Utility-Aware Participant Selection with Budget Constraints for Mobile Crowd Sensing

Shanila Azhar[1], Shan Chang[1(✉)], Ye Liu[1], Yuting Tao[1], Guohua Liu[1], and Donghong Sun[2]

[1] Computer Science and Technology, Donghua University, Shanghai, China
{415029, yeliu, taoyuting}@mail.dhu.edu.cn, {changshan, ghliu}@dhu.edu.cn

[2] Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing, China
sundonghong@tsinghua.edu.cn

Abstract. Mobile Crowd Sensing is an emerging paradigm, which engages ordinary mobile device users to efficiently collect data and share sensed information using mobile applications. The data collection of participants consumes computing, storage and communication resources; thus, it is necessary to give rewards to users who contribute their private data for sensing tasks. Furthermore, since the budget of the sensing task is limited, the Service Provider (SP) needs to select a set of participants such that the total utility of their sensing data can be maximized, and their bid price for sensing data can be satisfied without exceeding the total budget. In this paper, firstly, we claim that the total data utility of a set of participants within a certain area should be calculated according to the data quality of each participant and the location coverage of the sensing data. Secondly, a participant selection scheme has been proposed, which determines a set of participants with maximum total data utility under the budget constraint, and shows that it is a Quadratic Integer Programming problem. Simulations have been conducted to solve the selection problem. The Simulation results demonstrate the effectiveness of the proposed scheme.

Keywords: Mobile Crowd Sensing · Utility · Budget · Data quality · Incentive

1 Introduction

In recent years, there is an enormous increase in the usage of handheld smart devices (i.e. smartphones, PDAs, tablets, smartwatches, and music headsets, etc.). According to Ericsson, the total number of worldwide smartphone subscriptions has reached 7.9 billion in Q2 2019, and it grows around 3% year on year [1]. In addition to the cellular communication standards (3G/4G), these mobile devices support numerous communication technologies Wi-Fi 802.11a/b/g/n, NFC and Bluetooth). They also come with high-performance processors and several compact size embedded sensors (e.g. camera, microphone, GPS, barometer, ambient light, gyroscope, accelerometer, magnetometer, and digital compass, etc.) to gather sensing data (e.g. traffic situation, temperature, and

X. Chu et al. (Eds.): QShine 2019, LNICST 300, pp. 38–49, 2020.
https://doi.org/10.1007/978-3-030-38819-5_3

noise level, etc.) from surrounding [2]. The advancement and pervasive usage of such mobile devices give an emergence of a new sensing paradigm called Mobile Crowd Sensing (MCS). MCS encourages ordinary users to collect sensing data from surroundings by simply using their smartphones, and to share that data for facilitating a large number of new applications; e.g. traffic management, monitoring [3], health monitoring [4], public safety [5], and even psychological survey questionnaires [6]. In typical MCS scenarios, after a server publishes a sensing task, the interested mobile users register for a sensing task, and then the server will select a set of users to join in the task according to a pre-determined selection scheme. The most commonly used metric for selecting participants is the data quality, which is however not enough. It is desired that those data are sensed over the whole target sensing area instead of from the same or neighboring locations, to avoid bias results. In this paper, we propose a novel participant selection scheme named Utility-aware participant selection (UPS), which considers the following key factors.

First, during the participation activities of users, mobile devices consume resources, i.e., battery, computing power, giving identity or location, etc. Many users may not probably like to contribute in the sensing activity due to the time and cost of resources utilized. Therefore, an incentive mechanism is essential to apprehend active and reliable MCS sensing and to encourage enough number of users to report their precise sensed data [7, 8], and thus guarantee high-quality sensed data. Considering that the server budget for a certain sensing task is limited, it desires a mechanism that maximizes the utility of sensing data under the constraint of a budget.

Second, the quality of data is an essential factor in MCS. Low-quality data might affect the performance of SP, resulting distrustful MCS system. We use the quality of data in measuring the utility of the MCS system. If a mobile device user holds high-quality data at a reasonable price, the user has more chances to be selected as a participant of a sensing task.

Third, since most sensing tasks are location-dependent, locations from where the sensing data is collected also need to be considered. Thus, it is important for SP to monitor and gather data from participants present at different locations in a grid to attain data reliability, as selecting the participants nearby each other or near the center of the grid can lead to an inefficient MCS system. Therefore, we also measure the distances between the users to maximize the location coverage when calculating the total data utility, as the farther the distance of participants from each other, the wider will be the area covered in a grid.

We define the total sensing data utility of a participant set, the value of which relates to both data quality of sensing data, and the distances between participants. If the participants in the set are far away from each other and hold high-quality data as well, then their total data utility should be high. We formulate the problem of selecting a participant set as maximizing the total utility under the constraint that the total data price required does not exceed the budget. The main contributions of this paper include:

- We consider the importance of quality and location coverage of sensing data of participants and introduce a new definition of data utility.

- We considered the data utility, data price of mobile device users, and budget of sensing tasks into account and propose a novel utility-aware participant selection scheme, to maximize the total data utility under the constraint of a budget. We prove that the selection problem can be formulated by a Quadratic Integer Programming (QIP) problem.
- We conduct simulations to show the efficacy of the proposed scheme. Simulation results depict that our scheme can achieve good performance in terms of the maximized data utility within the satisfactory budget cost.

The remainder of this paper is organized as: In Sect. 2 we discussed some related work. In Sect. 3, we present our proposed MCS system model. In Sect. 4 we give the problem formulation. In Sect. 5, we describe the performance evaluation and simulation setup and results. Finally, Sect. 6 concludes our work.

2 Related Work

Many of the participants' recruitment schemes in the MCS system have been investigated. The selection scheme plays an extensively important role in the success of any MCS system. The high density of smartphone users in an area, allows the MCS system to select a set of participants and result in better performance. To achieve high-quality data, a simple solution is to select as many participants as possible [9].

Numerous systems and experimental studies show the experimental study on the MCS coverage area and the general structure of participant recruitment [10]. *Chon et al.* [11] has performed an experimental study on the scaling and coverage properties and show likely results that MCS can provide relatively high coverage levels especially given area with large size. Besides, many theoretical studies on different sensing task allocation and participant selection problems have some tradeoffs of task completed, sensing cost, efficiency, and user incentive [12, 13].

Other researchers focused on the incentive mechanisms and auction methods for rewards and biding, respectively, to motivate and ensure the contribution of participants, also it encourages them to sense and send good quality data. Sun *et al.* [14] designed a reputation aware incentive mechanism to get the maximum weighted social welfare of the system and guarantee the individual rationality and truthfulness. Lee et al. [15] proposed an iterative reverse auction-based incentive mechanism, where mobile users sell their sensed data to the service provider with their demanded bid price. Chen et al. [16] used a double auction method in incentive mechanism for smartphone users as well as for sensing multiple tasks. Jaimes *et al.* [17] used the reverse auction method and considered the smartphone users covering location information to select the participants, and considered the budget limits of a service provider. Feng *et al.* [18] also used reverse auction in incentive mechanism design and considered the location information.

Most of the participants' recruitment schemes have the same purposes to develop a cost-effective selection scheme with high user density. These schemes require auction methods to assign or negotiate incentives with smartphone participants [19]. However,

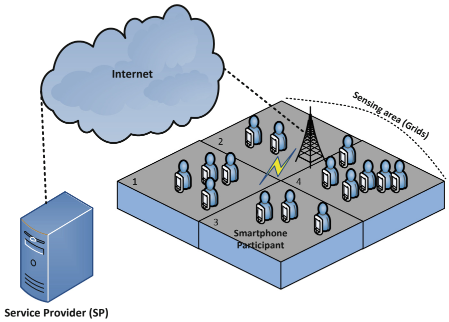

Fig. 1. MCS system.

until now the quality of utility and the sensing data from recruiting participants is still neglected. We considered these challenging parameters in our work.

3 System Model

We consider the MCS system as shown in Fig. 1. The service provider SP intends to collect the data in an area through the contribution of the set of mobile participants. SP announces the sensing task through mobile Apps. Mobile device users, who choose to conduct sensing task download and install these Apps and participate in the sensing process. The entire sensing area is divided into M grids denoted by set $G = \{g_1, g_2 .., g_M\}$. In a grid $g_k(1 \leq k \leq M)$, the number of users interested to participate in the sensing task is n_k, and a mobile device user $P_i(1 \leq i \leq n_k)$ is associated with numerous attributes.

– Bid price b_i: which is the payment/reward that participant P_i hopes to receive from SP for conducting a sensing task.
– Data quality q_i: which reflects the accuracy and truthfulness of sensed data provided by P_i at a certain location. SP assigns sensing tasks to users based on their report of quality of data. The data quality affects the performance of MCS systems, low-quality data can degrade the efficiency of SP and the reliability of the MCS system.
– Location l_i: which is the precise location where P_i conducts sensing tasks.

We give Fig. 2 to illustrate our MCS model. It depicts the main activities performed during the entire sensing process. The MCS model comprises two major modules; Service provider (SP) and Mobile Participants (P_i). Following steps illustrate the step by step MCS sensing process labeled above in our sensing model (assuming all interested participants already have necessary Apps on their devices);

- The SP advertises a sensing task in a region and intends to recruit a set of participants within the region.
- Mobile device users interested in sensing tasks within the vicinity send a registration request to SP and announces SP about their locations, data qualities and bid price.

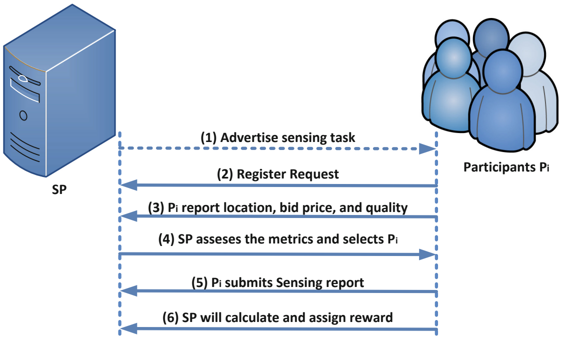

Fig. 2. System model for the MCS system.

- SP selects the subset of participants using a global algorithm satisfying the constraints of bid price and budget.
- Winning set of participants submits the sensing report.
- SP assesses the submitted data reports, and assign rewards to the participants.

4 Participant Selection

To guarantee high-quality MCS applications, the SP desires to collect sensing data with high utility. Since the budget of a sensing task is limited, it is impossible to recruit all participants who are interested in the task. Thus, the primary objective of SP is to attain maximum data utility within a limit of the total Budget. We also argue that the utility of data is not only related to the quality of sensing data but also the location where the data are sensed. Given a set of participants, it is desired that their data are high-quality and cover the overall area that needs to be monitored. Thus, we introduce a novel definition of the utility of data from a participant set.

Definition 1. The Utility of Data of a Participant Set W can be measured as:

$$U_W = \sum\nolimits_{P_i \in W} q_i * \sum_{\substack{i \neq j \\ P_i, P_j \in W}} d_{ij}, \text{ where } d_{ij} = \left\| \left(l_i - l_j \right)^2 \right\| \tag{1}$$

In the definition, we consider q_i to represent the quality of data submitted by the participants and d_{ij} denotes the Euclidean distance between the participant P_i and P_j, because we intend to recruit those participants that have high-quality data and also we aim to select those participants that are not located in the same place. For instance, if two users with high data quality are selected, but both are near to the same location, it will not assure high-quality data from the entire grid. Instead, we want participants to be distributed all over the grid to maximize the distance between them and achieve data reports from locations at farther distances in a grid. We give the following example to better explain Data Utility:

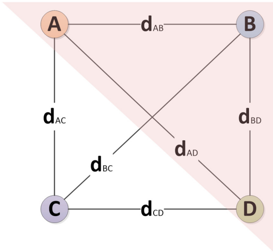

Fig. 3. Participants $\{A, B, C, D\}$ w.r.t their distances from each other.

Let's consider, SP decides to select participants A, B, and D as shown in Fig. 3. The Data Utility of the participant set $\{A, B, D\}$ achieved for this selection can be formulated by substituting it in Eq. (1) and expressed as;

$$\begin{aligned} U_{ABD} &= q_A * (d_{AB} + d_{AD}) + q_B * (d_{BA} + d_{BD}) + q_D * (d_{DA} + d_{DB}) \\ &= d_{AB} * (q_A + q_B) + d_{BD} * (q_B + q_D) + d_{AD} * (q_A + q_D) \end{aligned}$$

Thus, we formulate the participant selection problem as follows:

Definition 2. For a set of participants W (containing n participants), given the bid price b_i and data quality q_i of each participant P_i, and the constraints of budget C, the participant selection problem can be formulated as the following optimization problem;

$$W' = arg_{W' \in \Pi} max \sum\nolimits_{P_i \in W'} U$$

$$\sum\nolimits_{P_i \in W'} b_i < C$$

Where, W' is a winning set of participants to perform a sensing task, and Π is the possible set of participants to perform the sensing task.

We find that the above problem can be rewritten as a Quadratic Integer Programming (QIP) problem. The objective of the QIP is to find an n-dimensional vector x, that will

$$maximize \frac{1}{2} x^T Hx$$

$$subject\ to\ Bx \leq C,$$

$$x \in \{0, 1\}$$

Each dimension of x is an indicator to show whether the corresponding participant in W is selected as a participant for the sensing task. If $x_i = 1$, it means that P_i is selected, otherwise, P_i is unselected. x^T denotes the vector transpose of x. H is an $n \times n$-dimensional symmetric matrix, in which $h_{i,j}$ (the element on the i-th row and j-th

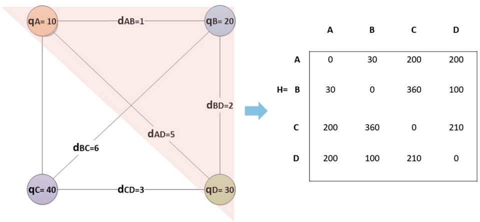

Fig. 4. The quality of each participant in W, and the distances among them, and the corresponding matrix H.

column) can be calculated by $h_{i,j} = d_{ij} * (q_i + q_j)$, i.e., the sum of data quality of participant P_i and P_j times the distance between them. A is an n-dimensional vector to represent the bid price of each participant, i.e., $B = [b_1, b_2, \ldots, b_n]$, and C is the total budget. Consider the above example again. The set of participants is $W = \{A, B, C, D\}$, the quality of each participant, and the distances between any two participants are shown in Fig. 4. Then, we have

$$H = \begin{bmatrix} 0 & 30 & 200 & 200 \\ 30 & 0 & 360 & 100 \\ 200 & 360 & 0 & 210 \\ 200 & 100 & 210 & 0 \end{bmatrix}$$

Assume that SP selects $W' = \{A, B, D\}$, i.e., $x^T = [1 \quad 1 \quad 0 \quad 1]$, then we can compute

$$U_{W'} = \frac{1}{2} x^T H x = \frac{1}{2} [(30 + 200) + (30 + 100) + (0) + (200 + 100)] = 330$$

The budget C is 0.8, and $B = [0.1, 0.2, 0.3, 0.4]$, thus we can verify that

$$Bx = [0.1 \quad 0.2 \quad 0.3 \quad 0.4] * \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = 0.6 < 0.8$$

It means the total SP expenses do not exceed the budget, i.e., the budget constraint is satisfied.

It is known that the QIP problem is NP-hard [20], which means that the optimal set of winning participants and the maximum data utility cannot be solved in polynomial time. Thus, we adopt a branch and bound algorithm proposed by Körner [21], to find an approximate optimal solution.

5 Performance Evaluation

In this section, to evaluate the performance of our scheme, extensive simulations have been conducted in terms of the optimum data utility with the budget constraints of SP. First, we compare the UPS scheme with the random selection method to show the superiority of our QIP solution; later, we analyze the performance trend for the varying number of mobile device users and the varying number of grids, respectively.

5.1 Simulation Setup

We consider the MCS system in which the whole sensing area is a 10 km × 10 km square. This square was further divided into several square grids. In the area, we randomly generate 100 mobile device users, i.e., they are randomly located in grids, and the distances between them are measured. The data quality of a user and the bid price of data both are real values between 0 and 1, and the budget of a sensing task is set as a real value between 10 and 50. We generate bid price and quality of sensing data of each participant using uniform distribution. Table 1 summarizes the settings of parameters in our simulations.

Table 1. Simulation parameters

Simulation area (km)	10 × 10
Number of Grids	4 to 100
Total no. of randomly generated users	10 to 100
Total budget C	$10 < C < 50$
Data quality q_j	$0 < q_j < 1$
Bid price b_j	$0 < b_j < 1$

5.2 Performance Comparison

We compare the UPS scheme with the following random selection scheme.

- **Random Selection Scheme:** Initially, we randomly select one mobile user and check the claimed bid price, if the bid price is smaller than the budget; we again randomly select another user and add it into the target group of participants, and compute the total bid price to see if the total bid price is still smaller than budget. We do it several times until the total bid price is larger than the budget when the bid price exceeds the limit the selection procedure is ended.

For the UPS we randomly generated users and consider the bid price, location, and quality of sensing data submitted by each participant. We run the simulation 10 round times and calculate the total utility achieved by UPS and Random selection. Figure 5 shows the utility curve of both the schemes for 10 rounds. The x-axis and y-axis represent the number of rounds and the total utility, respectively. It can be seen that the

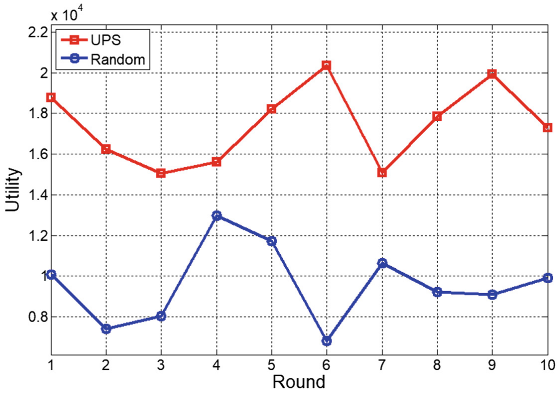

Fig. 5. Achieved utilities of UPS and random selection in 10 rounds.

UPS achieves much larger utility values in all rounds comparing with the random selection scheme.

5.3 Impact on Utility Due to Number of Participants

Figure 6 shows the impact of an increasing number of mobile users on the utility. In this set of simulations, we consider the effect of a larger number of users on the utility in association with the Budget range. The x-axis and y-axis represent the number of participants and the utility, respectively. Where the number of mobile users increases in an area from 10 to 100 at a step of 10. We determine the utility for various ranges of the budget value, from 10 to 50. For the small number of users under a given budget, most of the users can be selected into the target group. As the number of users increases, for a lower budget more and more participants exclude from the target grid (as Fig. 6 shows for B = 10). However, for the higher budget value, it can support many more

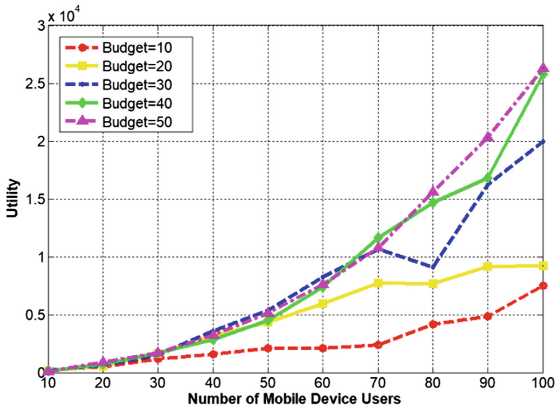

Fig. 6. Utilities vs. numbers of participants and budgets.

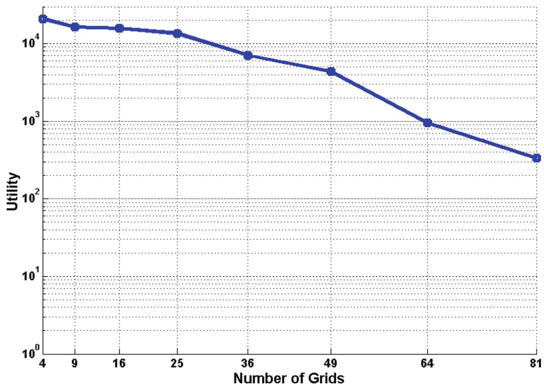

Fig. 7. Utilities vs. numbers of grids.

participants to be selected in the target group. We see at the value of budget 50, the utility increase significantly, even with 100 participants the majority of the users can still be selected into the group.

5.4 Impact on Utility Due to Number of Grids

In this set of simulations, we modify the number of grids in the simulation area. We divide the longitude and latitude directions of the area into 2 to 10 segments, respectively. Consequently, the corresponding numbers of grids increase as a squared factor from 2^2 to 10^2. We set the total budget in the whole area as 30, and evenly distribute the total budget to each grid. Figure 7 shows the impact of the numbers of grids on the utility. We can see that with an increase in the number of grids, the total utility of all grids decreases. It is because dividing the area into grids implies that each grid has its own local optimal set of participants. When we combine those sets into a whole participant set and calculate the total utility, the utility will be smaller than that of a global participant set obtained by solving a global QIP problem directly. Furthermore, more grids mean more participants might be excluded from participant sets, as more constraints of the budget should be satisfied. It should be noticed that a large grid means more users located in it. Thus, the computation complexity will be relatively high. To get a better trade-off between the computational complexity and the utility, a reasonable number of grids are 5^2.

6 Conclusion

Taking optimization of the data utility into consideration, a novel scheme for participant selection has been presented to collect the well-measured sensing data from the set of participants. The selection scheme significantly improved the participant recruitment process by assessing the data quality of the contributed data and their sensing location. Besides, the total budget of the MCS was stabilized by the bid price constraint. The evaluation of scheme performance was carried out by comparing the UPS scheme with a random selection scheme, determining the impact of the number of participants on the utility, and the impact of the grids on the utility. Furthermore, the presented results significantly validated the effectiveness of the proposed scheme in terms of utility, quality of data, and budget restraint.

In the end, this work opens some research directions for the future, we can customize it with other auction methods, or incentive mechanisms, etc. Also, the integration of reputation assignments in the MCS system can ensure and encourage a more refined selection of participants.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China (Grant No. 61972081, 61672151, 61772340, 61420106010), National Key Research and Development Project (Grant No. 2016QY12Z2103-2), Shanghai Rising-Star Program (Grant No.17QA1400100), National Key R&D Program of China (Grant No. 2018YFC1900700), Shanghai Municipal Natural Science Foundation (Grant No. 18ZR1401200), the Fundamental Research Funds for the Central Universities (Grant No. EG2018028), DHU Distinguished Young Professor Program and 2017 CCF-IFAA Research Fund.

References

1. Ericsson: Ericsson Mobility Report, pp. 1–3, August 2019
2. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. IEEE Commun. Mag. **48**(9), 140–150 (2010)
3. Wang, X., et al.: A city-wide real-time traffic management system: enabling crowdsensing in social internet of vehicles. IEEE Commun. Mag. **56**(9), 19–25 (2018)
4. Kalogiros, L.A., Lagouvardos, K., Nikoletseas, S., Papadopoulos, N., Tzamalis, P.: Allergymap: a hybrid mHealth mobile crowdsensing system for allergic diseases epidemiology: a multidisciplinary case study. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 597–602. IEEE, March 2018
5. Roitman, H., Mamou, J., Mehta, S., Satt, A., Subramaniam, L.V.: Harnessing the crowds for smart city sensing. In: Proceedings of the 1st International Workshop on Multimodal Crowd Sensing, pp. 17–18. ACM, November 2012
6. Schobel, J., Pryss, R., Reichert, M.: Using smart mobile devices for collecting structured data in clinical trials: results from a large-scale case study. In: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, pp. 13–18. IEEE, June, 2015
7. Yang, D., Xue, G., Fang, X., Tang, J.: Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, pp. 173–184. ACM, August 2012
8. Jaimes, L., Vergara-Laurens, I., Labrador, M.A.: A location-based incentive mechanism for participatory sensing systems with budget constraints. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom 2012), pp. 103–108, March 2012
9. Mendez, D., Labrador, M., Ramachandran, K.: Data interpolation for participatory sensing systems. Pervasive Mob. Comput. **9**(1), 132–148 (2013)
10. Reddy, S., Estrin, D., Srivastava, M.: Recruitment framework for participatory sensing data collections. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive 2010. LNCS, vol. 6030, pp. 138–155. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12654-3_9

11. Chon, Y., Lane, N.D., Kim, Y., Zhao, F., Cha, H.: Understanding the coverage and scalability of place-centric crowdsensing. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3–12. ACM, September 2013
12. Zhao, D., Ma, H., Liu, L.: Energy-efficient opportunistic coverage for people-centric urban sensing. Wirel. Netw. **20**(6), 1461–1476 (2014)
13. Xiong, H., Zhang, D., Wang, L., Chaouchi, H.: EMC 3: energy-efficient data transfer in mobile crowdsensing under full coverage constraint. IEEE Trans. Mob. Comput. **14**(7), 1355–1368 (2014)
14. Sun, J., Pei, Y., Hou, F., Ma, S.: Reputation-aware incentive mechanism for participatory sensing. IET Commun. **11**(13), 1985–1991 (2017)
15. Lee, J.S., Hoh, B.: Sell your experiences: a market mechanism based incentive for participatory sensing. In: 2010 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 60–68. IEEE, March 2010
16. Chen, C., Wang, Y.: SPARC: strategy-proof double auction for mobile participatory sensing. In: 2013 International Conference on Cloud Computing and Big Data, pp. 133–140. IEEE, December 2013
17. Jaimes, L.G., Vergara-Laurens, I., Labrador, M.A.: A location-based incentive mechanism for participatory sensing systems with budget constraints. In: 2012 IEEE International Conference on Pervasive Computing and Communications, pp. 103–108. IEEE, March 2012
18. Feng, Z., Zhu, Y., Zhang, Q., Ni, L.M., Vasilakos, A.V.: TRAC: truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 1231–1239. IEEE, April 2014
19. He, S., Shin, D.H., Zhang, J., Chen, J.: Toward optimal allocation of location dependent tasks in crowdsensing. In: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 745–753. IEEE, April 2014
20. Billionnet, A., Elloumi, S., Plateau, M.C.: Quadratic 0–1 programming: tightening linear or quadratic convex reformulation by use of relaxations. RAIRO-Oper. Res. **42**(2), 103–121 (2008)
21. Körner, F.: Integer quadratic optimization. Eur. J. Oper. Res. **19**(2), 268–273 (1985)