The Effectiveness of Centralised Management for Reducing Wasted Effort in Participatory Sensing

Mattias Linnap and Andrew Rice
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
ml421@cam.ac.uk, acr31@cam.ac.uk

***Abstract*—Participatory sensing projects typically rely on time-based polling or volunteer-initiated data collection. We have investigated the potential benefits of *managed sensing*, in which requests for information are targeted at particular volunteers as they visit locations of interest.**

We performed a study comprising of 24 volunteers who were asked to report sightings of bees whilst visiting gardens over a 400m by 400m area. We found that compared to a baseline periodic polling strategy, the centralised management approach achieved 8.6% larger sampling coverage of the visited area with 4 times less wasted participant effort.

We investigated the parameter sensitivity of the results in simulation. We show that the results are statistically significant, and that participant response time to questions is the most influential parameter. Centralised management is more effective than simpler approaches if participants respond within 3 minutes, and equivalent for longer response times. Both the experimental and simulated results are informative for parameter and management choices for future participatory sensing studies.

I. Introduction

Participatory sensing collects data about the physical world with human senses [1]. People can contribute observations on phenomena that are infeasible or expensive to detect with electronic sensors. Most participatory sensing projects to date are *unmanaged*: once recruited to the study, participants have to make their own decisions on when to contribute. Others include various forms of *management*: guidance and reminders during the sensing process (Section II).

Our framework for managed participatory sensing, YouSense, is inspired by *model-based sensing*, which has been previously used to optimise the data value and energy use trade-off in electronic sensor networks [2]. The current knowledge about the world, and the value of additional knowledge from any location are represented as spatial models. The mobile client continuously tracks the participants' location, and prompts them with questions when they are in the optimal location to contribute data (Section III).

We evaluated the effectiveness of three management approaches in an hour-long study of mapping bees with mobile phones in the University of Cambridge Botanical Gardens. 24 participants were split into three groups with different management software. The *timed* group was prompted to contribute every 2 minutes, the *individual* group was prompted at locations where they had not yet contributed, and the *centralised* group was prompted at locations where nobody in their group had yet contributed (Section IV).

The effectiveness of the three management approaches was measured by *coverage* (fraction of area with answers out of visited area) and *wasted effort* (fraction of answers which were not required to satisfy the researchers' goals). In studies with monetary incentives, wasted effort increases the cost of running the study, and in other studies may lead to annoyance and loss of good-will of the participants. The results show that the fully centralised management approach achieves a 8.6% higher coverage and 4 times less wasted effort compared to the timed group, and 4.4% higher coverage and 3.1 times less wasted effort compared to individual management (Figure 1).

It is infeasible to run controlled outdoor experiments for every parameter in participatory sensing. We investigated the parameter sensitivity of the results in simulation, and identified *response time* of the participants (time taken to answer after being prompted with a question) as the main factor in the relative effectiveness of management approaches (Section V).

The main contributions of this paper are:

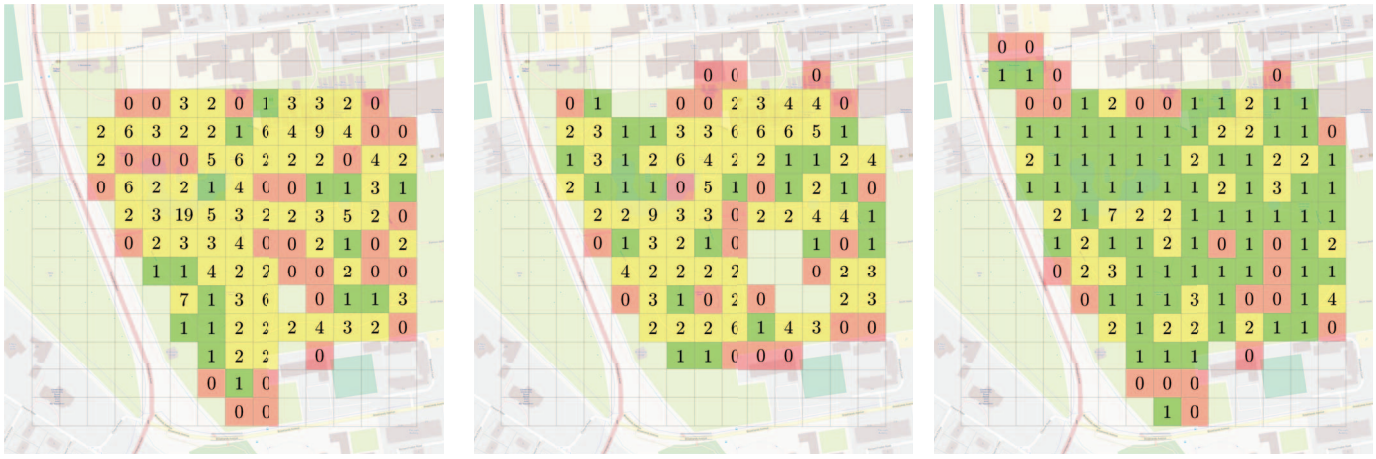
- Evaluation of timed, individual and centralised management for participatory sensing in a controlled outdoor experiment, showing a $4\times$ reduction in wasted effort and 8.6% higher coverage area.
- Sensitivity analysis of the results in simulation, evaluating the effects of participant response times, question intervals, and participant group assignment. Centralised management is more effective than other methods at participant response times below 3 minutes.

II. Related Work

Participatory sensing has been used to collect data on phenomena that are difficult to sense with electronic sensors: flowering of plants [3] and potholes [4] in streets, or subjective metrics such as feelings of happiness [5]. Participatory sensing projects differ in their approaches to *management*: whether and how the participants are instructed and offered guidance or reminders during the sensing process.

Most projects to date are *unmanaged*: once recruited to the study, participants are expected to notice events of interest and contribute observations without further guidance. Some use time-based management to request new contributions from participants. For example, the Mappiness study prompts participants with question at three random times per day [5]. The EmotionSense project uses various electronic sensor triggers (accelerometer movement, sound, phone calls, arriving at home location) to prompt participants to report their emotions in different social and life contexts [6].

978-1-4799-2736-4/14/$31.00 ©2014 IEEE



(a) Timed (b) Individual (c) Centralised

Fig. 1: Grid cells visited by participants in the three control groups in the outdoor experiment. Red cells received no answers, yellow cells were over-sampled, and green cells received the desired number of answers. Centralised management increased the fraction of visited cells with answers, and reduced participant effort wasted on unnecessary answers.

The choice of management triggers introduces bias into the resulting dataset. In social science studies about the participants subjective life experience, the chosen electronic sensor triggers determine the context that participants contribute reports, and can lead to statistically significant differences in observed moods, social interaction and opinions [7].

Volunteered geographical information (VGI) studies survey the physical world rather than the participants themselves [8]. People spend 60-65% of their time at home, and 20-25% at work or school [9], and time-based triggers can result in highly non-uniform geographic distribution of the reports. Correcting sampling bias with electronic sensor triggers is easier in VGI compared to subjective life experience studies, as the primary context (time and location) of the sampled phenomena are available from a mobile phone's GPS. Location-based triggers have been used to sample car fuel consumption [10], human-readable place labels [11], and noise in urban environments [12]. Triggers can be exposed to the participants to encourage movement towards desired locations [13].

Techniques similar to real-time management have been used in selecting participants for participatory sensing studies. Previously collected GPS traces of potential participants can be used to recruit a group that is likely to visit the largest total area [14], and analysis of reports submitted in previous studies can be used to recruit a group that is likely to regularly contribute accurate data [15] [16].

The choice of management approach influences the number of questions participants are asked. In some studies too frequent prompts may merely annoy participants, but studies giving out monetary incentives or limited experimental equipment have a direct maximum budget for effort available [14]. Micro-payment incentives for each question answered increase participant engagement and completion rate of mobile participatory sensing studies [17], and can be even structured as auctions [18]. Centralised management could be used to directly trade off the cost of sensing and the real-time value of an additional answer from a particular location.

III. Centralised Management for Participatory Sensing

Our YouSense framework targets questions to participants who are in the optimal location to contribute new useful data, while minimising the effort wasted on contributions that are not useful to the researchers running the study.

The question targeting is based on *model-based sensing*, which has been previously used to optimise the data value and energy use trade-off in electronic sensor networks [2]. The system makes a decision to notify a participant based on three models:

- The *world model* captures the current knowledge: participants' answers plus any prior beliefs or external data sources.
- The *value model* captures the researchers' goals: given the current world model, how valuable is an additional answer from a location.
- The *cost model* sets a limit on notifying participants. In studies with monetary incentives, it can be based on the rewards and budget available, in others it can be a fixed number per day or hour.

The models in YouSense can be configured to match the prior expectations and goals of the study organisers. In this study, the world model was represented as a 14×14 cell grid of 27.8×28.2 meters each (Figure 1), and the value model was set to value one answer from each cell. The cost model was set to allow notifying participants with questions no less than 2 minutes apart. Other studies may use different world model representations (for example a probability field over space) or different value models (for example answers from X unique participants, or Y answers that agree) based on the prior beliefs and expectations of the study organisers. YouSense uses a continuous location tracker to match the participants' current location to the spatial models.

The study compared three approaches to management: *timed*, *individual* and *centralised*.

With the *timed* approach, participants are prompted with a question every 2 minutes. Time-based notifications are the base case from past studies such as Mappniess [5].

With *individual* management, the system tracks the current location and previous answers from a single participant. They are prompted with a question when entering a grid cell with no previous answers (but not more than once every 2 minutes).

With *centralised* management, the system tracks the location and previous answers from all participants. The world and value models are synchronised between participants' mobile phones once per minute. They are prompted with a question when entering a grid cell where no participant has answered previously (but not more than once every 2 minutes).

The individual and centralised management approaches use continuous location tracking, and the centralised approach synchronises the models between participants over a network. These increase energy use on participants' phones, and require the researchers to maintain a management server. The more complex management methods are worthwhile only if they result in more efficient use of the participants' effort than simple time-based questions.

IV. Evaluation Methods

A. Outdoor Experiment

The effectiveness of the three management approaches was evaluated in an hour-long mobile participatory sensing study in the University of Cambridge Botanical Gardens. 24 participants walked the gardens for 43 to 73 minutes, and were asked to the question "Do you see any bees?".

The participants were split evenly between the three approaches. They were unaware of the groups, or how and when the system prompts them with questions. The movement paths chosen by the participant influence the results. For an ideal controlled study of management approaches, members of each group would follow the exact same path in teams of three, and respond to questions with uniform speed. As a practical approximation, the participants were asked to walk naturally and explore the gardens in groups of no more than 2-3 people, and participants who were likely to walk together were assigned into different control groups.

B. Parameter Sensitivity Analysis in Simulation

The effectiveness of the management approaches is influenced by many parameters, including the participant response time (time taken to answer after being prompted with a question), the interval between questions, the number of participants, and the assignment of participants to management approaches. It is infeasible to run a controlled outdoor experiment for every combination. Instead, parameter sensitivity was evaluated with discrete event simulation.

The simulator implements the same models and management approaches as the outdoor study (Section III). Participant movement is simulated based on the 1Hz GPS traces collected during the outdoor experiment, and their response time to questions is modelled using a fixed delay. The simulator includes the 1-minute interval for world and value model synchronisation between participants.

TABLE I: Summary of the outdoor results. All three groups visited a similar area in their walk in the gardens. Centralised management increased coverage area (cells with answers normalised to visited cells) by 8.6% compared to the unmanaged timed group, and reduced the effort wasted on answers from duplicate locations by 4 times. The per-participant values are given as mean $\pm$ standard deviation.

	Timed	Individual	Centralised
Total cells visited by the group	104	97	114
Mean cells visited per participant	58.3 ± 10.4	55.5 ± 12.0	61.1 ± 16.1
Covered area (cells with answers)	75 (72.1%)	74 (76.3%)	92 (80.7%)
Missed area (0 answers)	29 (27.9%)	23 (23.7%)	22 (19.3%)
Over-sampled area (2+ answers)	59 (56.7%)	51 (52.6%)	25 (21.9%)
Total answers by the group	224	188	129
Mean answers per participant	28.0 ± 2.2	23.5 ± 4.1	16.1 ± 4.2
Useful answers	75 (33.5%)	74 (39.4%)	92 (71.3%)
Wasted answers	149 (66.5%)	114 (60.6%)	37 (28.7%)
Wasted due to movement	34 (15.2%)	32 (17.0%)	33 (25.6%)
Wasted due to lack of management	115 (51.3%)	82 (43.6%)	4 (3.1%)

C. Evaluation metrics

The effectiveness of management approaches is evaluated on two main metrics:

- *Coverage area*: the fraction of visited cells with answers. Coverage is normalised to visited area to negate the bias of different movement paths across groups.
- *Wasted effort*: the fraction of answers which were not required to satisfy the value model. In this study, all answers after the first from a cell were wasted according to the chosen value model.

V. Results

Table I summarises the results of the outdoor study. The fully centralised management approach achieves a 8.6% higher sampling coverage of the visited area, and 4 times less wasted effort compared to the timed group (no location tracking), and 4.4% higher coverage and 3.1 times less wasted effort compared to individual approach.

A. Coverage Area

Centralised management increases the coverage area. With centralised management, participants contributed answers from 80.7% of the visited cells, compared to 76.3% for individual and 72.1% for timed management (Table I). As the participants in each group walked a different path (and therefore had different number of potential answer locations), these numbers are normalised to the number of cells visited by the group. Figure 1 shows the cells visited by participants of each group, along with the number of answers from each cell.

The modest benefit of centralised management for increasing coverage area is due to the chosen cost model: questions were asked up to every 2 minutes. Participants in all groups were prompted with questions frequently enough to still have an opportunity to answer from most of the visited cells.

Simulation results show that for all management approaches, there is a linear relationship between the question

interval and coverage area. More questions lead to higher coverage, but centralised management becomes relatively more effective at longer intervals. With 1 minute intervals, the coverage achieved by all methods is equal, but with 5 minute intervals centralised management covers 1.2 times more area.

B. Wasted Effort

Besides maximising coverage area, a goal of the centralised management is to minimise participant effort. Participant effort is classified as *useful* or *wasted* based on the value model chosen for the study. In this study, the value model considers the first answer from each cell as useful, and any following answers as wasted effort. In other studies, the YouSense framework can be configured with different value models. Studies offering monetary incentives can put a direct monetary value on a data contribution from a particular location, and measure the wasted effort directly as economic return on investment.

Table I shows the three management approaches differ greatly in the number of questions asked from the participants, but the number of *useful* answers is similar for all groups. The main benefit of management is preventing participant effort wasted on answers from already sampled cells.

With centralised management, 28.7% of answers were from a cell where data was already available and therefore considered wasted effort, compared to 66.5% for timed questions. Individual management had 60.6% wasted effort. In theory, centralised management should eliminate answers from locations where the value of new contributions is zero. The analysis of participant behaviour in Section V-D shows that this is due to participants moving between being notified with a question and submitting an answer.

Figure 2 shows the timeline of useful and wasted answers for all three groups during the study. Over time, the centralised approach prompts participants with fewer questions, as they re-visit locations where themselves or others have already contributed data. The rate of useful answers contributed is expected to decrease over time as the experimental area becomes saturated with answers, but this threshold was not reached in the hour-long study.

Wasted effort also depends on the interval between questions. Intervals ranging from 0 to 5 minutes were evaluated in simulation. With more frequent questions, the timed approach is more likely to ask multiple times in the same location, directly increasing wasted effort. The individual management approach prevents the same participant from answering twice from a location, but with more frequent questions, different participants are still more likely to answer from locations where their movement paths overlap (Figure 3).

C. Effects of participant assignment

In the outdoor study, the 24 participants were assigned into three groups, 8 for each management approach. Each participant walked a different path in the gardens, and the groups visited a different total number of cells. The effects of participant assignment were evaluated in simulation with 333 randomised group assignments. The simulation results confirm that the increase in coverage area and decrease in wasted effort is due to better management, not the particular participant groups chosen in the outdoor experiment.

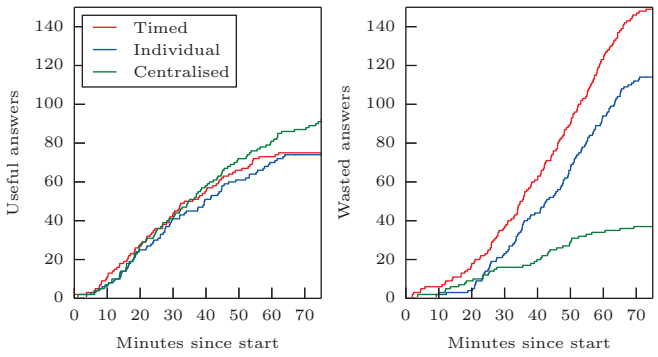


Fig. 2: Cumulative useful and wasted answers over time in the outdoor experiment. Useful answers were contributed and the area coverage increased at the same speed for all three groups. Centralised management reduces effort wasted on duplicate answers by suppressing question prompts when the participant is at a location with previous answers.

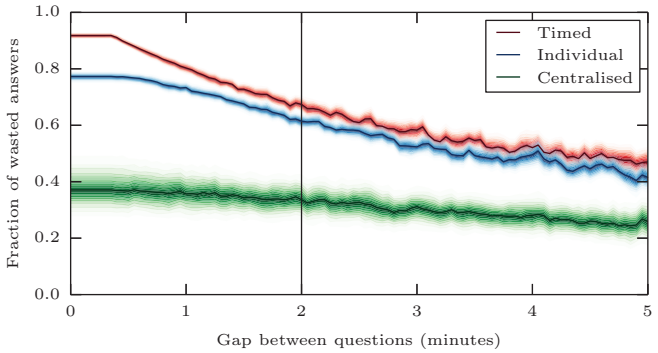


Fig. 3: Simulation shows that centralised management is more effective at small question intervals (more frequent questions). The black vertical line marks the 2-minute question interval used in the outdoor experiment. Intervals under 20.1 seconds do not increase the number of wasted nor useful answers, as that was the participant response time model chosen for the simulation. The fuzzy lines are density histograms of the results over 40 000 random participant group assignments.

The results are statistically significant: 1-way 3-group ANOVA on 333 samples per group has $F = 673.3$, $p < 0.001$ for coverage area and $F = 12377.4$, $p < 0.001$ for the wasted effort metrics. The Tukey HSD test also confirms with $p < 0.001$ that the differences between each pair of the result metric distributions of the three management approaches are significant, and they are extremely unlikely to be drawn from the same distribution. The high confidence in the benefits of management is due to the large sample size in simulation, which would be infeasible to achieve in outdoor experiments.

D. Effects of participant response time

Answers arrive from already sampled cells for two reasons: the participants were prompted in with a question in a cell with previous answers (insufficient management), or participants moved to a cell with previous answer after being

prompted. The YouSense framework for participatory sensing management can only target *questions* to locations where new *answers* would be most valuable, but it cannot compel the participants to answer from there promptly.

We have identified participant *response time* (time taken to answer after being prompted with a question) as the most influential factor in the effectiveness of centralised management. All types of management had $32-34$ answers from cells with previous answers due to participants moving between the notification and submitting an answer (Table I). However, centralised management almost eliminates effort wasted due to prompts in cells with previous answers. The 4 questions in cells with existing answers for centralised management are a result of synchronisation delays of up to 1 minute between participants' mobile phones.

Table II shows the time taken and distance moved between YouSense prompting a participant with a question, the participant opening the answer screen, and the participant submitting an answer. Figure 4 shows the histogram of response times. These distributions remained consistent throughout the study.

Response time varied across groups. The timed group received most questions and answered with a mean of 16.8 seconds, the individual group 20.6 and centralised group in 25.3 seconds. The majority of the time taken to answer was spent on reacting to a notification and opening the app. After opening the app, all groups answered with a mean of 4.5 seconds. Post-study interviews with the participants revealed that the regularity of questions for the timed group encouraged faster responses. Some participants in the timed group kept their phone in the hand throughout the study, as questions arrived at a regular 2 minute interval. Participants in the individually managed and centralised groups tended to only take the phone out of their pocket when hearing the notification.

We investigated the effect of response time on management effectiveness in simulation (Figure 5). With response times over 3 minutes, both the individual and centralised methods lose their ability to target *answers* to valuable locations in our study, as the participants move too far between the question and answer. These results indicate that knowing the expected response time is important for deciding on the management method for participatory sensing studies: if the delays are large, a simple timed approach is sufficient.

E. Where are the bees?

The participatory sensing question used in the study was "do you see any bees", with options "No", "Yes, one" and "Yes, many". Although only the location and timeliness of the answers, not the answer itself matters for this study of effectiveness of managed participatory sensing, the results match the prior expectations of the pollination scientists in the Botanical Gardens. Figure 6 shows the map of the answers, with most bees reported near the flower beds considered attractive to bees by the botanists.

VI. Discussion and Conclusion

A model-based approach to participatory sensing enables a study organiser to provide a quantitative model for the

TABLE II: Participant response times and distance moved between receiving a question and contributing an answer in the outdoor experiment. The results were not normally distributed, so standard deviations are not informative (Figure 4). For reference, the grid cell size was 28 meters, and the average accuracy reported by GPS 5.9 ± 5.1 meters.

	Timed	Individual	Centralised	All
Mean response time (s)	16.8	20.6	25.3	20.1
Median	12.8	15.6	16.7	14.6
95th percentile	36.4	37.8	93.9	45.1
Mean movement distance (m)	7.1	10.0	10.9	9.0
Median	4.8	5.3	5.0	5.0
95th percentile	19.6	28.9	40.4	26.6

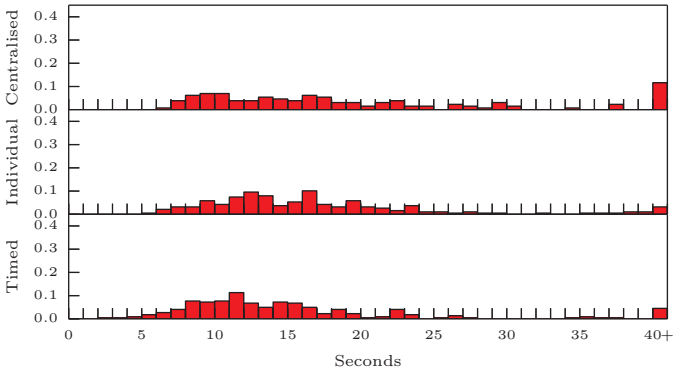


Fig. 4: Distributions of response times in the outdoor experiment. The timed group was the fastest to react to a notification and submit an answer. The vertical axis is normalised to the total number of answers contributed by the group. The final column is the sum of all delays longer than 40 seconds.

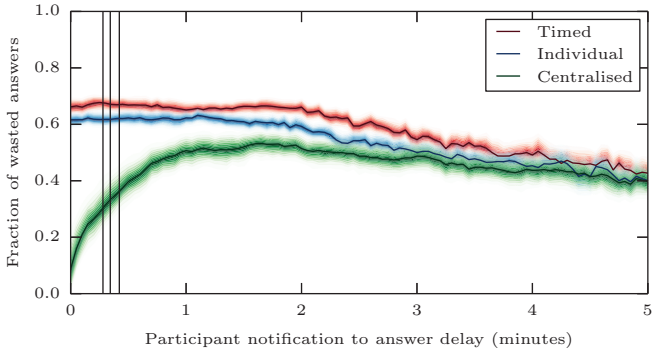


Fig. 5: Simulation shows that centralised management is more effective with fast response times. If participants take over 3 minutes from being notified with a question to contribute an answer, it is no more effective than the timed approach. The black vertical lines marks the mean response times observed for each group in the outdoor experiment. The fuzzy lines are density histograms of the results over 40 000 random participant group assignments.

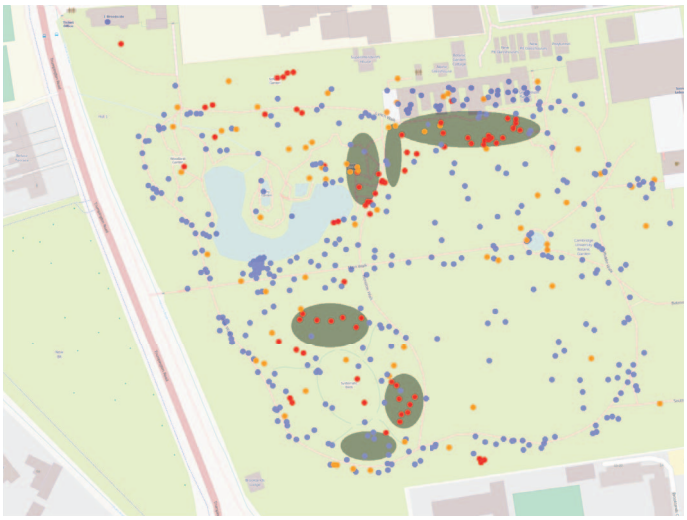


Fig. 6: Answers from all participants in the outdoor experiment. Red is "many bees", orange "one bee" and blue "no bees". The dark ellipses mark areas with flower beds that pollination scientists expected to be most attractive to bees.

current knowledge about the *world*, the *value* of additional data contributions, and the *cost* of human effort. The YouSense centralised management framework uses these models together with location tracking and real-time data synchronisation to ensure that the participants are prompted to contribute only when they are in the optimal situation to do so.

In the outdoor experiment, centralised management reduced the participant effort wasted on answers from duplicate locations by 4 times. Analysis of the results in simulation indicated that the most important factor in the effectiveness of management is the participants' response time to questions, which in turn relates to movement between being prompted with a question and submitting an answer. With response times under 3 minutes, centralised management is more effective, and over 3 minutes equivalent to the other approaches.

Centralised management increases the energy and mobile network use on the participants phones, and requires the researchers to maintain a management server. We hope the results in this paper allow researchers planning studies to make an informed decision on whether to implement centralised management, individual location-based management, or a simpler time-based notification approach.

We believe there are further participant behaviour patterns and study parameters worth investigating for their effect on centralised management: transport modality, movement density and overlaps between participants, size and duration of the study, and alternative representations of world, value and cost models. Further work into value and cost models are particularly interesting for studies involving monetary incentives. In those studies, the framework can directly compute the return-on-investment on prompting a participant, or run an auction for human contributions. To ensure that centralised management is capable of targeting questions to valuable locations, the system could be augmented with movement path prediction [19], or offer rewards based on time taken to answer by the participant.

Acknowledgement

The authors thank Oliver Chick, Daniel Thomas and Daniel Wagner for reviewing drafts of this paper, Ramsey Faragher for statistical advice, and Beverley Glover for advice on running experiments in the Botanical Gardens.

References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. Srivastava, "Participatory sensing," in *World Sensor Web Workshop*, Boulder, Colorado, USA, 2006.

[2] A. Deshpande, C. Guestrin, and S. Madden, "Using probabilistic models for data management in acquisitional environments," in *Conference on Innovative Data Systems Research*, Asilomar, California, USA, 2005.

[3] K. Meymaris, S. Henderson, P. Alaback, and K. Havens, "Project BudBurst: Citizen Science for All Seasons," in *American Geophysical Union, Fall Meeting*, San Francisco, California, USA, 2008.

[4] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *MobiSys*, New York, New York, USA, 2008.

[5] G. MacKerron, "Happiness economics from 35000 feet," *Journal of Economic Surveys*, vol. 26, no. 4, pp. 705–735, 2012.

[6] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "EmotionSense : A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research," in *UbiComp*, Copenhagen, Denmark, 2010.

[7] N. Lathia, K. K. Rachuri, C. Mascolo, and P. J. Rentfrow, "Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods," in *UbiComp*, Zurich, Switzerland, 2013.

[8] M. Goodchild, "Citizens as sensors: the world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[9] P. Baumann, W. Kleiminger, and S. Santini, "The Influence of Temporal and Spatial Features on the Performance of Next-place Prediction Algorithms," in *UbiComp*, Zurich, Switzerland, 2013.

[10] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: A participatory sensing fuel-efficient maps application," in *MobiSys*, San Francisco, California, USA, 2010.

[11] D. Kim, K. Han, and D. Estrin, "Employing user feedback for semantic location services," in *UbiComp*, Beijing, China, 2011.

[12] E. D'Hondt, M. Stevens, and A. Jacobs, "Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring," *Pervasive and Mobile Computing*, vol. 9, no. 5, pp. 681–694, 10 2012.

[13] J. Rula and F. E. Bustamante, "Crowd (Soft) Control," in *HotMobile*, San Diego, California, USA, 2012.

[14] S. Reddy, K. Shilton, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using context annotated mobility profiles to recruit data collectors in participatory sensing," in *Location and Context Awareness*, Tokyo, Japan, 2009.

[15] S. Reddy, K. Shilton, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Evaluating participation and performance in participatory sensing," in *UrbanSense*, Raleigh, North Carolina, USA, 2008.

[16] S. Reddy, V. Samanta, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "MobiSense - Mobile Network Services for Coordinated Participatory Sensing," in *Autonomous Decentralized Systems*, Athens, Greece, 2009.

[17] M. Musthag, A. Raij, D. Ganesan, S. Kumar, and S. Shiffman, "Exploring micro-incentive strategies for participant compensation in high-burden studies," in *UbiComp*, Beijing, China, 2011.

[18] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing," in *MobiCom*, Istanbul, Turkey, 2012.

[19] J. Krumm and E. Horvitz, "Predestination: Inferring destinations from partial trajectories," in *UbiComp*, Orange County, California, USA, 2006.