Recruitment Framework for Participatory Sensing Data Collections

Sasank Reddy, Deborah Estrin, and Mani Srivastava

Center for Embedded Networked Sensing
University of California at Los Angeles, USA
sasank@ee.ucla.edu, destrin@cs.ucla.edu, mbs@ee.ucla.edu

Abstract. Mobile phones have evolved from devices that are just used for voice and text communication to platforms that are able to capture and transmit a range of data types (image, audio, and location). The adoption of these increasingly capable devices by society has enabled a potentially pervasive sensing paradigm - participatory sensing. A coordinated participatory sensing system engages individuals carrying mobile phones to explore phenomena of interest using in situ data collection. For participatory sensing to succeed, several technical challenges need to be solved. In this paper, we discuss one particular issue: developing a recruitment framework to enable organizers to identify well-suited participants for data collections based on geographic and temporal availability as well as participation habits. This recruitment system is evaluated through a series of pilot data collections where volunteers explored sustainable processes on a university campus.

Keywords: Mobile Computing, Participatory Sensing, Urban Sensing

1 Introduction

The recent proliferation of mobile smart phones combined with the ease of deployment of web services for storage, processing and visualization, has ushered in a new pervasive data collection model - participatory sensing [1–3]. By enabling people to investigate previously difficult to observe processes with devices they use everyday, participatory sensing brings the ideals of traditional community based data collection and citizen science to an online and mobile environment; offering automation, scalability, and real-time processing and feedback [4, 5]. In participatory sensing, individuals explicitly select the sensing modalities to use and what data to contribute to larger data collection efforts. Example initiatives that are enabled by participatory sensing include our pilot data collections, where individuals collected photos of assets that documented recycling behavior, flora variety, and green resources to learn more about sustainability at a university.

However, advancing participatory sensing from a potential to a coordinated reality remains a major challenge. Finding a fit between diverse users and participatory sensing projects mirrors traditional selection for volunteer work based on interest and skill. But because participatory sensing is organized virtually, identifying particular participants (individuals who collect, analyze, and share their data) for campaigns (targeted data collection efforts) can be partially automated. Identification can rely not only on participants' reputations as data collectors based contribution habits, but can also be enhanced by incorporating participants' availability in the area of interest [6–8]. Specific attention is payed to the fact that humans have self-will, exhibit varied data collection performance, and have mobility traits that are opportunistic in nature [9].

This paper proposes a recruitment framework for participatory sensing data collections. Our work makes the following contributions: (a.) identifies availability and data collection performance as core attributes needed to match participants to campaigns, (b.) details models and algorithms that can be used to represent the recruitment factors, and (c.) evaluates the usefulness of the proposed recruitment mechanisms through pilot data collections. The rest of the paper is organized as follows: Section 2 illustrates example campaigns and motivates the recruitment problem. Section 3 provides an overview of the approach taken to address the recruitment challenge. Section 4 describes related work, and system details are given in Section 5. The paper ends with an evaluation of the recruitment framework and a discussion passage in Section 6 and 7 respectively.

2 Motivation and Application Examples

The application area for our data collections was an effort to learn more about sustainability practices at a university. A series of campaigns that documented various resource use issues were initiated. The data collections were enabled by a system consisting of a mobile phone client (Android G1 and Nokia n95) along with web services for data storage (Flickr and sensor database), analysis (Python application server), and visualization (Google Maps and Charts). Figure 1 a.) contains the mobile phone and web feedback page user interfaces. The campaigns involved taking geo-tagged photos, Figure 1 b.), and are described below:

- **GarbageWatch:** The campus needs to divert 75% of its waste stream from landfills, and effective recycling can help reach this goal. Participants documented the contents of outdoor waste bins through photo documentation. By analyzing the images, one can determine if recyclables (paper, plastic, glass, or aluminum) are being disposed of in waste bins, and then identify regions and time periods with low recycling rates.
- **What's Bloomin:** Water conservation is a high priority issue for the campus and efficient landscaping can help. This campaign involved taking geo-tagged photos of "blooming" flora. Having this inventory enables facilities to replace high water usage plants with ones that are drought tolerant. This flora catalog does not exist since the landscape is managed by many groups.
- **AssetLog:** For sustainable practices to thrive on a campus, the existence and locations of "green" resources needs to be documented. These resources include bicycle racks, recycle bins, and charge stations. But with expansion and re-construction activities, an up to date list is not available. Thus, this campaign tasked individuals to capture photos of these sustainability assets.

Fig. 1. System User Interface Design and Campaign Image Examples

Participatory sensing campaigns seek individuals willing to collect data about a particular phenomenon. A recruitment service takes campaign specifications as input and recommends participants for involvement in data collections. Campaign specifications may involve a number of factors including participants' device capabilities, demographic diversity, and social network affiliation. However, this work concentrates on a specific set of requirements for recruitment: participants' reputations as data collectors and availability in terms of geographic and temporal coverage. Also, our campaigns have an overall budget associated with them which may include resources needed to run the data collections along with compensation when incentives are provided for participant involvement. In our system, reputation is limited to considering participants' willingness (given the opportunity, is data collected) and diligence in collecting samples (timeliness, relevance and quality of data). Availability is learned from previously collected context-annotated mobility traces (i.e. streams of location, time, and transportation mode) in the campaign coverage area. Thus, the recruitment step would be used by campaign organizers to select participants who achieve the highest data collection utility while adhering to the set campaign budget. Overall, our recruitment framework is best suited for campaigns that have systematically defined data collection guidelines and are constrained in terms of coverage.

The sustainability campaigns are used to illustrate the features of the recruitment system. For these campaigns, well-suited participants are ones that regularly walk on campus during daytime hours and cover as much of the campus area as possible. Individuals that run, bike, or drive may be less likely to notice the resources of interest, and collecting clear photos is difficult at night. Furthermore, it is important that participants are willing to make observations when given the opportunity and that these samples are relevant and high quality.

3 System Overview

The process of recruiting volunteers for participatory sensing campaigns is analogous to recruiting volunteers or employees in non-virtual environments. Drawing on this similarity, we have created a recruitment framework, illustrated by Figure 2, that consists of three stages: the qualifier, assessment, and progress review.

- **The Qualifier:** Participants for campaigns must meet minimum requirements. For availability, prerequisites are based on destinations and routes within time, space, and transportation mode constraints. For participation reputation, requirements are measures of sampling likelihood, quality, and validity over several campaigns or by campaign-specific calibration exercises.
- **The Assessment:** Once participants that meet minimum requirements are found, the recruitment system then identifies which subset of individuals maximize coverage over a specific area and time period while adhering to the required transportation modes. Participants have costs and there exists a campaign budget which are both considered when selecting participants.
- **The Progress Review:** As a campaign runs, the recruitment system must check participants' coverage and data collection reputation to determine if they are consistent with their base profile. This check can occur periodically, and if the similarity of profiles is below a threshold, organizers should be alerted so that they can provide feedback or recruit additional participants.

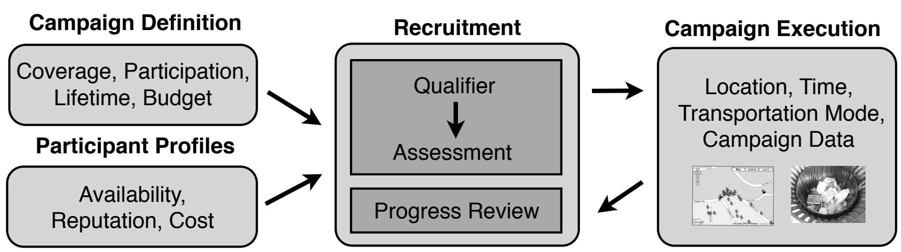


Fig. 2. Recruitment Framework Inputs, Outputs, and Steps

The design of the recruitment system takes into account the private nature of availability and participation data. Thus, the three-stage framework works to be parsimonious by limiting both the amount and granularity of information that is shared. Also, our system is designed to be run in coordination with a personal data vault where all participant information is stored and external queries on this data are strictly opt-in [10, 11]. For the qualifier and progress review, the query results sent to the data vault will simply be aggregate results of whether conditions or thresholds are met. In the case of the assessment, more detailed data in regards to mobility profiles needs to be shared with the recruitment system since coverage is based on collective participant mobility, but the data is limited to a particular spatial region, time span, and transportation mode.

4 Related Work

An overview of related work in terms of models, algorithms, and systems that share properties similar to the participatory sensing recruitment system is provided. First, models used to represent mobility and reputation that exist are reviewed. Then, details about systems that share a similar purpose of selecting resources for a task based on set conditions are documented.

4.1 Mobility Models

Location Summarization for Personal Analytics. There has been a significant amount of work in regards to coming up with clustering algorithms to summarize the most significant destinations of a user based on location traces [12–14]. The location traces can come from a GPS receiver, access point (GSM or WiFi) mappings, or hybrid setups that combine GSM, WiFi, and GPS. To derive the signification destinations, consecutive location points within a certain time period are aggregated into clusters. Also, certain systems use map matching and reverse geo-coding to add additional contextual information (semantic meaning) to the clusters [15]. This information has been used to create “gazetteers” (geographical dictionaries) for individuals. In terms of the recruitment framework, the qualifier step will have to use a similar clustering scheme as these systems since the granularity of summarization is on the destination and route level.

Location Prediction to Adapt Applications. Mobile quality of service (QoS) and location based services (LBS) have used location prediction in order to improve and enable applications. The mobile QoS work mainly concentrates on creating systems that provide predictive and adaptive bandwidth reservation for mobile phone users based on their short term mobility. These models take a very microscopic view on mobility, concentrating on determining which “cells” a user might travel based on transition patterns from previous cells, time spent in the current cell, and speed/trajectory information [16, 17]. Most LBS use the current location of a user for application adaptation, for instance in traffic, entertainment, and shopping settings. But researchers have proposed to make LBS more relevant to the user’s next destination. For instance, [18] models transitions of individuals using Markov processes, and [19] incorporates factors such as land-use information to help with destination prediction. Although the recruitment service does not require this type of short term prediction, the underlying algorithms to model historical location data is relevant.

Mobility Based Networking. The Mobile AdHoc Networking (MANET) community has used mobility models that simulate the movements of individuals to test out performance of networking protocols [20, 21]. Early work concentrated on using a random waypoint model where a node is specified with certain speed, direction, and duration of travel and then simulated to generate mobility patterns by randomly changing these factors after a period of time. Recently, these models have gotten more sophisticated with the inclusion of geographical constraints and historical information, but they are still mainly useful for generating statistically equivalent traces and not for modeling existing real world traces.

Delay tolerant networking (DTN) has also used mobility models in order to manage routing of messages so that systems would work in situations that do not have continuous network connectivity. These systems rely on creating location matrices that model the presence of an individual at different locations and then compare the profiles of users to figure out where to disseminate a message so that it will eventually end up in the target location [22, 23]. Similar type of modeling has been used by the Reality Mining project to learn about location habits of groups [24]. Overall, this work is relevant to the project review step in the recruitment process since mobility profile similarity checks are needed.

4.2 Reputation Models

Summation and Average. The simplest reputation models are ones that are summation and average based. In this setup, ratings are aggregated, by summing or averaging, to create an overall single reputation score [25]. An example of a summation system is eBay where ratings, which can be either -1 (negative), 0 (neutral), and 1 (positive), are added together [26]. Amazon instead uses averaging and relies on a "star" rating system that ranges from 1-5 where 1 is poor and 5 is excellent [27]. The advantage of these models is that they are easy to understand since a single number represents reputation, but the disadvantage is that they provide a primitive view on an individual's actions and can cover up negative ratings if many positive ratings exist in proportion [28].

Discrete Trust Models. An alternative scheme to having reputations being a numerical value is to use discrete labels. For example, the Slashdot web site aggregates ratings on actions, such as story submissions, postings, moderation activities, into "karma" tiers for participants that include terrible, bad, neutral, positive, good, and excellent [29]. Although this model is helpful for individuals to quickly determine a meaning for a reputation measure, it is not mathematically tractable and has no method to determine reputation confidence [28].

Bayesian Systems. Reputation models based on Bayesian frameworks have been popular for peer-to-peer networks and sensor systems [25, 30]. Particularly, these systems rely on ratings, either positive or negative, and use probability distributions, such as the Beta distribution, to come up with reputation scores [31]. By taking the expectation of the distribution, reputation can be determined. The confidence in this reputation score is captured by analyzing the probability that the expectation lies within an acceptable level of error. Additional features are easily enabled, such as aging out old ratings by using a weight factor when updating reputation and dealing with continuous ratings by employing an extension involving the Dirichlet process [30, 31]. Overall, this Bayesian framework, specifically with the Beta distribution, seems to be appropriate to model participant data collection habits.

4.3 Selection Services

Crowd-Sourcing Sites. Many crowdsourcing services on the web have requirements that need to be met before individuals can take part in a task [32]. Sites like Amazon Mechanical Turk and GURU.com, which are systems that provide a marketplace to get commissioned work done, keep detailed statistics tracking the performance of workers. In Amazon Mechanical Turk, work done by a participant is evaluated in terms of whether it was accepted or rejected by requesters [33]. In GURU.com, the technical skill, creativity, timeliness, and communication capabilities of a worker are kept through a star-based rating system based on feedback from work requesters [34]. Our work builds on this idea of monitoring user behavior and provides metrics to evaluate participation and performance of individuals involved in data collection.

Sensing Systems. Sensor network research has taken place in regards to selecting and placing static devices to maximize coverage [35]. Similarly, work exists to coordinate robotic motion for sensing purposes [36]. Unfortunately, the algorithms for these systems do not apply directly to the recruitment problem since mobility of individuals is not always controllable and there exists variability in when and how sampling occurs. Previous work related to mobile phone opportunistic sensing either concentrate on creating protocols to recognize when sensing should be activated based on pre-defined zones [6, 7, 37] or choosing how much sensing should occur depending on privacy restrictions [38]. Our work differs in that the data collection recruitment problem is directly addressed with participant availability, reputation, and coverage/participation inconsistency considered. Also, our system does not rely on knowing prior distribution information or having detailed statistical models of the phenomenon of interest.

5 System Details

The steps involved in both availability as well as participation and performance based recruitment are detailed below. Specifically, we focus on the inputs and outputs of each of the different steps in the recruitment framework. Also, we detail the models and algorithms involved in the framework.

5.1 Coverage Based Recruitment

Mobility Information. Coverage based recruitment relies on transforming raw participant mobility data into building blocks that can be used for processing. The system assumes that participants have previously collected location traces in the form of latitude, longitude, and time points for a period of time that represent their "typical" behavior (e.g. for a profile week). The location traces could be augmented with sensor-based information which can help in adding context such as transportation mode (still, walking, running, biking, or driving) [39–41]. Having this type of data collected by participants is not far fetched; services already exist that rely on location check-ins and traces [42, 43].

Qualifier. The transportation mode annotated location traces are transformed into significant destinations and routes for the qualifier. The system pre-processes the data by normalizing it to a set sample rate (for instance, every 30 seconds) and fills in missing values when the GPS signal is lost. For large spatial gaps, the points are filled by generating likely traces using the Google Maps API. Then, location points within a certain time period (at least 15 minutes) and distance bound (50 meters) are grouped into "stays" [44]. Density based clustering is used to group stays within a certain distance (250 meters) into "destinations" [14]. Routes are points between destinations and are aggregated using hierarchical clustering where the average minimum point segment distance is the comparison metric [45]. Qualifier queries use these building blocks to create filters, such as participants that have at least 5 destinations in a certain area in a week or individuals that have 7 unique walking routes during day time weekday hours.

Assessment. Next, in the assessment step a subset of qualified individuals that maximize coverage are identified. Formally, the assessment is an instance of the budgeted maximum coverage problem [46]. A participant pool, $P = \{p_1, p_2, ..., p_n\}$, exists with non-negative costs, $\{c_i\}_{i=1}^{n}$. Spatial and temporal blocks with an associated transportation mode, $E = \{e_1, e_2, ..., e_n\}$, are present. The blocks have utilities, $\{u_i\}_{i=1}^{n}$, defined for the campaign as well. The goal is to find a subset of participants $P^* \subseteq P$, such that the utility of elements covered, $U(P^*)$, is maximized while the cost of the subset, $C(P^*)$, is under a set campaign budget, B [38, 46]. Hence, the optimization can be stated as:

$$argmax\ U(P^*)\ subject\ to\ C(P^*) \leq B \tag{1}$$

This optimization is NP-hard since selecting a participant for the subset changes the utility for the rest not included. Thus, to find the best solution, all subset combinations must be searched. Since the utility function is sub-modular (adding a participant helps more if fewer are already selected) and non-decreasing (utility of subset is less than the set it is derived from), the greedy algorithm can find an adequate solution when costs are identical (at least 63% from the optimum) [47]. If costs are not identical, the benefit-cost greedy algorithm can be used where the ratio of utility to cost is used as the metric to pick participants [38]. Alternatively, this algorithm can help find the least costly subset to achieve a coverage goal.

Progress Review. While a campaign runs, check-ups are needed to ensure that participant mobility is consistent with the profile used for recruitment. Thus, in the progress review the similarity of mobility profiles is checked. To model mobility for the progress review, a time span (one week) is represented using an association matrix, A, consisting of $m\ x\ n$ entries [22, 48]. The m rows indicate spatial blocks (e.g. 10000 meter^2 grids) while the n columns model distinct time periods (days). An entry in the matrix is the proportion of time spent in a location performing set transportation modes within the time period selected. A day is chosen as the representative time period while a week is the time span based on previous work on human location patterns [24, 49].

Since it is only necessary to compare the dominant mobility patterns, a summarization technique for the association matrix is needed. Thus, Singular Value Decomposition (SVD) is applied to the association matrix: $A = U \cdot \Sigma \cdot V^t$. In this decomposition, U, the left eigenvectors, are referred to as eigenbehaviors and represent patterns that are common across different time periods (days), and the singular values Σ represent the variance represented by each pattern. Consecutive time spans (weeks) are compared by taking the cosine similarity of the behavior vectors weighted by the singular value importance [22]. Hence, if there exists two eigenbehaviors, U_{t1} and U_{t2}, representing different time spans, $t1$ and $t2$, with singular value importance, W_{t1} and W_{t2}, the similarity metric is:

$$Similarity(U_{t1}, U_{t2}) = \sum_{i=1}^{rank(U_{t1})} \sum_{j=1}^{rank(U_{t2})} w_{t1_i} w_{t2_j} |U_{t1_i} \cdot U_{t2_j}| \quad (2)$$

Similarity is indexed from 0 (least similar) to 1 (most similar) by normalizing on the base eigenbehavior similarity.

5.2 Participation and Performance Based Recruitment

Inspired by reputation metrics in other domains (Section 4.2), we divide data collector reputation into two classes: cross-campaign and campaign-specific. Cross-campaign indicators, such as the number of campaigns volunteered, participated in, and abandoned, provide a granular view of a participant's experience across many campaigns. Campaign-specific metrics measure the quality and quantity of samples that can be expected for a specific data collection. In our work, we concentrate on campaign-specific measures, specifically on participation likelihood. Other examples include timeliness, relevancy, and quality of samples.

Timeliness represents the latency between when a phenomenon is sampled (or occurs) and when it is available for analysis. It is influenced by user and upload delay. Relevancy indicates how well the sample describes the phenomenon of interest. It ranges from describing the item that is desired to not being related at all. Quality represents the ability of a processing module to determine a particular feature for further classification. Participation likelihood describes whether an individual took a sample when given the opportunity. These measures can be automatically quantified or might require human intervention. The campaign organizer defines a utility function that combines the importance of each metric to determine the overall reputation for a participant on a per campaign basis.

Modeling. The Beta distribution is adopted for campaign-specific reputation since it can be stored and updated efficiently, estimate stochastic (due to the randomness of the system) and epistemic uncertainty (due to lack of knowledge about the randomness of the system), and have features such as aging added on top easily. The distribution is indexed by alpha (α) and beta (β), which define the number of successful and unsuccessful events and is expressed as follows:

$$f(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad (3)$$

A participant's reputation can be found by calculating the expectation of the Beta distribution (stochastic uncertainty), $E(\alpha, \beta) = \alpha \;/\; (\alpha + \beta)$. Confidence in this reputation score (epistemic uncertainty) is the posterior probability given the actual expectation value lies within an acceptable level of error found by calculating the area under the Beta curve [50]. Alpha and beta are set to 1 initially, which results in a uniform distribution where all values are considered equally likely. Distributions that represent more evidence for a hypothesis are peaked at the expectation compared to ones with less evidence. Also, if continuous ratings are needed an extension of Beta involving the Dirichlet process can be used [30].

Qualifier. Most campaigns will not have a prior participant reputation data as related to the specific data collection that is of concern. Thus, it is necessary to go through a "calibration" exercise so that evidence is gathered for the qualifier step. This exercise commonly involves having an expert gather ground truth on set paths and directing participants to traverse them as well. In cases where an expert cannot be involved, participants simply get compared against each other on these paths. The contributions are evaluated in the Beta framework and the qualifier step removes individuals that do not have a certain reputation level.

Progress Review. As a campaign runs, the participation and performance of individuals could change. For example, individuals might be initially very diligent about data collection but then change their behavior due to loss of interest or schedule tensions. Thus, it is important to be able to check reputation based on the most current information. The Beta distribution provides the ability to consider discounting old information by using an aging factor, w. This aging is done by discounting existing reputation values at set intervals when updates occur [31]. Essentially, alpha and beta are transformed as follows:

$$\alpha_{new} = w_{age} * \alpha_{old} + \alpha_{obtained}; \quad \beta_{new} = w_{age} * \beta_{old} + \beta_{obtained} \tag{4}$$

6 Evaluation

This section analyzes the models and algorithms involved in coverage and reputation based recruitment. The sustainability campaigns provide the data for the evaluation. Importance is placed on highlighting the features of the framework.

6.1 Campaign Deployment Information

The sustainability campaigns were initiated by engaging individuals from campus student groups. Individuals were given a phone, trained on what to identify, and how to use the data collection software. Participants ran a campaign for at least one week although many continued for additional days (results shown in Table 1). Before the campaigns started, all individuals performed calibration exercises where they would go on pre-defined routes to collect data. These routes were also traversed by "experts" who gathered ground truth. During the campaign, participants did not receive instructions on where and when to sample.

Campaign Type	Total Images	Total Users	Average Per User	Maximum Per User	Minimum Per User
GarbageWatch	1752	31	56	231	7
What's Bloomin	4041	22	183	398	4
AssetLog	1488	16	93	266	11

Table 1. Campaign Participation Data

6.2 Coverage Based Recruitment

The usefulness of coverage based recruitment is analyzed with the GarbageWatch campaign. Specifically, we focus on the assessment and progress review stages. Participants have already passed the minimum qualification of having routes and destinations on the campus since they all belonged to the university community.

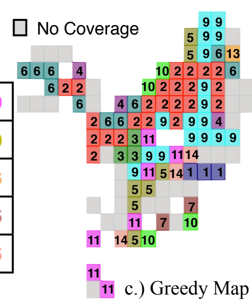
Assessment: Evaluating the Best Coverage of the Campus. For GarbageWatch, the spatial zones of interest are campus waste bin locations, temporal span is daytime weekday hours, and the transportation mode is walking. The size of the spatial blocks was set to 10000 m^2, which was empirically chosen based on GPS error and waste bin density, and the temporal block granularity was set to 1 hour so recycling behavior over time can be monitored. Three participant selection methods were compared: random, naive, and greedy. Random selects individuals for campaigns arbitrarily. Naive represents a heuristic where selecting participants is based on which individual covers the most blocks overall without considering what existing selected participants have covered. Greedy chooses participants that maximize utility while taking into consideration the coverage by existing selected participants. Thus, for greedy the participant utilities need to be re-calculated after an individual is selected. For evaluation purposes, the block utilities are all the same, the participant costs are set to 1, and the budget is limited to 15. In essence, 15 individuals are chosen from the pool of 31.

Algorithm	Space Blocks Covered	Space Time Blocks Covered	Space Blocks Not Covered
Random	80	397	34
Naive	81	578	33
Greedy	83	591	31

a.) Coverage Results for Different Algorithms

User	ST Blocks				
1	64	6	43	11	30
2	58	7	37	12	30
3	53	8	37	13	26
4	48	9	36	14	25
5	44	10	35	15	25

b.) Coverage Using Greedy



c.) Greedy Map

Fig. 3. Algorithm Comparison for GarbageWatch Campaign Coverage

The coverage results for the algorithms are shown in Figure 3 a.). The number of spatial temporal blocks is 6840 made up off 114 spatial blocks (based on 10000 m^2 granularity) and 60 time blocks (12 daytime hours per day for 5-weekday span). Furthermore, Figure 3 b.) shows specific coverage information for the greedy case, and Figure 3 c.) illustrates the greedy results on a map with the participant with the most coverage for a spatial block taking ownership. Random selection performs much worse then either the naive or greedy algorithms, specifically picking participants that have less spatial and temporal coverage and more spatial blocks not covered by anyone. The greedy algorithm performs better then just the naive heuristic. If more coverage overlap existed between participants, the performance of the greedy algorithm would be even higher. In general, considering availability when selecting participants is important. Otherwise, large coverage gaps could exist, and the opportunities available for sensing could be low. Also, the more complex instance of this problem, with variable costs for participants and different utilities for spatial and temporal blocks, can be handled by using a variant of the greedy algorithm where the benefit to cost ratio is used to evaluate participants during the selection process [38].

Progress Review: Comparing Coverage Profiles Over Time. As a campaign runs, participants availability might deviate from their established profiles. Thus, campaign organizers should be able to run checks on mobility profile consistency so that actions, such as recruiting additional individuals or providing feedback to existing participants, can take place if there is coverage loss. This progress review consistency check is especially important for long running campaigns since schedules might shift. The usefulness of the progress review is shown by analyzing two participants involved in the sustainability campaigns. One participant had a very stable schedule while the other had a significant shift occur. The mobility profile check is run by calculating similarity between eigenbehaviors of two weeks using SVD. Participants' mobility is modeled using an association matrix which is 114 (number of spatial blocks when considering a spatial granularity of 10000 m^2) by 5 (number of weekdays in a week) in size that takes into account daytime walking instances on campus during a week.

The mobility map and similarity score of Participant #9 is shown as Figure 4 a.), and based on interviewing the individual, we find that the participant mainly travels between two main hubs on campus and does not typically deviate. Thus, the similarity score of 0.85 based on comparing eigenbehaviors between the two weeks makes sense. In some cases, an individual might have a shift in schedule or a change in the way they travel. This was the case with Participant #2 who changed their transportation mode between their residence and campus from walking to driving between weeks. As shown in Figure 4 b.), this individual's similarity score is only 0.34. Overall, the SVD based similarity measure is effective to learn about major availability changes. Also this method has the advantage of summarizing mobility patterns in a compact manner - aggregating weeks of similar mobility data into a few dominant eigenbehaviors [8, 22, 48].

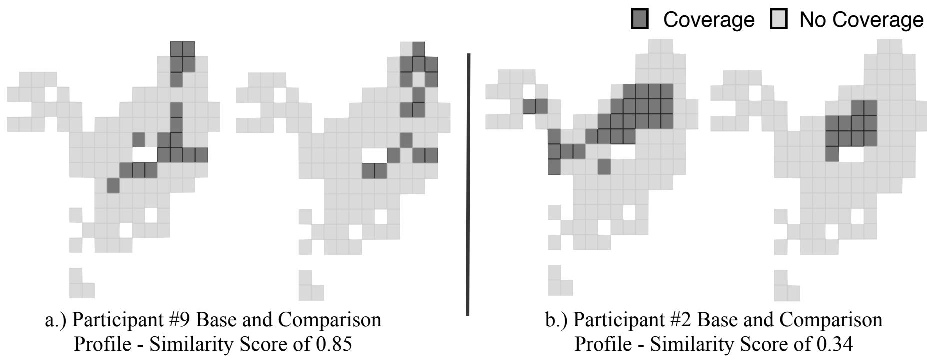


Fig. 4. Progress Review Consistency Check for Campaign Coverage

6.3 Participation and Performance Based Recruitment

Another factor to consider during recruitment is participants' reputations as data collectors. Although factors such as sample timeliness, relevancy, and quality can play a role in reputation, in our sustainability campaigns we found these elements to be less applicable since automatic image uploading was occurring, the items that needed to be sensed were distinctive, and participants took very few unusable images. Thus, we concentrate on whether a participant is likely to contribute a sample if they had an opportunity. This metric is used to exercise the features of the Beta distribution in the qualifier and progress review stages.

Qualifier: Running Calibration Exercise for Initial Reputation. Initially, for the three campaigns, no prior information existed in terms of sampling reputation. Thus, calibration exercises were implemented to get an initial sense of a participant's likelihood to capture a sample if they had a chance. This was done by having three specific routes that participants had to traverse for each campaign. Ground truth information was obtained along these paths by an "expert". For the case of GarbageWatch, opportunities to sample were places where waste bins existed. Similarly, for What's Bloomin the opportunities were related to places where flowers existed, and for AssetLog, each route was associated with a color and items of that color were samples of interest. The calibration routes were chosen to be paths that individuals on campus are familiar with. The calibration is run by participants once at the beginning of a campaign.

When designing a calibration exercise, an important factor to consider is whether there are enough sampling opportunities to be able to be confident of the reputation that is derived. For instance, in the case of the AssetLog campaign, if one route is only considered instead of all three to calculate initial reputation, then the campaign organizer might not have confidence in the reputation prediction provided. For example, Figure 5 shows Beta distributions for a participant where one route is compared to all three routes. As Figure 5 a.)

shows, even though the reputation of the participant is high with a score of 0.77 (likely to sample the phenomenon when given the chance), our confidence in his ability is low since the number of check points for sampling is small when considering only one route. When all three routes are used, Figure 5 b.), the confidence we have in the overall reputation of 0.81 is much higher. In fact, the confidence is at a level of 0.97 with all three routes considered as compared to just 0.61 when one is used. The confidence score was calculated by taking the area under the Beta curve with an acceptable error of 0.1 around the mean reputation.

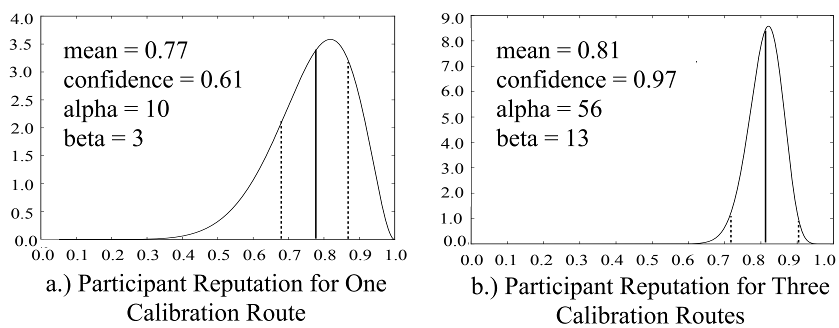


Fig. 5. Calibration Reputation for Participant in AssetLog Campaign

A question that comes up is whether these calibration exercises are useful as a predictor of sampling behavior during the actual campaign. To test this we compare the reputation gathered from the calibration exercises to the reputation derived when the participant ran the actual campaign. Since mobility traces were collected while the participants were performing the campaigns, we analyzed when they took images compared to when they had an opportunity. For GarbageWatch, prior information on all the waste bins locations existed and for the What's Bloomin and AssetLog campaigns, collective knowledge gathered from the participants submissions were used as ground truth. Table 2 shows the average of the percent difference of reputation for each participant in the three campaigns. The values are calculated by taking the difference between the calibration reputation and the reputation derived from the campaign and then averaging per campaign. The results, an average of 12.5% in reputation difference when considering all campaigns, indicate that the calibration exercises are reasonable approximations for participants real campaign reputations.

	GarbageWatch	What's Bloomin	AssetLog
Reputation Difference	10.3%	12.4%	14.8%

Table 2. Comparison of Calibration to Real Campaign Reputation

Progress Review: Checking Reputation Over Time. Since there is a chance that sampling behavior could change as campaigns run, it is important to check participant reputations at set points as part of a progress review. Introducing aging on top of the Beta distribution can help with this checkup since it can be used to obtain a more current indication of an individual's reputation. We exercise this feature by analyzing the contributions of an individual that was involved in What's Bloomin for an additional week. Figure 6 shows the reputation, along with the Beta curves, of this participant based on their first week, second week, and then two methods to combine their weeks (with and without aging). During the first week, the participant's reputation to sample when given the opportunity was 0.46. But on the second week, their reputation is much lower at a level of 0.15. At the two week period, if all contributions were considered equally, the participant's reputation would be 0.30 but this is not indicative of the recent performance change. Instead if an aging factor of 0.75 (where 1.0 represents keeping all history and 0.0 is only considering the current information) is used to discount past reputation daily, then the end reputation is 0.14 which is a better indicator of the recent behavior shift.

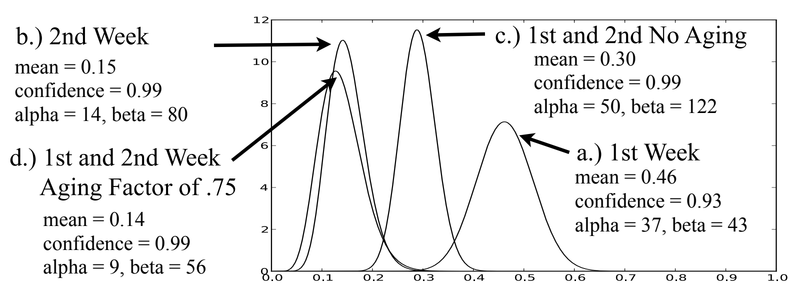


Fig. 6. Reputation for Participant Considering Aging Factor

7 Discussion

This section concludes the paper by summarizing lessons learned for campaign recruitment based on the evaluation results. Also, feedback provided by participants on their experience of performing campaigns is presented. Finally, future work that makes the recruitment system more flexible and adaptive is reviewed.

7.1 Recruitment Framework Analysis

The evaluation results reveal some important lessons for the recruitment framework. When analyzing the performance of the different algorithms during the assessment stage (Section 6.2), we find that selecting individuals based on using the greedy algorithm significantly improves coverage over random selection but only slightly compared to the naive approach. If there existed more coverage overlap between participants, the performance gap between greedy and naive algorithms would widen. This indicates that our recruitment framework is more useful when campaigns have a limited geographic scope (neighborhoods, city blocks) and have participants with higher mutual coverage.

Several individuals participated in multiple campaigns. When participant performance, in terms of sampling likelihood, was compared across campaigns, the individuals on the extremes, either on the high end where their reputation was above normal or vice versa, generally remained at those levels (top or bottom 5 in one campaign stayed in that same range in the others). This indicates that there is potential in using previous performance in similar campaigns to bootstrap reputation models. But a larger study with more varied participants needs to done to verify this conclusion. Also, in our campaign set, participants grew tired of collecting samples if the campaign lasted for an extended period of time. When individuals performed the campaigns for an additional week, their reputation was much lower. This points to the usefulness of the progress review step to check up on participants especially in long running campaigns.

7.2 Participant Experience Feedback

Participants were asked to fill out post-campaign surveys on their experience in performing the data collections. In terms of capturing data on the mobile phones, participants indicated that it was important that the act of data capture should be streamlined so that it can be repeated rapidly. Many participants also wanted mobile visualizations to help them participate more effectively. For instance, individuals desired a map interface colored by campaign coverage needs and an augmented reality browser to help discover nearby locations for participation. When asked if they would change their routines to participate in campaigns, most indicated that they would be willing to adhere to minor diversions but drastic changes would require extra incentives. Finally, participants stated that daily contribution summaries and in situ reminders would help increase participation.

7.3 Future Work

There are many opportunities to enhance the recruitment framework. The current system relies solely on past coverage and participation behavior. But contributors might be aware of impending changes in their schedule or habits. Individuals could specify a level of service they are willing to offer, and organizers could weight this projection based on participants' profiles and past negotiation fulfillments. Another area of exploration is whether more complex incentive models can help fix sensing gaps caused by inconsistent participants. Bonuses can be given if participants fill immediate campaign needs, and incentives can be scaled depending on context. Finally, the recruitment system should explicitly consider participant sampling bias. Ground truth from independent sources and parameters learned from all participant submissions can quantify this behavior.

Acknowledgments. This work is supported in part by NSF Cooperative Agreement #CCR-0120778 and NSF Grant #CNS-0627084. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding entities.

References

1. Campbell, A., Eisenman, S., Lane, N., Miluzzo, E., Peterson, R.: People-centric urban sensing. In: Proceedings of WiCOM, IEEE (2006) 18–32
2. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.: Participatory sensing. In: Proceedings of WSW, ACM (2006) 1–5
3. Paulos, E., Honicky, R., Hooker, B.: Citizen Science: Enabling Participatory Urbanism. Urban Informatics: The Practice and Promise of the Real-time City (2008)
4. Shallwani, S., Mohammed, S.: Community-based participatory research: Training Manual for Community-Based Researchers. Aga Khan University (2007)
5. Cooper, C., Dickinson, J., Phillips, T., Bonney, R.: Citizen science as a tool for conservation in residential ecosystems. Ecology and Society **12**(2) (2007) 11
6. Lu, H., Lane, N., Eisenman, S., Campbell, A.: Bubble-sensing: Binding sensing tasks to the physical world. Pervasive and Mobile Computing (2009)
7. Gaonkar, S., Li, J., Choudhury, R., Cox, L., Schmidt, A.: Micro-blog: Sharing and querying content through mobile phones and social participation. In: Proceedings of Mobisys, ACM (2008) 174–186
8. Reddy, S., Shilton, K., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Using Context Annotated Mobility Profiles to Recruit Data Collectors in Participatory Sensing. In: Proceedings of LOCA, Springer (2009) 52–69
9. Paxton, M., Benford, S.: Experiences of participatory sensing in the wild. In: Proceedings of Ubicomp, ACM (2009) 265–274
10. Shilton, K.: Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. Communications of the ACM **52**(11) (2009) 48–53
11. Hong, J., Landay, J.: An architecture for privacy-sensitive ubiquitous computing. In: Proceedings of Mobisys, ACM (2004) 177–189
12. Ashbrook, D., Starner, T.: Using GPS to learn significant locations and predict movement across users. Personal and Ubiquitous Computing (2003) 275–286
13. Kim, M., Kotz, D., Kim, S.: Extracting a mobility model from real user traces. In: Proceedings of Infocom, IEEE (2006) 1–13
14. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personal gazetteers: an interactive clustering approach. In: Proceedings of GIS, ACM (2004) 266–273
15. Liao, L., Fox, D., Kautz, H.: Location-based activity recognition using relational Markov networks. In: Proceedings of IJCAI, AAAI (2005)
16. Bhattacharya, A., Das, S.: LeZi-update: an information-theoretic approach to track mobile users in PCS networks. In: Proceedings of Mobicom, ACM (1999) 1–12
17. Soh, W., Kim, H.: Dynamic guard bandwidth scheme for wireless broadband networks. In: Proceedings of Infocom, IEEE (2001) 572–581
18. Hariharan, R., Toyama, K.: Project Lachesis: parsing and modeling location histories. In: Proceedings of GIScience, Springer (2004) 106–124
19. Krumm, J., Horvitz, E.: Predestination: Inferring destinations from partial trajectories. In: Proceedings of Ubicomp, ACM (2006) 243–260

20. Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. Wireless Communications and Mobile Computing **2**(5) (2002) 483–502
21. Lee, K., Hong, S., Kim, S., Rhee, I., Chong, S.: SLAW: A Mobility Model for Human Walks. In: Proceedings of Infocom, IEEE (2009) 855–863
22. Hsu, W., Dutta, D., Helmy, A.: CSI: A Paradigm for Behavior-oriented Delivery Services in Mobile Human Networks. ACM Transactions on Networking (2008)
23. Ghosh, J., Beal, M., Ngo, H., Qiao, C.: On profiling mobility and predicting locations of wireless users. In: Proceedings of REALMAN, IEEE (2006) 55–62
24. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing **10**(4) (2006) 255–268
25. Schlosser, A., Voss, M., Bruckner, L.: Comparing and evaluating metrics for reputation systems by simulation. In: Proceedings of Trust in Agent Societies. (2004)
26. eBay: The worlds marketplace. `http://ebay.com`.
27. Amazon: Online shopping center. `http://amazon.com`.
28. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems **43**(2) (2007) 618–644
29. SlashDot: News for nerds. `http://slashdot.com`.
30. Ganeriwal, S., Balzano, L., Srivastava, M.: Reputation-based framework for high integrity sensor networks. ACM Transactions on Sensor Networks **4**(3) (2008) 15
31. Jøsang, A., Ismail, R.: Beta reputation system. Bled eConference (2002) 324–337
32. Howe, J.: The rise of crowdsourcing. Wired Magazine **14**(6) (2006)
33. Amazon: Amazon mechanical turk. `http://mturk.com`.
34. GURU.com: Freelancer network. `http://guru.com`.
35. Slijepcevic, S., Potkonjak, M.: Power efficient organization of wireless sensor networks. In: Proceedings of ICC, IEEE (2001) 472–476
36. Kansal, A., Kaiser, W., Pottie, G., Srivastava, M., Sukhatme, G.: Reconfiguration methods for mobile sensor networks. ACM Transactions on Sensor Networks (2007)
37. Kapadia, A., Triandopoulos, N., Cornelius, C., Peebles, D., Kotz, D.: AnonySense: Opportunistic and privacy-preserving context collection. LNCS **5013** (2008) 280
38. Krause, A., Horvitz, E., Kansal, A., Zhao, F.: Toward Community Sensing. In: Proceedings of IPSN, ACM (2008) 481–492
39. Zheng, Y., Liu, L., Wang, L., Xie, X.: Understanding transportation mode based on GPS data for Web applications. ACM Transactions on the Web (2009)
40. Mun, M., Estrin, D., Burke, J., Hansen, M.: Parsimonious Mobility Classification using GSM and WiFi Traces. In: Proceedings of EmNets, IEEE (2008)
41. Reddy, S., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Determining transportation mode on mobile phones. In: Proceedings of ISWC, IEEE (2008) 25–28
42. Everytrail: Geotagging with everytrail. `http://everytrail.com`.
43. FourSquare: Explore your city. `http://playfoursquare.com`.
44. Kang, J., Welbourne, W., Stewart, B., Borriello, G.: Extracting places from traces of locations. Mobile Computing and Communications Review **9**(3) (2005) 58–68
45. Froehlich, J., Krumm, J.: Route prediction from trip observations. SAE (2008)
46. Khuller, S., Moss, A., Naor, J.: The budgeted maximum coverage problem. Information Processing Letters **70**(1) (1999) 39–45
47. Nemhauser, G., Wolsey, L., Fisher, M.: An analysis of approximations for maximizing submodular set functions. Mathematical Programming **14**(1) (1978) 265–294
48. Eagle, N., Pentland, A.: Eigenbehaviors: Identifying structure in routine. Behavioral Ecology and Sociobiology **63**(7) (2009) 1057–1066
49. Gonzalez, M., Hidalgo, C., Barabasi, A.: Understanding Individual Human Mobility Patterns. Nature **453**(7196) (2008) 779–782
50. Teacy, W., Patel, J., Jennings, N., Luck, M.: Coping with inaccurate reputation sources. In: Proceedings of AAMAS, ACM (2005) 997–1004