Using Context Annotated Mobility Profiles to Recruit Data Collectors in Participatory Sensing

Sasank Reddy, Katie Shilton, Jeff Burke,
Deborah Estrin, Mark Hansen, and Mani Srivastava

Center for Embedded Networked Sensing (CENS)
University of California, Los Angeles
{sasank,kshilton,jburke,mbs}@ucla.edu
destrin@cs.ucla.edu,cocteau@stat.ucla.edu
http://urban.cens.ucla.edu

Abstract. Mobile phones and accompanying network layers provide a platform to capture and share location, image, and acoustic data. This substrate enables participatory sensing: coordinated data gathering by individuals and communities to explore the world around them. Realizing such widespread and participatory sensing poses difficult challenges. In this paper, we discuss one particular challenge: creating a recruitment service to enable sensing organizers to select well-suited participants. Our approach concentrates on finding participants based on geographic and temporal coverage, as determined by context-annotated mobility profiles that model transportation mode, location, and time. We outline a three-stage recruitment framework designed to be parsimonious so as to limit risk to participants by reducing the location and context information revealed to the system. Finally, we illustrate the utility of the framework, along with corresponding modeling technique for mobility information, by analyzing data from a pilot mobility study consisting of ten users.

Key words: Participatory Sensing, Mobility Modeling, Location Based Services

1 Introduction

Mobile phones and the cellular infrastructure are increasingly used for more than just communication. These platforms are being employed as tools to understand the habits and environments of individuals and communities. Many mobile phones are already equipped with acoustic, image, and location sensors in the form of microphones, cameras, and GPS, Wi-Fi, or cellular positioning, and contain a Bluetooth interface that can connect to external sensors. Phones also enable users to enter text to describe or record events. Emerging services take advantage of these diverse modalities to support active living, dietary monitoring, and analysis of environmental impact [1–3]. A complimentary vision, referred to as participatory or urban sensing, proposes tools to engage potentially large numbers of the general public in coordinated data gathering [4, 5].

Intended to enable and encourage anyone to gather and investigate previously invisible data, participatory sensing can support advocacy - “making a case” through distributed documentation of a community need or an issue.

Mobile phones allow for easy, convenient and widespread data collection. Complimentary tools take advantage of phone networks and enable anyone to coordinate participants and initiate data collection “campaigns” that focus on social, political, or urban processes. Several challenges must be met for participatory sensing applications to flourish [6, 7]. Among these is recruitment: enabling organizers, who may be community groups or simply motivated individuals, to select an interested and well-suited set of participants for a campaign based on the needs and specifications of the case they want to make. Campaign organizers might base recruitment around a number of factors. In this paper, we focus on enabling selection based upon the geographic and temporal availability of participants. Specifically, we describe a framework that enables organizers to select a group of participants based on a campaign’s contextual (further detailed in Section 3), spatial, and temporal coverage constraints. Furthermore, this recruitment system is designed to be parsimonious in its usage of context and location information to minimize disclosure risk for participants.

The paper is organized as follows: Section 2 describes a motivating application that illustrates why availability-based recruitment is useful. Section 3 discusses the recruitment problem in more detail. Section 4 overviews the design goals of the system and how the proposed recruitment stages adhere to these goals. System details are presented in Section 5, and we evaluate the recruitment framework using results from a pilot mobility study in Section 6. The paper ends with a discussion of future work items in Section 7.

2 Motivational Applications

Not all participatory sensing campaigns will need to select a limited pool of well-suited participants. Many campaigns will benefit from engaging as many people as possible, gaining the advantages of diverse backgrounds, interest, and availability [4]. Some campaigns, however, may face constraints that prevent them from incorporating all interested volunteers. If campaign organizers provide financial incentives for participation, distribute hardware to volunteers, or train individuals in specialized data collection, organizers may need to limit the number of possible participants or focus on select subset. Coverage-based recruitment can help organizers make selection decisions according to participants’ geographic and temporal availability.

An example campaign that benefits from coverage-based recruitment is inspired from the sustainability initiative at UCLA. In recent years, there has been a strong focus on raising campus awareness of environmental sustainability [8]. In coordination with the Education for Sustainable Living Program, we are undertaking a project to create a “green map” of campus sustainability resources. Volunteers will chart recycling bins and bicycle racks by taking geotagged photos [9]. They will also document negative impacts such as improper waste disposal (recyclables in regular bins) and inefficient energy usage (lights continuously left on outdoors during daytime). Having images along with location information is important since it provides a visual reference of the point of interest and enables individuals to later annotate with additional data, for example noting the number of slots in a bike rack or the type of recyclable thrown in a regular waste basket. The campus sustainability campaign will take advantage of the geographic coverage that a selected group of volunteers can provide by asking volunteers to contribute information that they see as they go about their regular routines. The participants will focus on mapping "tasks" that run weeks at a time. The information that is collected will be used to improve current sustainability processes (e.g. by suggesting better placement of recycle bins or informing facilities where efficiency problems exist) as well as to help educate the UCLA community of areas for improvement.

The campus sustainability campaign will benefit from coverage-based recruitment because it depends on observations in a limited region (UCLA) over a long period (several weeks). Further, during the campaign, we will provide participants with the mobile devices and accompanying data plan necessary to perform the data collection. We will also hold training sessions to help participants understand the mapping tasks and goals. For these reasons, we must focus resources on a select number of participants. Thus, it is important to recruit participants whose availability, in terms of context (in this case transportation mode), space, and time, matches campaign needs. The campaign will remain open for anyone on campus to participate with their own mobile phones, but the hardware resources and training will be limited to a select few.

In this campaign, well-suited participants regularly walk on campus during the daytime (individuals that run, bike, or drive may be less likely to notice points of interest, and collecting clear photographs is more difficult at night); cover as much area as possible; and are consistent in their routes and routines. Recruiting participants whose availability matches the campaign coverage constraints will provide the best chance of documenting sustainability resources using a limited pool of individuals in a constrained time frame. Throughout the rest of the paper, the campus sustainability campaign is used as an example to explain the details of the recruitment framework.

Although we detail only one campaign as an example here, many other data collection initiatives could benefit from a coverage-based recruitment system. Examples include the CycleSense initiative, which tasks bike commuters to collect information about the quality of bike routes and paths, and the Walkability project, designed to gather safety issues of walking paths in local neighborhoods [10, 11].

3 Problem Description and Challenge

Like many crowd-sourcing services on the web [12], a campaign seeks interested participants willing to volunteer their time to help with the data collection task. For certain campaigns it might be appropriate to focus on a specific set of volunteers from an overall pool. In this situation, choosing volunteers wisely becomes critically important. Note that this parallels the recruitment that occurs in web services that provide a marketplace for commissioned work, such as Amazon Mechanical Turk and GURU.com, in that well-suited individuals are preferred [13, 14]. Organizers of campaigns may wish to consider a number of factors. For example, organizers could request that participants have specific sensor capabilities (e.g. camera, microphone). Organizers may also wish to recruit participants who have certain performance standards based on previous involvement in campaigns [15] and are willing to be flexible to conform to sensing requests. This paper, however, focuses on a third requirement: the geographic and temporal availability of participants.

At a technical level, the recruitment problem in participatory sensing is similar to that of static sensor selection and placement [16] and robotic motion coordination for sensing [17]. The distinction is that participatory sensing must consider human mobility, which is not directly controllable. However, mobility-based recruitment does takes advantage of the fact that people's routes and locations are often regular and repetitive [18, 19]. This work differs from existing systems that use mobile phones for sensing by concentrating on selecting a set of participants so that geographic coverage can be maximized. Our approach also takes into account variation of participant mobility over time. Previous work has geared task assignment for opportunistic in-situ sensing [20, 21] or has focused on initiating sampling around specific location "bubbles" (regions) [22, 23]. Furthermore, our work focuses on campaigns targeting phenomenon that are not easily expressed as Gaussian processes or do not have spatial and temporal distributions that are known a priori [24]. For instance, locations of sustainability assets are often non-uniformly distributed and instances of improper resource (energy, waste, water) usage are dependent on processes that are dynamic based on space, time, and other external factors (which will need to be learned through data collection).

We outline a recruitment engine that uses campaign specifications provided by an organizer to select a limited set of potential volunteers based on participants' previously-gathered mobility profiles. A mobility profile is derived from the participants' context-annotated mobility traces: streams of previously-collected location, time, and context data. Location and time are obtained via GPS receivers embedded in mobile phones or from cellular network infrastructure [25]. Context includes a number of inferences drawn from sensors available on the mobile phone, but we specifically concentrate on the transportation mode (walking, running, biking, or in motorized transport) of an individual which can be obtained by analyzing GPS traces [26].

To define mobility-based requirements for their campaign, a campaign organizer would limit a campaign to geographic regions (e.g. UCLA), temporal boundaries (e.g. between 2/1/09 and 6/1/09), and specific time period (e.g. weekdays between 8 a.m. and 6 p.m.). Organizers could also specify important contextual information, such as preferred modes of transportation. The recruitment engine would match these specifications with participants' mobility profiles.

The challenge for a engine is finding a subset of volunteers whose combined mobility profiles best fulfills the coverage requirement. Finding the appropriate set of participants requires iterating through all subset combinations of individuals considering ordering since availability could be redundant. And since participant routes and habits may differ over the course of a campaign, the recruitment system needs to analyze whether the mobility profiles of participants change as the campaign runs and alert organizers to possible coverage problems.

4 Design Goals

Recruiting volunteers for participatory sensing is somewhat analogous to recruiting volunteers or employees in non-virtual environments. Considerations include geographic and temporal availability as well as qualifications and willingness. Drawing on this similarity, we have created a coverage-based recruitment system that consists of three distinct stages: the qualifier, interview, and progress review, modeled after real-world recruitment processes. Figure 1 shows each of these three recruitment steps with the perspective of campaign actions. Furthermore, an outline of the three stages exists below.

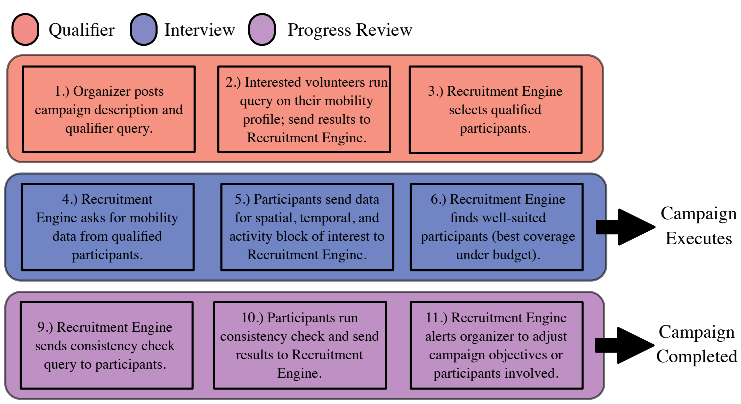


Fig. 1. Campaign Recruitment Flow with Recruitment Steps Labeled

- **The Qualifier:** To be considered for a campaign, potential participants must have a certain number of significant locations, routes, and desired transportation modes within a certain time and space.
- **The Interview:** The recruitment engine compares participants that meet initial qualification requirements to determine which limited subset of individuals maximize coverage over a specific area and time period.
- **The Progress Review:** As a campaign runs, the recruitment engine checks the coverage consistency of participants by periodically evaluating their mobility during the campaign against their qualifying profile. If similarity scores

are below a defined level, the engine alerts campaign organizers so they may take action (check for coverage, provide feedback, etc.)

During the design of the recruitment system, we were concerned with how best to handle querying and sharing of participant mobility information. Because this information describes an individual's routines and habits, sharing mobility profiles presents a number of exposure risks to individuals [27–29]. Thus, a major design principle employed for the recruitment system was parsimony: sharing the minimal amount of information needed to meet an application's goals [30, 31]. The recruitment system uses a framework that limits the amount and granularity of mobility information shared during each step of the recruitment process. In this section, we detail how our system achieves this design goal by describing how we minimize what information is shared, with whom it is shared, and how long it is retained during each stage of the recruitment process. All of these considerations are important factors in aiding participants to understand and control the privacy of their data [32]. We assume that all mobility information used to generate a profile resides with the participant in a private data store, and participants must opt-in to allow queries to be run on this store.

If a participant is interested in a campaign, they can opt to run the qualifier query on their mobility information. These queries are created by campaign organizers and run on the participant's data store. Only aggregate results are shared with a campaign organizer (e.g. whether geo-spatial qualifications have been met or how consistent a profile is). The campaign organizer therefore has no access to a participant's detailed location information during the qualifier stage. Only during the interview stage does data from the mobility profile need to be shared with campaign organizers. In order to find a cover set among the participants, the recruitment engine needs access to location information. At this stage, however, not only have participants expressed interest in a campaign, they have a good chance of being chosen to participate. If a participant trusts a campaign organizer and has applied to their campaign, more information sharing may be justifiable. Parsimony, however, remains important. Participant mobility data shared with the recruitment engine is limited to a particular spatial region (e.g.UCLA) and time span (e.g. weekdays). Also, rather than sharing granular location information, the recruitment engine provides the organizer with a generalized coverage map of the chosen region (see Figure 4 in Section 6.3). If the coverage area is small, an organizer may be able to infer participants' routes, but a targeted coverage area limits exposure of locations not relevant to the campaign. Once the interview process has ended, the system deletes the shared location data and only maintains the participant subset coverage data for the duration of the campaign. Finally, like the qualifier, the progress review is run on the participant's store, and only aggregate measures of consistency are shared with campaign organizers. An organizer will know, for example, if a participant follows similar routines as determined by the interview process, but exact routes and locations will not be reported.

5 System Details

The steps involved in the recruitment process and the pre-processing procedures needed to build mobility profiles are detailed in this section. We place particular emphasis on describing the underlying algorithms employed. We also explain the inputs and outputs of each stage along with parameters that need to be specified for execution of the steps in the framework.

5.1 Building Mobility Profiles

Each of the recruitment stages depends on the transformation of a participant's raw mobility information into elements that can be used for recruitment. We assume that participants have previously used their mobile phones to collect raw data in the form of GPS traces (latitude, longitude, time). Services already exist to collect and manage this type of data [33, 34], and we expect that historical location data will become increasingly available. Mobility information can also be obtained by having individuals annotate maps manually, but in this paper we focus on mobility data that is verified through actual in-field data collection. We plan to handle heterogeneous sources of mobility information along with allowing individuals to specify different granularities of data as future work (Section 7).

The recruitment engine can process raw mobility data to infer significant locations and routes along with associated transportation modes. Over time, this processed data can form participant mobility profiles. The steps required to build these profiles include:

- **Pre-Processing:** Since sampling rate may vary, the engine first normalizes the GPS logs to a set sample rate (every 30 seconds in our experiments) and fills in missing values (e.g. when an individual goes indoors and loses GPS signal). For cases where there is a gap in data and the points before and after are significantly different, the engine generates a likely route using Google Route generator, taking into account time and distance traveled [35].
- **Destination and Route Inference:** Next, the engine finds places where users spent a continuous amount of time (15 minutes) within a certain distance (50 meters) and considers this a "stay" [36]. Stays within a distance (250 meters) are clustered into "destinations" using density based clustering [37]. Destinations are then used to divide a GPS log into significant locations and routes. Time and distance thresholds were chosen based on varying the parameter combinations and analyzing which were most effective at representing participants' notions of destinations. Routes are clustered using average minimum point-segment distance as the comparison method (threshold of 100 meters) and performing hierarchical clustering [38].
- **Route Transportation Mode Inference:** Once routes have been identified, they need to be labeled in terms of transportation mode. The routes are first segmented based on change points (speed close to zero, loss of GPS signal) [26]. Then features such as the average, maximum, and minimum speeds, and total distance for each segment are calculated, and used to classify in

terms of transportation modes with likely transitions considered [26]. Also, we have explored using additional sensors on phones, such as accelerometers and GSM/WiFi radios, to help with classification [30, 39].

5.2 Running the Qualifier

Campaign organizers can use participants' mobility profiles to select volunteers from as many interested, qualified candidates as possible. This is analogous to a job or volunteer hiring process where recruiters only want to see resumes from a pool of candidates who meet a set of minimum requirements. A participatory sensing organizer would therefore post a description of a campaign including its purpose, type of data needed, lifetime, and contribution expected. The organizer would also specify qualifications based on geographic and temporal coverage by using attributes such as the total number, uniqueness, transportation mode, and the minimum time needed for locations and routes in a specific spatial region and time period. The qualifier can be defined based on routes that have start/end points that are in a specific zone as well.

The recruitment engine would run the qualification filter on interested participants' mobility profiles, and only share with the organizer whether an individual meets the criteria. Interested participants who do not explicitly meet the qualification may choose to share more detailed information, such as how many more routes or locations they would need to meet the qualifier, so that the organizer has the choice to include them if needed. The consistency of a participant's mobility information can also be used as part of the qualification. We discuss this concept in more detail in the progress review stage below.

5.3 Running the Interview

Once the recruitment engine has identified a pool of participants who meet minimum qualifications, the next step is the interview. In traditional employment or volunteer recruitment, the interview phase gives employers a better understanding of the skills possessed by a set of candidates. Similarly, the participatory sensing interview evaluates the mobility qualities of a set of potential participants, and calculates which subset would maximize a coverage utility function.

In a technical sense, the coverage-based interview process is an instance of the budgeted maximum coverage problem [40]. Essentially, we have a participant pool $P = \{p_1, p_2, ..., p_n\}$ with individual non-negative costs $\{c_i\}_{i=1}^{n}$ defined over a spatial and temporal block elements $E = \{e_1, e_2, ..., e_n\}$ with associated utilities $\{u_i\}_{i=1}^{n}$. The goal is to find a subset of participants, $P^* \subseteq P$, for the campaign such that the utility of elements covered by subset, $U(P^*)$, is maximized while the cost of the subset, $C(P^*)$, is under a set budget, B, [24, 40, 41]. In summary, we are solving the following optimization problem:

$$argmax\ U(P^*)\ subject\ to\ C(P^*) \leq B$$

The budget is modeled as the resources needed to engage participants in a campaign. Each participant has a cost, which might include both compensation and a system opportunity cost for the individual's participation (i.e. if adding a participant elicits organizational or administrative costs for a campaign). The spatial and temporal blocks, E, are defined by specifying a region and time of interest (e.g. UCLA, weekdays from 8 am - 6 pm), spatial and temporal granularities (e.g. 10000 meter^2 blocks, 5 minute time spans), and a set of transportation modes. Utility weights are associated with each block and are specified by the campaign organizer. To get a more intuitive sense of how a utility weights can be assigned, we consider two cases: uniform weighting and priority based. In the first, all spatial and temporal blocks have equal weights, so the utility of each block could be set uniformly to 1 and the maximum aggregate utility is simply the number of blocks. The second case is priority based where certain blocks are more important to be sensed. This can be reflected by assigning higher utility values to the priority blocks and lower values to less important blocks.

The optimization problem of finding the best subset of participants is proven to be NP-hard [40, 42]. Selecting a participant to be involved in a subset changes the utility for the participants not included, and one would have to search through all combinations of subsets to find the best solution. Since our utility function is both sub-modular (adding a participant helps more if fewer participants are included in the subset and less if there are more participants are in the subset already) and non-decreasing (the utility of the subset is less than the set that it came from), the standard greedy algorithm is guaranteed to find a constant fraction (63%) solution when the costs of the participants are the same [42]. In the more complex setting where the costs of participants varies, a benefit-cost greedy algorithm where the ratio of utility to cost is considered when adding participants could be employed to obtain a solution [41]. Also, the dual problem of minimizing cost of the set given a particular coverage requirement can also be solved using the greedy algorithm with a log-factor approximation of the optimal value.

5.4 Running the Progress Review

The end of the interview process results in selecting a set of participants to execute the campaign. For certain campaigns, there may need to be additional checkups to ensure that the mobility of participants is consistent with their pre-established profiles and coverage over the area of interest is being maintained. This is especially true for campaigns with a long lifetime, during which people's routines may change. Thus, a "progress review" can check a participant's original mobility profile for similarity with current behavior. The campaign organizer sets the intervals for, and frequency of, reviews.

In order to check for similarity, the recruitment engine models a participant's mobility profile for a particular period of time using an m x n matrix with a set of corresponding transportation modes [19, 43]. The m rows in the matrix represent spatial blocks (e.g. 10000 meter^2 grids, zip codes, census blocks) while the n columns model distinct time spans. An entry in the matrix would be the proportion of time spent performing a set of transportation modes within the time period of interest in a particular spatial block for a time span. Based on previous work that explores daily human location patterns, we chose a day as a representative time span for our analysis [18, 19, 43]. This matrix is normalized by applying the arcsine-root transformation [44]. Throughout the rest of the paper, we refer to this data construct as an "association matrix", A.

Since we are interested in dominant mobility patterns, a summarization method is necessary. An effective technique to analyze multi-dimensional data for major patterns is Principal Component Analysis (PCA). PCA finds a set of orthogonal axes (eigenvectors) that linearly transform the original data so that the largest amount of variance is explained [45]. The first coordinate axis explains the most variation in the data and the subsequent axes explain a decreasing amount of variation. The eigenvalues obtained from PCA explain the amount of variation for which each eigenvector accounts. To obtain the dominant patterns, the original data is projected onto the eigenvectors weighted by the eigenvalues. Oftentimes when analyzing location association data, the resulting projections are referred to as "eigenbehaviors" [19, 43].

In practice, a more flexible technique to find eigenbehaviors is Singular Value Decomposition (SVD) [45]. Given the association matrix A, the SVD would be:

$$A = U \cdot \Sigma \cdot V^t$$

In this decomposition, U and V are the left and right singular vectors and Σ is the diagonal matrix of singular values. Thus, in the terms of PCA, the singular values are the square roots of the corresponding eigenvalues, the columns of V represent the eigenvectors, and the columns of U are the eigenbehaviors [19, 43, 45]. Since the eigenbehaviors represent patterns that are common across different time spans (days) and the singular values represent the importance of each pattern, one can compare consecutive time periods by taking the cosine similarity of the behavior vectors weighted by the singular value importance (percentage of variance represented by a singular value) [43]. Hence, if there exists two eigenbehaviors, U_{t1} and U_{t2}, representing different time periods, $t1$ and $t2$, with singular value importance, W_{t1} and W_{t2}, the similarity metric would be defined as:

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

This measure of similarity [43] (or consistency) is indexed from 0 (least similar) to 1 (most similar) by normalizing on the base eigenbehavior similarity. Ideally, not all eigenbehavior vectors need to be used. If a large variation of the behaviors is explained by a few vectors, then the remainder of the components can be ignored. This SVD based approach has been used previously to analyze videos, documents, and cellular/WiFi association patterns for clustering purposes [19, 43, 46, 47]. We instead focus on using this technique as a measure of an individual's consistency of mobility patterns over time.

6 Experiments

Because we are in the preliminary stages of several campaigns, we base our evaluation of the coverage-based recruitment system on mobility profiles generated during a pilot mobility study. Running this data through our recruitment process, with the campus sustainability campaign as context, illustrates the performance and usefulness of the different stages of our approach.

6.1 Generating Sample Mobility Profiles

We recruited ten participants for the pilot study and asked them to carry a Nokia n95 to automatically capture and upload a GPS time series (latitude, longitude, and speed every 30 seconds). We used this information to create transportation annotated mobility profiles for the participants. The study was conducted over a time-span of sixty five days where seven users were recruited at the start and another three were phased in during the mid-point (due to equipment availability). Participants were requested to log information at all times, but had the right to turn off their devices at any time (Figure 2). Participants were all affiliated with UCLA so that a common area of coverage would be available.

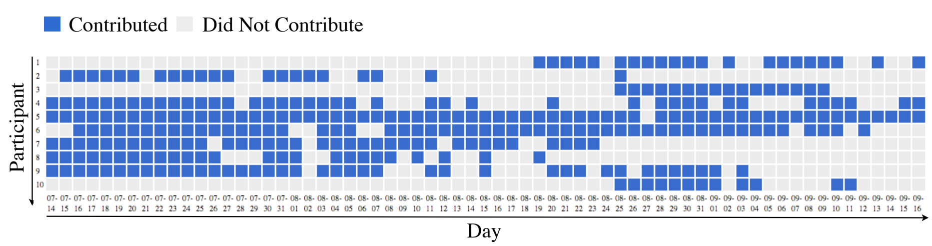
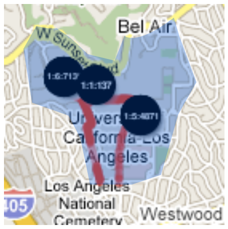


Fig. 2. Mobility Study Participation Statistics (Organized by Participants/Days)

6.2 Qualifier: Evaluating UCLA Destinations and Walking Routes

For the campus sustainability campaign, well-suited candidates would have walking routes to and from places on campus. Thus, a qualifier filter for this campaign could be: "Participants that have at least 4 daytime walking routes with destinations starting or ending at UCLA during a week." Using this qualifier on pilot participants' sample 5-weekday stretches indicates that seven participants (#3, #4, #5, #6, #7, #8, #9) would qualify for the campaign while three (#1, #2, #10) would not. Figure 3 shows the mobility information during a 5-weekday span for one participant that matches the qualifier (#8) and one that does not (#2). Further, the figure provides the walking route statistics for each participant. User #8 has many walking routes on campus (this individual lives on campus and prefers walking to other transport modes). User #2 has only one walking route (this participant drives to campus and parks under their office).



a.) Participant #8 TAM Data

b.) Participant #2 TAM Data

Participant	# Walk Routes
1	2
2	1
3	6
4	9
5	8
6	4
7	8
8	6
9	5
10	2

c.) Qualifier Filter Results

Fig. 3. Mobility Information for Participants and Qualifier Results

6.3 Interview: Evaluating Best Coverage of UCLA

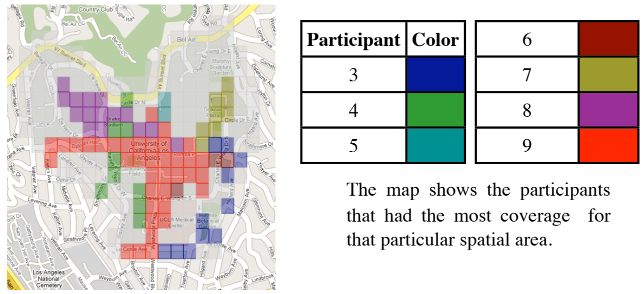
To evaluate the coverage of the UCLA campus by qualified participants, we perform a set cover for a daytime hours during a 5-weekday span where the spatial granularity is set to 10000 m^2 blocks, the time block granularity is 1 hour, and the activity is walking. The results of maximizing coverage, with the budget set to infinity, and individual participants costs and block utility weights all set evenly, are shown in Figure 4.

The results indicate that participant #3 is the most valuable since he spans the most spatial/temporal blocks. In terms of spatial area alone, participants #8 and #9 provide the most coverage. These two individuals both live near campus (in close proximity) and often walk to UCLA from home. Notice that participant #9's score is much higher than #8's. This is due to the fact that #9 and #8 have common coverage, and the common areas count toward #9's score, since she spans more blocks overall. The interview process can also adapt to campaign budgets and participant costs. For instance, if the campaign had a budget of five credits, where each individual had a cost of one credit, then the interview process would eliminate participants #5 and #4.

In the above analysis, the specifications for the spatial and temporal blocks were chosen in a "generic" fashion to illustrate the capabilities of the interview process based on the mobility data available. But for the sustainability campaign, different specifications might exist for the attributes. For instance, emphasis could be placed in certain regions (around specific buildings) and time periods (during peak campus occupancy), the time granularity could be changed ("day" level indexes of time as opposed to individual hours and spatial blocks contoured to walking paths on campus).

Participant	Coverage Score
3	101
4	88
5	51
6	44
7	74
8	87
9	96

a.) Maximum Cover Results

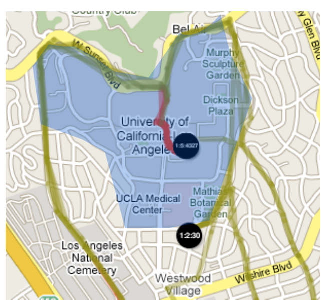


b.) Map of Coverage Results

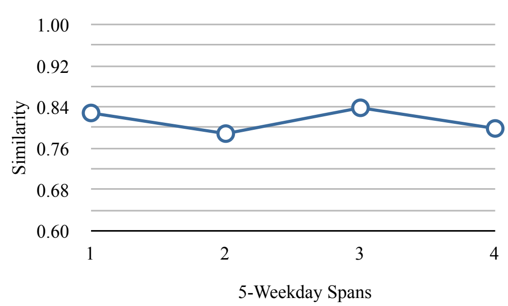
Fig. 4. Maximum Walking Coverage of UCLA During Daytime Weekday Hours

6.4 Progress Review: Comparing Similarity of Profiles Over Weeks

If participants' availability deviates from their established mobility profiles during a campaign, the organizer should be alerted. The inconsistent participant could be checked to see if their coverage still meets coverage requirements by running the interview process or the organizer might need to recruit additional participants to help with the campaign. The progress review consistency "check up" is especially important for long running data collections, such as the campus sustainability campaign, since participants' behavior can change over time.



a.) Participant #5 Base Profile

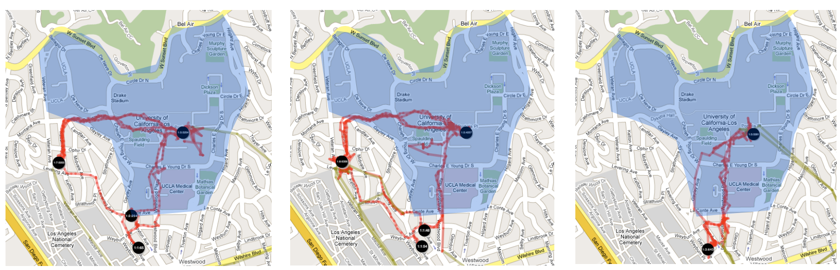


b.) Participant #5 Similarity Results

Fig. 5. Consistency of Participant #5's Mobility Profile (Base Compared to 5 Weeks)

To demonstrate the utility of the performance review, the availability changes of two participants (#5 and #9) in the pilot mobility study is analyzed. Note that since the spatial granularities are set again to 10000 m^2 blocks and a 5-weekday span is analyzed, the association matrix is 208 (representing the blocks at UCLA) by 5 (number of days) in size. Participant #5 provided GPS information for the longest period, and an interview determined that the individual was very consistent in his routes and habits over the time span. To test whether the progress review technique would corroborate this consistent behavior, we employed SVD to obtain eigenbehaviors based on daytime walking instances for the first 5-weekday span, and then compared them with the eigenbehaviors from the next four 5-weekday spans at the UCLA location. The top three eigenbehaviors were used, in this comparison and the one below, since they represented over 90% of the variance (using all five components results in similar consistency measures as considering the top three). Figure 5 (a.) illustrates the mobility information for the base week, and Figure 5 (b.) depicts the similarity results for the progress review. The average similarity over the time span was 0.82.

Not all users are as consistent as Participant #5. For instance, Participant #9's mobility drastically changed over the campaign period. An interview revealed that the participant moved her home address during the time period. Thus, her daily routes and locations changed significantly. We performed a consistency check comparing two consecutive 5-weekday spans, between which the move occurred, to her base profile (Figure 6). As expected, the first week is fairly consistent with the base profile, achieving a similarity score of 0.95. The second week is significantly different, achieving a similarity of only 0.67.



a.) Participant #9 Base Profile

b.) Comparison #1 with Similarity Score of 0.95

b.) Comparison #2 with Similarity Score of 0.67

Fig. 6. Consistency of Participant #9's Mobility Profile (Base Compared to 2 Weeks)

As shown above, it is not always necessary to use all the eigenbehaviors when performing a consistency check since dominant components could exist that explain the majority of the variance. To further illustrate this point, we analyzed all participant mobility data in regards to the UCLA campus and daytime walking patterns with 10000 m^2 spatial blocks which results in a 208 by n coverage days association matrix. Table 1 shows the number of days in which there was a coverage match for each participant along with the fraction of variance represented by the top three eigenbehaviors. Based on the analysis, the top three components can represent an average of 0.76 fraction of the variance of the participants' behavior. These dominant patterns represent walking behavior influenced by location choices for work and going to different places for dining.

Table 1. Participant UCLA Walking Totals and Variance Fraction Represented by Top 3 Components

Participant	# of Days	Variance Fraction
1	5	0.90
2	8	0.70
3	4	0.94
4	13	0.62
5	28	0.64
6	8	0.74
7	8	0.70
8	12	0.69
9	15	0.78
10	4	0.89

7 Future Work

We are currently building the overall system to manage campaigns: enabling organizers to easily create, manage, organize data collections, and enabling participants to find and participate in campaigns in which they are interested in. This final section details future work on campaign recruitment. This includes exploration of making the recruitment process clear for participants involved, as well as enhancements to make the recruitment more flexible and effective.

7.1 Designing for System Legibility

Because a coverage-based recruitment system demands that participants share information about their locations, habits and routines, privacy is an important concern. We define privacy as a negotiation between participants and organizers of what information to share or withhold. Privacy decision-making is contextual and individually variable [48–50].

To negotiate information disclosure, individuals should understand the benefits and repercussions of sharing their data. To make an informed decision, they must understand who is asking for the data (identity); what the data will reveal about them (granularity); what the organizer wants to use the data for (purpose); and how long data will be retained by a requesting organizer (retention). Through each stage, the system communicates to participants the nature of the query they are running, what information the query shares and with whom, and how long the information can be retained by the campaign organizer. The system should also communicate the identity and reputation of a campaign organizer to potential participants, so that participants can decide whether to trust an organizer with their mobility information. Communication between the system and the participant is essential to the system's legibility: the ways in which a system enables people of all technical backgrounds to make informed disclosure decisions. Our ongoing work explores methods to represent disclosure to participants to aid informed decision-making.

7.2 Recruitment Process Enhancements

In this paper, we focused on coverage-based recruitment, but there are other elements that future work will add to the recruitment process as well. In addition to a participant's sensing capabilities and their past campaign performance and participation, we plan to integrate social network membership or external credentials as factors in recruitment [15]. Furthermore, we will consider a participant's responsiveness, which includes both their willingness and flexibility to perform specific sensing requests that might deviate from their own objectives when tasked by the campaign organizer.

Another area to investigate is how the level of parsimony affects the uncertainty of a participant's qualification for a campaign. This will most affect the interview stage of the recruitment process where mobility profile information needs to be shared with the recruitment engine as opposed to results of a query. If a campaign organizer has coverage constraints based on coarse level mobility information, it will be less of a disclosure risk for participants but it might make figuring out which set of participants to recruit much more difficult. On the other hand, creating coverage requirements that require fine-grained information might result in obtaining a more well-suited participant set, but has the downside of a higher privacy risk. We are interested in designing tools so that organizers can balance participant disclosure risk and qualification uncertainty.

The current recruiting system has a fairly rigid structure: a campaign is defined and participants are compared against organizer specifications based on profile information. To make the system more adaptable, a negotiation step, where participants can specify a level of service to which they are willing to commit, could be incorporated as well. Also, the system could allow participants to specify different incentive criteria including not having any specified at all, being rewarded based on performance, or just obtaining a flat reward for participation as a whole. Additionally, coverage-based recruitment relies on having access to historical mobility information, but this data might not always be available. Thus, the framework should be able to incorporate manually-specified availability as well. An approach we will explore is placing confidence weights on different sources of coverage information. Also, the mobility information that is specified might range in terms of granularity, so the system would need to represent availability uncertainty based on the resolution of data provided. Finally, we are interested in exploring how the recruitment system can operate during the campaign as a run-time tool for the designer. Thus, exploring the robustness and accuracy of the mobility analysis technique (employing eigenbehaviors) when faced with daily or sub-weekly updates will be a point of analysis along with exploring how feedback and participant changes affect reaching the campaign coverage goals.

8 Acknowledgments

This work is supported by NSF Grant CNS-0627084 and IIS-0832873. Hardware is provided by Nokia Research Labs under the Sensor Planet initiative.

References

1. Consolvo, S., McDonald, D., Toscos, T., Chen, M., et al.: Activity sensing in the wild: a field trial of ubifit garden. Computer Human Interaction (2008)
2. MycaNutrition: MyFoodPhone Nutrition. http://mycanutrition.com (2008)
3. CENS: Personal Environmental Impact Report. http://peir.cens.ucla.edu (2008)
4. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.: Participatory Sensing. World-Sensor-Web, Sensys (2006)
5. Eisenman, S., Lane, N., Miluzzo, E., Peterson, R., Ahn, G., Campbell, A.: MetroSense Project: People-Centric Sensing at Scale. In: World-Sensor-Web, Sensys
6. Campbell, A., Eisenman, S., Lane, N., Miluzzo, E., , et al.: The Rise of People-Centric Sensing. IEEE Internet Computing **12**(4) (2008) 12–21
7. Parker, A., Reddy, S., Schmid, T., , et al.: Network System Challenges in Selective Sharing and Verification for Personal, Social, and Urban-Scale Sensing Applications. Hotmobile (2006)
8. UCLA Insitute of the Environment: Sustainability. http://sustain.ucla.edu/ (2008)
9. GreenMap: Green map - think global, map local. http://greenmap.org/ (2008)
10. Shilton, K., Ramanathan, N., Reddy, S., Samanta, V., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Participatory Design of Sensing Networks: Strengths and Challenges. Participatory Design Conference (2008)
11. LiveablePlaces: Liveable Places, Making the Connections Community Workshop. http://liveableplaces.org (2008)
12. Brabham, D.: Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence **14**(1) (2008) 75
13. Amazon: Amazon mechanical turk. http://mturk.com (2008)
14. GURU.com: Freelancers at online service marketplace. http://guru.com (2008)
15. Reddy, S., Shilton, K., Burke, J., et. al.: Evaluating Participation and Performance in Participatory Sensing. UrbanSense, Sensys (2008)
16. Ganesan, D., Cristescu, R., Beferull-Lozano, B.: Power-efficient sensor placement and transmission structure for data gathering under distortion constraints. ACM Transactions on Sensor Networks (TOSN) **2**(2) (2006) 155–181
17. Kansal, A., Kaiser, W., Pottie, G., Srivastava, M., Sukhatme, G.: Reconfiguration methods for mobile sensor networks. Transactions on Sensor Networks (2007)
18. Gonz, M., et al.: Understanding individual human mobility patterns. Nature **453**(7196) (2008) 779
19. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing **10**(4) (2006) 255–268
20. Miluzzo, E., Lane, N., Campbell, A.: Virtual Sensing Range (Poster Abstract). In: Proc. of ACM 4th Intl Conf. on Embedded Networked Sensor Systems. (2006) 1–3
21. Eisenman, S., Lane, N., Campbell, A.: Techniques for Improving Opportunistic Sensor Networking Performance. In: Distributed Computing in Sensor Systems: 4th IEEE International Conference, Dcoss 2008 Santorini Island, Greece, June 11-14, 2008, Proceedings, Springer (2008)
22. Lu, H., Lane, N., Eisenman, S., Campbell, A.: Bubble-Sensing: A New Paradigm for Binding a Sensing Task to the Physical World using Mobile Phones. Workshop on Mobile Devices and Urban Sensing, IPSN (2008)
23. Gaonkar, S., Li, J., Choudhury, R., Cox, L.: Micro-Blog: sharing and querying content through mobile phones and social participation. Mobisys (2008)
24. Krause, A., Horvitz, E., Kansal, A., Zhao, F.: Toward Community Sensing. In: ACM Conference on Information Processing in Sensor Networks (IPSN). (2008)

25. Hightower, J., Borriello, G.: A survey and taxonomy of location systems for ubiquitous computing. IEEE Computer **34**(8) (2001) 57–66
26. Zheng, Y., Liu, L., Wang, L., Xie, X.: Learning transportation mode from raw gps data for geographic applications on the web. WWW Conference (2008)
27. Anthony, D., Henderson, T., Kotz, D.: Privacy in Location-Aware Computing Environments. Pervasive Computing, IEEE **6**(4) (2007) 64–72
28. He, Q., Wu, D., Khosla, P.: The Quest for Personal Control over Mobile Location Privacy. IEEE Communications Magazine (2004)
29. Shahabi, C., Khoshgozaran, A.: Location Privacy in Geospatial Decision-Making. Lecture Notes in Computer Science **4777** (2007) 1
30. Mun, M., Estrin, D., Burke, J., Hansen, M.: Parsimonious Mobility Classification using GSM and WiFi Traces. IEEE EmNets (2008)
31. Shilton, K., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Participatory Privacy in Urban Sensing. Workshop on Mobile Device and Urban Sensing, IPSN (2008)
32. Waldo, J., Lin, H., Millett, L.: Engaging Privacy and Information Technology in a Digital Age. National Academies Press (2007)
33. Yahoo: FireEagle - Take Location Data to the Web. http://fireeagle.com (2008)
34. MapMyRide: Map Cycling and Biking Routes. http://mapmyride.com (2008)
35. Google: Google Maps API. http://maps.google.com (2008)
36. Kang, J., Welbourne, W., Stewart, B., Borriello, G.: Extracting places from traces of locations. Mobile Computing and Communications Review **9**(3) (2005) 58–68
37. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personal gazetteers: an interactive clustering approach. In: ACM GIS, ACM New York, NY, USA (2004) 266–273
38. Froehlich, J., Krumm, J.: Route Prediction from Trip Observations. Society of Automative Engineers **2193** (2008) 53
39. Reddy, S., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Determining Transportation Modes on Mobile Phones. ISWC (2008)
40. Khuller, S., Moss, A., Naor, J.: The budgeted maximum coverage problem. Information Processing Letters **70**(1) (1999) 39–45
41. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: ACM Knowledge discovery and data mining, ACM Press New York, NY, USA (2007) 420–429
42. Nemhauser, G., Wolsey, L., Fisher, M.: An analysis of approximations for maximizing submodular set functionsI. Mathematical Programming **14**(1) (1978) 265–294
43. Hsu, W., Dutta, D., Helmy, A.: Mining Behavioral Groups in Large Wireless LANs. ACM Mobicom (2007)
44. Osborne, J.: Notes on the use of data transformations. Practical Assessment, Research & Evaluation **8**(6) (2002)
45. Shlens, J.: A tutorial on principal component analysis. Systems Neurobiology Laboratory, University of California at San Diego, December (2005)
46. Deerwester, S., Dumais, S., Furnas, G., Landauer, T., Harshman, R.: Indexing by latent semantic analysis. American Society for Info. Science **41**(6) (1990) 391–407
47. Gong, Y., Liu, X.: Video summarization using singular value decomposition. In: CVPR. Volume 2., IEEE Computer Society; 1999 (2000)
48. Cohen, J.: Privacy, Visibility, Transparency, and Exposure. University of Chicago Law Review **75**(1) (2008)
49. Nissenbaum, H.: Privacy as Contextual Integrity. Wash. Law Review **79**(1) (2004)
50. Zimmer, M.: Privacy and Surveillance in Web 2.0: A study in Contextual Integrity and the Emergence of Netaveillance. Society for Social studies of Science (2007)