York and New Jersey in October/November 2012. The second was collected during hurricane Irene, one of the most expensive hurricanes that hit the Northeastern United States in August 2011. The third one was collected from Cairo, Egypt during the violent events that led to the resignation of the former president in February 2011. In these traces, many claims were generated to describe the events that happened in the physical world, which have unique ground truth. For the granularity of the events, we first divided the data trace into different time intervals (e.g., a day) and then applied the tweet clustering function described in Section IV in each interval to cluster tweets that describe the same event together. These traces are summarized in Table II.

| Trace | Sandy | Irene | Egypt Unrest |
|---|---|---|---|
| Start Date | 11/2/2012 | 8/26/2011 | 2/2/2011 |
| Time duration | 14 days | 7 days | 18 days |
| # of tweets | 12,931 | 269,308 | 93,208 |
| # of users twitted | 7,583 | 207,562 | 13,836 |
| # of follower-followee links | 37,597 | 3,902,713 | 10,490,098 |
| # of users crawled | 704,941 | 2,510,316 | 5,285,160 |

TABLE II. STATISTICS OF THREE TRACES

The Apollo tool was fed each data trace above, while executing each one of the compared filtering algorithms. The output of filtering was manually graded in each case to determine match with ground truth. Due to man-power limitations, we manually graded only the 150 top ranked claims by each algorithm using the following rubric:

- *True claims:* Claims that describe a physical or social event that is generally observable by multiple independent observers and corroborated by sources external to the experiment (e.g., mainstream news media).
- *Unconfirmed claims:* Claims that do not meet the definition of true claims.

Note that, the unconfirmed claims contain the false claims and some possibly true claims that cannot be independently verified by external sources. Hence, our evaluation presents a *pessimistic* performance estimates, taking all unconfirmed claims as false. We also note that there could be possible cyclic dependency between news media and twitter as many news sites start to use twitter as an important data source. However, we found the claims that were officially reported by several mainstream media are those events that actually happened in the real world.

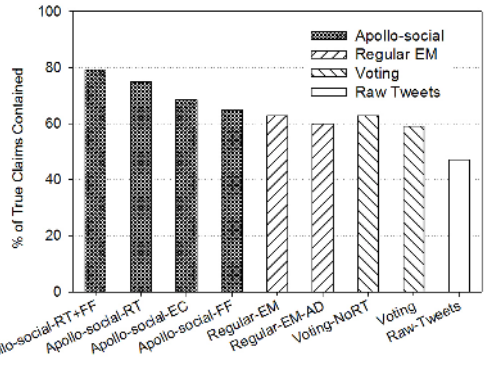


Fig. 2. Evaluation on Hurricane Sandy Trace

Figure 2 shows the result for the hurricane Sandy trace. We observe that Apollo-social generally outperformed the regular EM schemes in providing more true claims and suppressing the unconfirmed claims. This is achieved by incorporating the source dependency into the analytical framework of expectation maximization to better handle non-independent sources and their claims. The performance advantage of Apollo-social compared to regular EM is significant (nearly 20%) if we use the combined social network information (i.e, RT+FF social network) constructed from follower-followee and retweet relationship between users. We observed that the performance of the Apollo-social using Epidemic Cascades (EC) to estimate the social network is between Apollo-social using RT and FF social network. The reason is the RT social network is generated from the retweet relationship from current data interval and is very dynamic to reflect current source dependency and while FF social network is generated from the follower-followee relationship independently from the data traces and is relatively static. The dynamics of source dependency of EC social network is between RT and FF social network.

We also observe the regular EM schemes with admission control perform slightly worse than the one without admission control. The reason is: since the social network in the Sandy trace is relatively dense, the admission controller dropping some sources reduces the amount of useful information. In contrast, the Apollo-social scheme is shown to handle the social links in a more approriate way. We also note the voting scheme has a reasonable performance on this trace. The reason

| # | Media | Tweet found by Apollo-social | Tweet found by Regular EM |
|---|---|---|---|
| 1 | Rockland County Executive C. Scott Vanderhoef is announcing a Local Emergency Order restricting the amount of fuel that an individual can purchase at a gas station. | Rockland County Orders Restrictions on Gas Sales - Nyack-Piermont, NY Patch http://t.co/cDSrqpa2 | **MISSING** |
| 2 | New York City Mayor Michael Bloomberg has announced that the city will impose an indefinite program of gas rationing after fuel shortages led to long lines and frustration at the pump in the wake of superstorm Sandy. | Gas rationing plan set for New York City: The move follows a similar announcement last week in New Jersey to eas... http://t.co/nkmF7U9I | RT @nytimes: Breaking News: Mayor Bloomberg Imposes Odd-Even Gas Rationing Starting Friday, as Does Long Island http://t.co/eax7KMVi |
| 3 | New Jersey authorities filed civil suits Friday accusing seven gas stations and one hotel of price gouging in the wake of Hurricane Sandy. | RT @MarketJane: NJ plans price gouging suits against 8 businesses. They include gas stations and a lodging provider. | **MISSING** |
| 4 | The rationing system: restricting gas sales to cars with even-numbered license plates on even days, and odd-numbered on odd days will be discontinued at 6 a.m. Tuesday, Gov. Chris Christie announced on Monday. | # masdirin City Room: Gas Rationing in New Jersey to End Tuesday # news | RT @nytimes: City Room: Gas Rationing in New Jersey to End Tuesday http://t.co/pYIVOmPo |
| 5 | New Yorkers can expect gas rationing for at least five more days: Bloomberg. | Mayor Bloomberg: Gas rationing in NYC will continue for at least 5 more days. @eyewitnessnyc #SandyABC7 | Bloomberg: Gas Rationing To Stay In Place At Least Through The Weekend http://t.co/mmqqjYRx |

TABLE III. GROUND TRUTH EVENTS AND RELATED CLAIMS FOUND BY APOLLO-SOCIAL VS REGULAR EM IN SANDY

is: we used a set of concrete key words (e.g., gas, station, fuel, etc.) for data collection, which results in a relatively "clean" input with less irrelevant information. As we shall see, the performance of voting drops siginificantly when the input tweet trace has more noise (e.g., Egypt trace as we will discuss later).

The above results show the *precision* of the top claims. Another relevant metric is *recall*. Unfortunately, this metric is hard to define because we have no objective way to exhaustively enumerate all relevant physical events in order to determine what exact fraction of them was reported.

Note that, the exact recall may be of less interest, since we are usually interested in only the milestones and key moments of an event as opposed to every possible detail. Therefore, we carried out experiments to evaluate an approximate recall metric. Specifically, we independently collected 5 important events reported by media during Sandy to see if they are captured in our top claims. We then scanned through the top ranked claims for each of the algorithms compared to find these events. Results for selected baselines are shown in Table III. We observed that all five events were covered by the top claims from the Apollo-social scheme, while two of them were missing from the top claims returned from the regular EM scheme.

We repeated the above precision and recall experiments on the Irene tweet trace and Egypt tweet trace. The precision results for Irene are shown in Figure 3. In Figure 3, we consistently observe that Apollo-social achieves non-trivial performance gain in reducing the number of unconfirmed claims and providing more useful information by using the social network information. Similar results are shown for the Egypt trace in Figure 4. For recall, collecting 10 media events on each case, we observed that Apollo-social found all 10 of them in the case of Irene and 9 in Egypt, compared to 7 and 7 by regular EM.

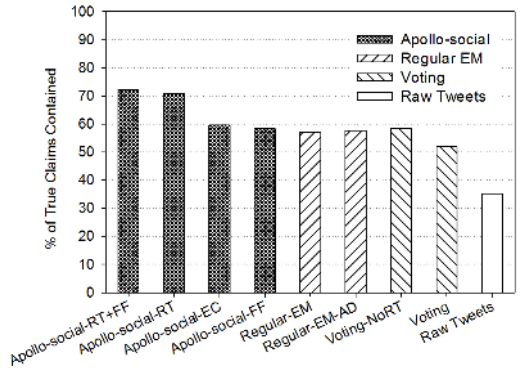


Fig. 3. Evaluation on Hurricane Irene Trace

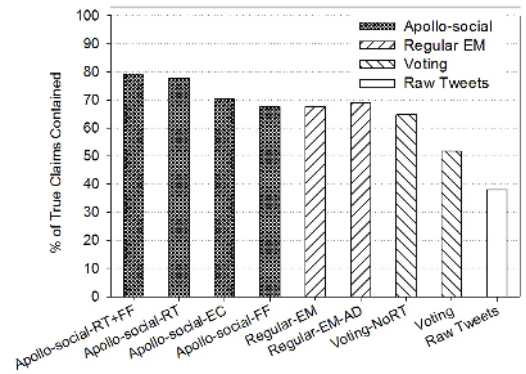


Fig. 4. Evaluation on Egypt Unrest Trace

Note that, in the experiments above, we do not report the number of claims that were verified to be *false* (i.e., false positives). This is because it is easier to verify that an event occurred than it is to verify that it did not. Prominent events that are true would typically be reported in the media. Obscure events would not be. It is therefore hard to verify if they really happened or not.

There was one exception to the above. Namely, in the Sandy example, we were able to collect ground truth on gas availability at a subset of New York and New Jersey gas stations at different points in time. This data was mainly obtained from two sources: (i) GasBuddy.com, which recruited spotters to go out and report the status of gas stations in New York and New Jersey during the gas shortage event in the aftermath of hurricane Sandy [16], and (ii) NYC_GAS_Finder, which generated updates on NYC gas stations serving gas based on credit card transaction data [31].

In order to match claims with the ground truth, we selected from the top claims those that (i) unambiguously mention fuel status, and (ii) explicitly describe a gas station location that matches one of the gas stations for which we have ground truth on *the day of the claim*. We considered the claim true if it matched ground truth status. Otherwise it was false. Thirty (30) of the top ranked Apollo claims could be verified this way. Of these, 29 were true matches, which is 97%. Hence, we have reason to believe that the number of unverified claims in other data sets actually contains a lot of true claims.

Finally, we also checked the average running time of the *entire* Apollo system (with different algorithms) took to process and analyze an hour's worth of data in each of three data traces we studied. The running time is measured from the point when the first tweet of the hour is processed to the point when all results of the hour are computed. The results are shown in Table IV. We observed, for all data traces, the processing time depends mainly on the scale of trace. Voting scheme ran the fastest, as might one expect, at the cost of more error in classifying claims correctly. The running speed of Apollo-social is quite comparable to the regular EM scheme. We also note that the running time of Apollo-social and other baselines is well below one hour, which verified the *real-time* feature of the Apollo system on the real-world data traces.

Note that, the values reported in Table 4 include the time taken for computing distance functions, clustering, and running the chosen estimation algorithm (i.e., Voting, Regular EM, or Apollo-social). This explains why voting comes so close to the other two algorithms in run time, despite the fact that they involve multiple iterations, while voting does not. A more detailed examination reveals that the bottleneck in Apollo lies in computing the distance functions between pairs of observations (as mentioned in Section 4.2). Hence, in total terms, the iterative expectation maximization algorithm described in this paper does not add much overhead.

| Algorithms | Sandy | Irene | Egypt Unrest |
|---|---|---|---|
| Apollo-social | 2.06 | 61.1 | 2.67 |
| Regular EM | 1.99 | 47.25 | 2.47 |
| Voting | 1.98 | 33.96 | 2.34 |

TABLE IV. RUNNING TIME (SECONDS) OF DIFFERENT ALGORITHMS ON AN HOUR'S TRACE

## VII. DISCUSSION AND LIMITATIONS

The paper presents encouraging results demonstrating that the accuracy of claims made by social sources can be estimated predominantly correctly even when such sources have unknown reliability and when data has uncertain provenance.

In performing this work, the authors encountered, on occasion, some interesting side-effects. Most prominently, while the main intention from considering source dependencies was to account for the reduced degree of corroboration when sources are related, an interesting side-effect was observed. Namely, the scheme tended to reduce the number of introspective (e.g., emotional and opinion) tweets, compared to tweets that presented descriptions of an external world. This was serendipitously quite appropriate of a "sensor network" that is not supposed to have emotions and opinions, but rather convey observations. Looking into the problem further, we noticed that emotions and slogans tended to be retweeted along social network topology pathways and hence, tended to be suppressed by our algorithm. In contrast, external facts (such as gas availability at a given location) were often observed independently by multiple sources, and hence not suppressed.

The observation offers many questions and opportunities for improvement, especially when it comes to modeling information propagation. The FF, RT and RT+FF are first approximations. They can be improved by building information propagation models that account for topic (e.g., sensational, emotional, and other news might propagate along different models), expertise, relationship reciprocity, mutual trust, and personal bias (e.g., the claim that "police is beating up innocent demonstrators", versus "demonstrators are attacking the police"). Note that distortions and biases of human sensors are quite persistent in terms of direction and amplitude, unlike white noise distortions. In a sense, humans exaggerate information in predictable ways. Understanding more about the community of sources can help better quantify the distortion and bias, leading to more accurate formulations of the maximum likelihood model.

In this paper, we focus on a binary claim model. It is shown to be a reasonable model to represent the physical events that are either true or false. However, there is also a large number of real world applications where observations from participants are *non-binary* (e.g., on-line review systems). Hence, we recently generalized our estimation framework to explicitly model the claims that have multiple mutually exclusive values [49]. Furthermore, we are now considering to generalize our model to better handle claims that have continuous values. We find that fuzzy logic could be a good modeling technique to apply in this case.

A separate problem is to deal with dynamics. When the network changes over time, how best to account for it in maximum likelihood estimation? A better formulation is needed that puts less weight on older data. Deception and malicious users also need to be addressed. Of particular interest are sources that gain advantage by acting reliably then change their behavior to deceive the system. Another obvious opportunity for extension lies in the binary sensor model. The current model is an

approximation. Having understood the performance of this model, is there a better sweet spot in trading accuracy for simplicity?

It is interesting to understand the impact of distance functions inside Apollo on the accuracy of estimation results. Distance function assess how close two tweets are. Current functions are based on syntax only (i.e., they compare words without interpreting them). How much benefit is attained by semantic analysis of different levels of sophistication? How does that benefit depend on the properties of the trace? These issues will be addressed in future work.

Finally, we should note that the human sensor model and the estimation framework developed in this paper is not limited to the applications that are based on Twitter. It can be also applied to a much broader set of crowdsourcing and mobile sensing applications, where the data are collected from both human sources and the devices on their behalf. Examples include traffic condition prediction using data from in-vehicle GPS devices and geo-tagging applications using participant's smart-phones. In these applications, humans or their devices represent the sources and measurements they report represent claims. The proposed estimation approach can be used to address similar data reliability and uncertain data provenance challenges in these applications.

## VIII. Conclusion

This paper presented an exercise in modeling social networks as sensor networks. A minimalist model was presented and its performance was evaluated. In this model, human sources represent sensors. The observations they make represent (data) claims. The sensing problem is to determine which claims are correct; which is to say, separate data from noise. This is similar to fusion problems in sensor networks, except for two challenges stemming from the nature of the human observer: first, the reliability of our human sensors is generally unknown *a priori*. Second, the provenance of reported observations is uncertain. The paper presented a maximum-likelihood solution to the sensing problem that is novel in addressing both of the above two challenges simultaneously. The solution was implemented in the Apollo tool and tested using data from Twitter. Test results show that the model offers sufficient accuracy in properly ascertaining the correctness of claims from human sources.

## Acknowledgements

Research reported in this paper was sponsored by the Army Research Laboratory, DTRA grant HDTRA1-10-1-0120, and NSF grants CNS 10-40380 and CNS 13-29886, and was accomplished under Cooperative Agreement Number W911NF09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

## References

[1] E. C. Barrett. *Introduction to environmental remote sensing*. Routledge, 2013.

[2] J. Blair, R. F. Czaja, and E. A. Blair. *Designing surveys: A guide to decisions and procedures*. SAGE Publications, Incorporated, 2013.

[3] I. Boutsis and V. Kalogeraki. Privacy preservation for participatory sensing data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, volume 18, page 22, 2013.

[4] J. Burke et al. Participatory sensing. In *Workshop on World-Sensor-Web (WSW): Mobile Device Centric Sensor Networks and Applications*, pages 117–134, 2006.

[5] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson. People-centric urban sensing. In *Proceedings of the 2nd annual international workshop on Wireless internet*, WICON '06, New York, NY, USA, 2006. ACM.

[6] A. E. Cano, S. Mazumdar, and F. Ciravegna. Social influence analysis in microblogging platforms–a topic-sensitive based approach. *Semantic Web*, 2011.

[7] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proc. WWW*, pages 675–684, NY, USA, 2011.

[8] D. Cuff, M. Hansen, and J. Kang. Urban sensing: out of the woods. *Commun. ACM*, 51(3):24–33, Mar. 2008.

[9] A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer Error-Rates using the EM algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):20–28, 1979.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[11] X. Dong, L. Berti-Equille, Y. Hu, and D. Srivastava. Global detection of complex copying relationships between sources. *PVLDB*, 3(1):1358–1369, 2010.

[12] S. B. Eisenman et al. The bikenet mobile sensing system for cyclist experience mapping. In *SenSys'07*, November 2007.

[13] K.-w. Fu and C.-h. Chan. Analyzing online sentiment to predict telephone poll results. *Cyberpsychology, Behavior, and Social Networking*, 2013.

[14] A. Galland, S. Abiteboul, A. Marian, and P. Senellart. Corroborating information from disagreeing views. In *WSDM*, pages 131–140, 2010.

[15] R. M. Gardner, D. L. Brown, and R. Boice. Using amazon's mechanical turk website to measure accuracy of body size estimation and body dissatisfaction. *Body Image*, 2012.

[16] GasBuddy. http://gasbuddy.com/.

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

[18] M. Gupta, Y. Sun, and J. Han. Trust analysis with clustering (poster paper). In *20th World Wide Web Conference (WWW'11)*, 2011.

[19] S. Hammi, V. Simonneaux, J. B. Cordier, D. Genin, M. Alifriqui, N. Montes, and L. Auclair. Can traditional forest management buffer forest depletion? dynamics of moroccan high atlas mountain forests using remote sensing and vegetation analysis. *Forest Ecology and Management*, 260(10):1861–1872, 2010.

[20] T. Higashino and A. Uchiyama. A study for human centric cyber physical system based sensing–toward safe and secure urban life–. In *Information Search, Integration and Personalization*, pages 61–70. Springer, 2013.

[21] B. Hull et al. CarTel: a distributed mobile sensor computing system. In *SenSys'06*, pages 125–138, 2006.

[22] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, page 8. ACM, 2013.

[23] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[24] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. Urban sensing systems: opportunistic or participatory? In *Proceedings of the 9th workshop on Mobile computing systems and*

*applications*, HotMobile '08, pages 11–16, New York, NY, USA, 2008. ACM.

[25] J. R. Lax and J. H. Phillips. How should we estimate public opinion in the states? *American Journal of Political Science*, 53(1):107–121, 2009.

[26] T. M. Lillesand, R. W. Kiefer, J. W. Chipman, et al. *Remote sensing and image interpretation.* Number Ed. 5. John Wiley & Sons Ltd, 2004.

[27] A. Mukhamedov, A. D. Gordon, and M. Ryan. Towards a verified reference implementation of a trusted platform module. In *Security Protocols XVII*, pages 69–81. Springer, 2013.

[28] S. Nath. Ace: Exploiting correlation for energy-efficient and continuous context sensing. In *Proceedings of the tenth international conference on Mobile systems, applications, and services (MobiSys'12)*, 2012.

[29] F. Nel, L. M.-J., P. Capet, and T. Dellavallade. Rumor detection and monitoring in open source intelligence: Understanding publishing behaviors as a prerequisite. In *Proc. Terrorism and New Media Conference*, 2010.

[30] P. Netrapalli and S. Sanghavi. Learning the graph of epidemic cascades. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '12, pages 211–222, New York, NY, USA, 2012. ACM.

[31] NYC Gas Finder. https://github.com/hirefrank/nycgasfinder.

[32] J. Pasternack and D. Roth. Knowing what to believe (when you already know something). In *International Conference on Computational Linguistics (COLING)*, 2010.

[33] J. Pasternack and D. Roth. Generalized fact-finding (poster paper). In *World Wide Web Conference (WWW'11)*, 2011.

[34] G.-J. Qi, C. C. Aggarwal, J. Han, and T. Huang. Mining collective intelligence in diverse groups. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1041–1052. International World Wide Web Conferences Steering Committee, 2013.

[35] K. K. Rachuri, C. Mascolo, M. Musolesi, and P. J. Rentfrow. Sociablesense: exploring the trade-offs of adaptive sampling and computation offloading for social sensing. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, MobiCom '11, pages 73–84, New York, NY, USA, 2011. ACM.

[36] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *19th international conference on World Wide Web (WWW'10)*, pages 851–860, 2010.

[37] Sense Networks. Cab Sense. http://www.cabsense.com.

[38] E. Seo, P. Mohapatra, and T. Abdelzaher. Identifying rumors and their sources in social networks. In *SPIE Defense, Security, and Sensing*, MD, USA, 2012.

[39] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57:5163–5181, 2011.

[40] L. Shi, S. Yu, W. Lou, and Y. T. Hou. Sybilshield: An agent-aided social network-based sybil defense among multiple communities. In *INFOCOM, 2013 Proceedings IEEE*, pages 1034–1042. IEEE, 2013.

[41] S. Splichal. Public opinion and opinion polling: Contradictions and controversies. *Opinion Polls and the Media: Reflecting and Shaping Public Opinion*, page 25, 2012.

[42] M. Srivastava, T. Abdelzaher, and B. K. Szymanski. Human-centric sensing. *Philosophical Transactions of the Royal Society*, 370(1958):176–197, January 2012.

[43] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. 2005.

[44] E. Toch, Y. Wang, and L. F. Cranor. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, 22(1-2):203–220, 2012.

[45] M. Uddin, M. Amin, H. Le, T. Abdelzaher, B. Szymanski, and T. Nguyen. On diversifying source selection in social sensing. In *Networked Sensing Systems (INSS), 2012 Ninth International Conference on*, pages 1 –8, june 2012.

[46] F. Viani, P. Rocca, G. Oliveri, and A. Massa. Pervasive remote sensing through wsns. In *Antennas and Propagation (EUCAP), 2012 6th European Conference on*, pages 49–50. IEEE, 2012.

[47] D. Wang, T. Abdelzaher, L. Kaplan, and C. C. Aggarwal. Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications. In *The 33rd International Conference on Distributed Computing Systems (ICDCS'13)*, July 2013.

[48] D. Wang, T. Abdelzaher, L. Kaplan, R. Ganti, S. Hu, and H. Liu. Exploitation of physical constraints for reliable social sensing. In *The IEEE 34th Real-Time Systems Symposium (RTSS'13)*, 2013.

[49] D. Wang, L. Kaplan, and T. Abdelzaher. Maximum likelihood analysis of conflicting observations in social sensing. *ACM Transactions on Sensor Networks (ToSN)*, Vol. 10, No. 2, Article 30, January, 2014.

[50] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal. On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing. In *The 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON 12)*, June 2012.

[51] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal. On credibility tradeoffs in assured social sensing. *IEEE Journal On Selected Areas in Communication (JSAC)*, 2013.

[52] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher. On truth discovery in social sensing: A maximum likelihood estimation approach. In *The 11th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN 12)*, April 2012.

[53] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.

[54] X. Yin, J. Han, and P. S. Yu. Truth discovery with multiple conflicting information providers on the web. *IEEE Trans. on Knowl. and Data Eng.*, 20:796–808, June 2008.

[55] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: Defending against sybil attacks via social networks. In *ACM SIGCOMM*, pages 267–278, 2006.

[56] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan. Human as real-time sensors of social and physical events: A case study of twitter and sports games. *CoRR*, abs/1106.4300, 2011.

[57] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma. Aspect-based opinion polling from customer reviews. *IEEE Trans. Affect. Comput.*, 2(1):37–49, Jan. 2011.