Gao, Hui; Xiao, Yu; Yan, Han; Tian, Ye; Wang, Danshi; Wang, Wendong

A Learning-based Credible Participant Recruitment Strategy for Mobile Crowd Sensing

Published in:
IEEE Internet of Things Journal

DOI:
10.1109/JIOT.2020.2976778

E-pub ahead of print: 01/01/2020

Document Version
Peer reviewed version

Please cite the original version:
Gao, H., Xiao, Y., Yan, H., Tian, Y., Wang, D., & Wang, W. (2020). A Learning-based Credible Participant Recruitment Strategy for Mobile Crowd Sensing. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2020.2976778

A Learning-based Credible Participant Recruitment Strategy for Mobile Crowd Sensing

Hui Gao, Yu Xiao, *Member, IEEE,* Han Yan, Ye Tian, *Member, IEEE,* Danshi Wang, and Wendong Wang, *Member, IEEE*

***Abstract*—Mobile Crowd Sensing (MCS) acts as a key component of Internet of Things (IoTs), which has attracted much attention. In an MCS system, participants play an important role, since all the data is collected and provided by them. It is challenging but essential to recruit credible participants and motive them to contribute high quality data. In this paper, we propose a learning-based credible participant recruitment strategy (LC-PRS), which aims to maximize the platform and participants' profits at the same time via MCS participation. Specifically, the LC-PRS consists of two mechanisms, that a learning-based reward allocation mechanism (L-RAM) first calculates the maximum offered reward for different locations based on the number of participants in each location. Under a budget constraint, the proposed L-RAM prefers to collect sensing data from locations which relatively few data has so far been collected. Furthermore, for each location, we develop a credible participant recruitment mechanism (C-PRM), which employs semi-Markov model and game theory to predict quality of data provided by each participant and to recruit participants based on the predictions and the maximum offered reward calculated by L-RAM. We formally show LC-PRS has the desirable properties of computational efficiency, selection efficiency, individual rationality and truthfulness. We evaluate the proposed scheme via simulation using three real datasets. Extensive simulation results well justify the effectiveness of the proposed approach in comparison with other two methods.**

***Index Terms*—Participant recruitment, deep reinforcement learning, mobile crowd sensing.**

I. Introduction

Recent advances in mobile computing technologies and Internet of Things (IoTs) technologies enable us to place more built-in sensors and wireless communication modules into a smart device or IoT device [1], [2]. With a rich set of embedded sensors and effective computational capabilities, smart mobile devices (e.g., smartphones, wearable devices, UAVs) are able to collect and share various types of data in urban environments [3], [4]. This paradigm is called "Mobile Crowd Sensing (MCS)" [5]–[7]. In a mobile crowd sensing campaign, people or entities who need sensing data are called "task publishers". When they request to collect some data, there are often accompanied by requirements for data collection such as the desired type, quantity and quality of data, we refer to them as "sensing tasks", or simply "tasks". Mobile users who claim their requested rewards to participate in collecting data, and operate sensors of mobile devices physically or subconsciously, are called "participants". Normally, there is also a central "platform" to recruit participants, process sensing data reported by them and send results back to task publishers.

Mobile crowd sensing has promising applications in many domains, e.g., indoor maps reconstruction, where participants collect sensing data (such as photos or videos) to reconstruct 3D maps [8]. However, a common challenge for most mobile crowd sensing applications is to identify participants who can contribute sensing data that meets the requirement of the task, then to motive them to collect and contribute high quality data [9]. Unfortunately, as participants are self-interested that report less valuable sensing data to minimize their efforts, or have no idea how to collect valuable sensing data, it is difficult for the platform to guarantee to receive valuable service. Therefore, how to design a participant recruitment strategy is one of the most important topics for discussion [10]–[12].

It is worth noting that, for some crowd sensing applications, e.g., indoor map construction, the task coverage is a critical requirement. As low quality data is not equal to useless data, sometimes high quality photos for example are useless, if the photos do not contribute more mapping information. Furthermore, because of its own characteristics, the application prefers to construct the whole indoor map frame at first, then refining details of every place. That means how to collect data from the most number of locations is a significant step. One method to motive participants is that the platform first predicts participants' trajectory, then recruits them based on their mobility pattern [13]–[15]. Another method is to incentivize participants to contribute sensing data from their locations, employing their satisfied rewards [16]–[18]. Therefore, how to encourage participants to contribute sensing data from their locations with the limited budget is an important issue, which will be dealt with in this paper.

Generally speaking, the main challenges of participant re-

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No.2019RC52), the National Key Research and Development Program of China (Grant No. 2019YFB1802603), and the National Natural Science Foundation of China (Grant No. 61802022 and 61802027).

Hui Gao is with School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (E-mail: gaohui786@bupt.edu.cn).

Yu Xiao (Corresponding author) is with Aalto University, Espoo, Finland (E-mail: yu.xiao@aalto.fi).

Han Yan, Ye Tian and Wendong Wang (Corresponding author) are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (E-mail: {yanhan, yetian, wdwang}@bupt.edu.cn).

Danshi Wang is with State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China (E-mail: danshi_wang@bupt.edu.cn).

cruitment are summarized as follows: firstly, rewards are usually paid by task publishers who only have limited budget. That means the platform needs to recruit participants and collect required amount of sensing data under budget constraint [19]. On the other hand, a participant only conducts sensing tasks when he/she is satisfied with the offered reward. Therefore, how to find a balance that satisfies both the platform and participants is an inevitable challenge [20], [21]. Secondly, sensing data is collected by participants who may not receive professional training on how to collect high quality sensing data. Even worse, there may be some dishonest participants who only want to gain rewards but do not contribute sensing data. Then how to recruit participants in order to collect high quality sensing data is another challenge [22]–[24].

To overcome the challenges mentioned above, in this paper, we propose a learning-based credible participant recruitment strategy (LC-PRS), which consists of two steps, i.e., a reward allocation step and credible participant recruiting step. Compared with the previous incentive mechanisms which did not consider task coverage requirement, high quality data collection or credible participant recruitment, or represent incentive mechanism, e.g., encouraging participants to collect sensing data from locations where are rarely sensed. [14], [25]–[32]. The proposed method aims at finding a balance between the limited budget and the coverage requirement of a task. As the budget is limited, the platform prefers to recruit participants located in the "important" grids where sensing data is rarely collected, in order to meet the coverage requirement of a task. More explanations for the important grid will be given in Section IV. We propose a threshold named maximum offered reward in the reward allocation step, only participants whose requested rewards are less than the maximum offered reward could have a chance to be recruited. Furthermore, the maximum offered reward is dynamically changed, based on the desired amount of sensing data and the number of participants.

For the credible participant recruiting step, with the purpose of collecting high quality sensing data, the semi-Markov model is first employed to predict quality of data provided by each participant. Then in order to help the platform make participant recruitment decision with the consideration of the maximum offered rewards calculated by the first step and the prediction results, we employ a game theory to analyse strategies of participants and the platform, respectively. Based on the analysed strategies, we finally find out ones which satisfy the Nash equilibrium.

The main contribution of this paper is summarized:

- We first model the profits of the platform and participants, and formulate the process of participant recruitment as a joint optimization problem, with the objectives of maximizing both the platform and participants' profits together. The problem proves to be a NP-hard problem.
- To deal with the optimization problem, a learning-based credible participant recruitment strategy is proposed, which consists of a learning-based reward allocation mechanism and a credible participant recruitment mechanism.
- We formally show that the proposed participant recruitment strategy has the desirable properties of truthfulness, individual rationality, budgetary feasibility and computational efficiency.
- We perform extensive simulations on three sets of real datasets. The results show that our approach outperforms previous works in terms of profits and the amount of high quality sensing data.

The rest of this paper is organized as follows. We discuss related research efforts in Section II. The system model is described in Section III. We introduce the learning-based credible participant recruitment strategy in Section IV. We present the simulation results in Section V. The practical issue is discussed in Section VI. Finally, we conclude the paper in Section VII.

II. Related Work

There has been much research on mobile crowd sensing [33]–[35], among which participant recruitment is an important issue [36]–[39]. As we mentioned in Section I, a participant recruitment issue is normally related with reward allocation and participant selection problems.

A. Incentive Mechanisms

Wang *et al.* proposed an online incentive mechanism that stimulated participants for a task based on the relative popularity among tasks by considering inequality of time-sensitive and location-dependent tasks, and allocated budgets to these tasks based on their required finish time and sensing locations [25].

Peng *et al.* incorporated the consideration of data quality into the design of incentive mechanism. The mechanism estimated the quality of sensing data, and offered each participant a reward based on his/her contribution. Authors extended the Expectation Maximization method which combines maximum likelihood estimation and Bayesian inference to estimate the quality of sensing data [26].

Gong *et al.* proposed their incentive mechanism for mobile crowd sensing systems, which incentivize strategic participants to truthfully report their private quality and data to the requester, and make truthful effort as desired by the requester [27].

Gao *et al.* focused on the incentive mechanism design for a vehicle based crowd sensing system, which contained a nontrivial set cover problem. Authors proposed a reverse auction based incentive to solve the problem [28].

Tao *et al.* employed a Stackelberg based game method to design their incentive mechanism. For every contributed sensing data, the platform calculated data utility, which took sensing time, distance and orientation into consideration. Then the platform gave participants rewards based on the data utility [29].

For a location-constrained MCS, Restuccia *et al.* proposed that the capability of participants to execute sensing tasks depended on their mobility pattern, which was often uncertain. They designed an incentive mechanism that employed reverse auction to recruit participants with uncertain mobility [40].

Xu *et al.* presented a vehicular location-constrained crowd sensing system. The system incentivized the participants to match the sensing distribution of the sampled data to the desired target distribution with a limited budget. They formulated the incentivizing problem as a knapsack problem and proposed an algorithm named iLOCuS to solve the problem [41].

Yan *et al.* designed a peer-based data exchanging model, in which relay nodes moved to certain locations to connect data providers and consumers for facilitating data delivery. Task publishers were willing to pay for the data and these rewards were given to both relays and participants. They designed an autonomous compensation game for relay nodes to make decisions of where to go individually [42].

Fan *et al.* proposed a joint trajectory scheduling and incentive mechanism for spatio-temporal UVCS systems. They designed an online incentive mechanism that decided whether to recruit a participant when he/she asked to contribute sensing data [43].

B. Recruitment Strategy

Usually, the final goal for designing incentive mechanisms is recruiting participants and then receiving their contributed sensing data. Therefore, it is essential to survey some state of art participant recruitment relative research.

Normally vehicle trajectories are predictable, which provide not only the current locations of the vehicles, but also their future mobility trajectories. Therefore, Wang *et al.* proposed vehicle recruitment algorithms, which considered the mobility of vehicles. Authors aimed to minimize the overall recruitment cost and employed a greedy algorithm to solve it [14].

Yi *et al.* modeled the participant recruitment problem as a unconstrained maximization problem without explicitly cost constraint, and a trade-off parameter introduced to control the recruited participant cost [30].

Chen *et al.* proposed that there were two kinds of redundancy in participant recruitment mechanisms. One was brought by the incomplete coverage assessment, while the other one was brought by the traditional participant selection process. In order to solve the issues, authors proposed a participant recruitment mechanism that selected a segment trajectory of participant, rather than selected the whole one [31].

Zhao *et al.* proposed a reputation based participant recruitment mechanism. The platform employed reputation value to calculate “quality risk” and then participants’ rewards. The quality risk value indicated the probability that participants contributed high quality sensing data. And the platform recruited participants based on the rewards [32].

Pouryazdan *et al.* studied two reputation calculation methods for recruiting participants, named 1) vote-based reputation calculation method which calculated reputation by means of a voting procedure, 2) anchor-assisted decentralized reputation calculation method which deployed some anchor nodes who were definitely trustworthy. Then authors introduced a metric named collaborative reputation scores, which were calculated by a weighted function of the vote-based and anchor assisted decentralized reputation components [44].

Jin *et al.* proposed a payment mechanism that aimed at selecting participant who could contribute reliable data. For data truth discovery, the participants whose data are closer to the aggregated results would be assigned higher weights, and the data from a participant with a higher weight would be counted more in the aggregation [45].

TABLE I
LIST OF IMPORTANT NOTATIONS

Notation	Explantation
B, $\mathscr{L}$, $\mathscr{S}$, $\mathscr{I}$, η	Budget, set of grids, set of stages, set of participants, the number of required pieces of sensing data
c_i^r, c^b, $c_l^f(s)$, $c_l(s)$	Requested reward of participant i, basic reward, the maximum offered reward and floating reward in grid l at stage s
$P_i^h(s)$, $P_i^t(s)$	Probability of high quality sensing data contributed by participant i, probability of recruiting i in at stage s
$U_{platform}$, U_i	Profits of the platform and a participant i
θ_i, g_i	Cost of participant i, profit gained from a piece of high quality sensing data
$\mathscr{V}$	Set of floating reward candidates
F, r_l^s, $c_l(s)$	state transition, reward function and policy function
n^s, a^s, r^s	state, action, reward of stage s
$\lambda_l(s)$	Shapley value of grid l at stage s
h, u	Sensing data quality
$W_i(\cdot)$	the semi-Markov kernel
$z_i(s)$, $x_i(s)$	The variant to denote whether i is recruited and whether he/she contributes high quality sensing data

Zhou *et al.* considered the fixed-wing UAV-aided location-constrained MCS system and investigated the corresponding joint route planning and task assignment problem from an energy efficiency perspective [46].

Sun *et al.* aimed to select participants and allocate sensing tasks to them so that the sensing costs undertaken by all participants were as balancing as possible, while the requirement of the task publisher for data reliability could be satisfied [47].

Wang *et al.* proposed an algorithm that selected participants to perform location-dependent tasks that spent as little time as possible for the participants. They designed an task allocation scheme and designated a specified path for each participant, considering both time-sensitivity, heterogeneity of the sensing tasks and people-variability of participants [48].

Liu *et al.* aimed at selecting participants who located in grids where were the more useful on data inference. They proposed a three-step strategy to recruit participants, that first selected some candidate participant sets, then estimated which grids were more useful on data inference according to the selected candidates, which finally recruited the proper participant set [49].

Compared with literature reviewed above, we propose a participant recruitment method which deals with sensing data quantity and quality issues. In order to collect enough sensing data, the method employs the reinforcement learning method to set the maximum offered reward dynamically based on the history collection performance. Any participant whose requested reward is lower than that of the maximum offered reward has a chance to be recruited. However, some authors employed the revise auction to design the incentive mechanisms, they did not consider malicious participants who could bid the lowest prices but did not permit to contribute high quality sensing data. Furthermore, compared with the former literature that focuses on collecting sensing data without considering data quality or coverage, the proposed method priors to recruit participants from the important grids which are rarely sensed, with the purpose of collecting data in the whole sensing region comprehensively. We carefully design a credible participant recruitment method which first predicts quality of data provided by each participant, then the method recruits participants based on the predictions and the maximum offered reward.

III. System Model

We consider a MCS system that consists of a single platform residing in the cloud and some participants locating in different grids. Every participant could be stimulated to a crowd sensing task via reward by the platform. The frequently used notations are summarized in Table I.

A. Profit Formulation

For the platform, when a task is received, together with the requirement of quantity and quality of data and budget, such as to collect photographs to reconstruct a 3D map under a limited budget B. The platform first divides the entire sensing region into a set of subregions or grids, which is denoted by $\mathscr{L} \triangleq \{1, 2, \ldots, L\}$. And the platform needs to collect no more than η pieces of sensing data from each grid $l \in \mathscr{L}$. With the purpose of deciding the maximum offered reward of each grid easily and dynamically according to the number of located participants, the platform then divides the entire sensing campaign into a set of stages, which is denoted by $\mathscr{S} \triangleq \{1, 2, \ldots, S\}$. Besides, multi-stage process offers participants who are not selected at one stage more opportunities to be selected at the next stage.

For any participant who prepares to perform the task in one stage, he/she will claim his/her location and requested reward at the very beginning of the stage. The platform will select several participants according to their requested rewards and locations, and the task requirement. There are a number of participants located in sensing region and denoted by the set $\mathscr{I} \triangleq \{1, 2, \ldots, I\}$. Each participant $i \in \mathscr{I}$ has a requested reward which is denoted by c_i^r. The probability of contributing high quality sensing data by each participant at stage $s \in \mathscr{S}$ is denoted by $P_i^h(s)$, which is stored by the platform. In order to collect high quality sensing data, on the other hand, the platform also calculates a value denoted by $P_i^t(s)$, which indicates the probability of recruiting the participant $i \in \mathscr{I}$ at stage $s \in \mathscr{S}$. The value of $P_i^t(s)$ will be calculated later in Section IV-B. The participant's cost that performs a task and contributes high quality sensing data is denoted by θ_i, where $i \in \mathscr{I}$. For the sake of simplicity, we assume there is no cost if a participant contributes low quality sensing data. The profit that the platform gains from every sensing data is denoted by g_i. The profits of the platform is defined as $U_{platform} = \sum_{i=1}^{I}\sum_{s=1}^{S}(g_i - c_i^r)x_i(s)$, where $x_i(s) = 1$ if a participant $i \in \mathscr{I}$ contributes high quality sensing data at stage s, otherwise $x_i(s) = 0$. And the profits of a participant $i \in \mathscr{I}$ can be defined as $U_i = c_i^r - \theta_i$.

B. Problem Formulation

The social welfare can be defined as $U_{platform} + \sum_{i=1}^{I}\sum_{s=1}^{S} U_i = \sum_{i=1}^{I}\sum_{s=1}^{S}(g_i - \theta_i)x_i(s)$. Based on that, the target of this paper is

can be formulated as:

$$\begin{aligned} \text{maximize: } & \sum_{i=1}^{I}\sum_{s=1}^{S}(g_i - \theta_i)x_i(s) \\ \text{subject to: } & \sum_{i=1}^{I}\sum_{s=1}^{S} c_i^r z_i(s) \leq B \\ & \sum_{i=1}^{I}\sum_{s=1}^{S} x_i(s) \leq L\eta, \end{aligned} \tag{1}$$

where $z_i(s) = 1$ represents that the platform recruits a participant i at stage s, otherwise $z_i(s) = 0$.

Theorem 1. *The proposed maximum target is an NP-hard problem.*

Proof. We consider a special case of (1) with that: the recruited participants always contribute high quality sensing data, for any participant $i \in \mathscr{I}$, there is $g_i - \theta_i = c_i^r$. Furthermore, the budget will be exhausted before the required amount of high quality sensing data is met. Then (1) could be rewritten as:

$$\begin{aligned} \text{maximize: } & \sum_{i=1}^{I}(g_i - \theta_i)z_i(s) \\ \text{subject to: } & \sum_{i=1}^{I} c_i^r z_i(s) \leq B. \end{aligned} \tag{2}$$

(2) is equivalent to the subset-sum problem, which is a well known example of NP-hard problem. In order to prove that, we turn (2) into a set of problems. Considering any family D of n sets, each set contains three cardinalities. We shall construct $g_1 - \theta_1, g_2 - \theta_2, \ldots, g_n - \theta_n$ and B such that there is a subset of the $(g_k - \theta_k)$ who can sum to B if and only if there exists a subfamily of D covering exactly the universe $\mathscr{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_{3m}\}$, where m is an integer and $m > 0$.

All sets in D can be represented as bit-vectors of length $3m$, e.g., $\{\alpha_1, \alpha_5, \alpha_6\}$ and $\{\alpha_2, \alpha_4, \alpha_6\}$ can be represented as 100011 and 010101, respectively. Based on this, we represent (2) in $(n+1)$-ary form, which is shown as $g_k - \theta_k = \sum_{\alpha_i \in \mathscr{S}_k}(n+1)^{i-1}$, and $B = \sum_{j=0}^{3m-1}(n+1)^j$, where $\mathscr{S}_k \subseteq \mathscr{S}$ and B corresponds to the sequence $11\ldots1$ (with length $3m$). We next prove that D has a subset with sum B, if and only if there exists a subfamily of D that exactly covers $\{\alpha_1, \alpha_2, \ldots, \alpha_{3m}\}$.

The sufficiency is proved as follows. Suppose that there exists a set of $\mathscr{S} = \{1, 2, \ldots, m\}$ that satisfies $\sum_{j \in \mathscr{S}}(g_i - \theta_i) = B$. In order to conduct this summation in $(n+1)$-ary arithmetic, we notice that only the digits 0 and 1 appear in the summands, and the number of summands is less than $n+1$, which means there is no "carry" happening in this addition. Consequently, if the function $\sum_{j \in \mathscr{S}}(g_i - \theta_i) = B$ is established, which means that there is only one number 1 exactly existing in each of the $3m$ positions, or, equivalently, the

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JIOT.2020.2976778

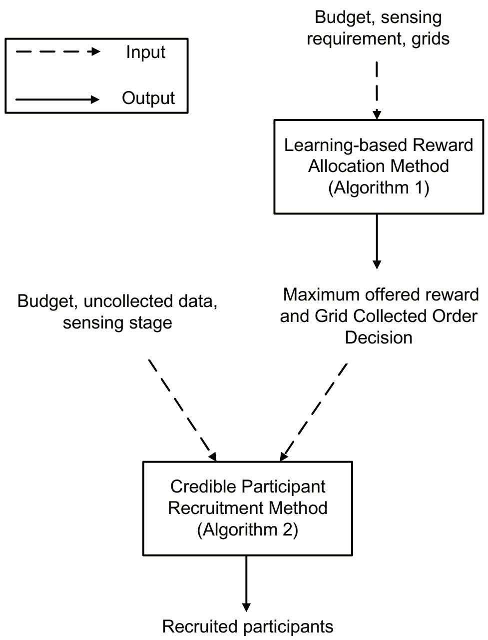

Fig. 1. Framework of proposed approaches.

subfamily $G - \Theta = \{\mathscr{S}_j : j \in \mathscr{S}\}$ covers $\{\alpha_1, \alpha_2, \ldots, \alpha_{3m}\}$ exactly.

The necessity is proved as follows. Since $G - \Theta$ covers $\{\alpha_1, \alpha_2, \ldots, \alpha_{3m}\}$ exactly, $\sum_{j\in\mathscr{S}}(g_i - \theta_i) = B$, which completes the proof of necessity. □

IV. Learning-based Credible Participant Recruitment Strategy

In this section, we first introduce our reward allocation method, then the quality prediction and participant recruitment method is proposed. Finally, we prove that our algorithm meets four desired properties, i.e., selection efficiency, individually rationality, computationally efficiency, and truthfulness. Fig. 1 illustrates the framework of our proposed approaches. Firstly, the reward allocation method takes the budget, sensing requirement and the number of grids as input, and produces the maximum offered reward and grid collected order as output. Then the participant recruitment method takes the budget and the number of sensing stages, together with the output of the former method as input, and produce the recruited participants as output.

A. Learning-based Reward Allocation Method

For the platform, the purpose of recruiting participant is buying their sensing data. And during this transaction, the price acts as a signal to reflect sensing data supply and demand, which depends on the demand of the platform and supply of participants. This is defined as "law of supply and demand" in market economy theories. In brief, if all other factors remain equal, the more quantity demand of the platform, the higher price of each piece of sensing data. On the other hand, the more quantity supply of participants, the lower price of each piece. Therefore, besides cost of collecting data, the price of every piece of sensing data also follows the law of supply and demand.

The purpose of proposing maximum offered reward is helping the platform decide the maximum reward offered to the participants, which follows the law of supply and demand. The maximum offered reward of every stage is denoted by $c_l^f(s)$, which is composed by two parts, namely: basic offered reward c^b and floating reward $c_l(s)$. The former one is fixed which indicates the ideal reward if the platform wants to collect the requested amount of sensing data under budget constraint. However, in reality, lower reward offered by the platform in some grids can attract participants to contribute sensing data, if there are more participants located, or the amount of collected sensing data nearly meets the task requirement. On the other side, the maximum offered reward should be higher in order to collect sensing data when the opposite condition happens. This is the reason why the floating reward exists.

Then how to decide how much floating reward offered for one piece of sensing data in every grid at each stage has become the main challenge. Here we introduce a deep reinforcement learning method to calculate the floating reward. Coarsely speaking, the proposed method involves a decision agent that repeatedly observes the current states of the participant recruitment, then takes an action among the available actions allowed in that state. After, the agent will transfer to a new state and obtain a reward.

1) State Space: $\mathscr{N} \triangleq \{\boldsymbol{n}^s = (N^1, N^2)\}$ denotes the state that indicates whether the platform recruits participants or not.

2) Action Space: $\mathscr{A} \triangleq \{a^s | a \in \mathscr{V}\}$ denotes the action set.

3) Probability Distribution and State Transition: $F : \mathscr{N} \times \mathscr{A} \times \mathscr{N} \to [0, 1]$ denotes the probability distribution $P\{\boldsymbol{n}^{s+1} | \boldsymbol{n}^s, \{a^s\}_{s\in\mathscr{S}}\}$ of a state transition, in which the current state is $\boldsymbol{n}^s$ and when action a^s is chosen, the state is transitioned to a new state $\boldsymbol{n}^{s+1}$.

4) Reward Function: $\mathscr{N} \times \mathscr{A} \to \mathbb{R}$ expresses the expected immediate reward received after the state is transitioned from $\boldsymbol{n}^s$ to $\boldsymbol{n}^{s+1}$, due to taking the action a^s, $s \in \mathscr{S}$, which is defined as: $r^s = e^{a^s} / \sum_{k=1}^{K} e^{v_k}$. Here we employ the softmax value to calculate the reward.

5) Problem Formulation: When state transition F and reward function r_l^s, $l \in \mathscr{L}$, $s \in \mathscr{S}$ is predetermined, for each stage s, our problem can be formulated as:

$$Q_l(\boldsymbol{n}^s) = \max_{a_l^s} \Big[r_l^s(\boldsymbol{n}^s, a_l^s) + \gamma \int_{\boldsymbol{n}^s \in \mathscr{N}} F(\boldsymbol{n}^s, a_l^s, \boldsymbol{n}^{s+1}) Q_l(\boldsymbol{n}^{s+1}) \Big], \tag{3}$$

and the optimal strategies of the floating reward is given by

$$c_l(s) = \arg\max_{a_l^s} \Big[r_l^s(\boldsymbol{n}^s, a_l^s) + \gamma \int_{\boldsymbol{n}^s \in \mathscr{N}} F(\boldsymbol{n}^s, a_l^s, \boldsymbol{n}^{s+1}) Q_l(\boldsymbol{n}^{s+1}) \Big], \tag{4}$$

Based on (4), the floating reward can be decided in grid l at stage s. It is worth noting that one special phenomenon may happen, that the sum of maximum offered reward of all grids may exceed the budget. Furthermore, as we mentioned in Section I, for some crowd sensing applications, e.g., indoor map construction, which prefers to construct the whole indoor

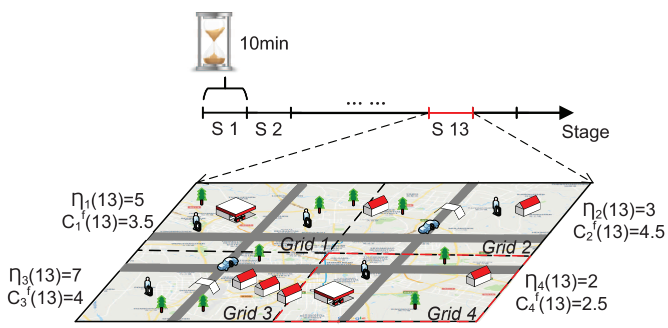

Fig. 2. Example of the proposed participant recruitment strategy.

map frame at first, then refining details of every place. For this reason, some grids where less amount of sensing data is rarely collected are more important. Therefore, a method is needed to help the platform recruit participants preferentially from these grids under the limited budget. Here the Shapley method is employed to identify which grids are important, which is shown as:

$$\lambda_l(s) = \sum_{\mathscr{L}' \subseteq \mathscr{L} \setminus \{l\}} \frac{|\mathscr{L}'|!\big(|\mathscr{L}| - |\mathscr{L}'| - 1\big)!}{|\mathscr{L}|!} f_m\big(\mathscr{L}' \bigcup \{l\}\big), \tag{5}$$

where $f_m\big(\mathscr{L}' \bigcup \{l\}\big)$ is the marginal value, which is shown as:

$$\begin{aligned} &f_m\big(\mathscr{L}' \bigcup \{l\}\big) = \\ &\quad \sum_{\tau=1}^{\eta_l(s)} \Big(\big(1 - \frac{\| \left[\eta_1(s), \eta_2(s), \ldots, \eta_{|\mathscr{L}'|}(s), \eta_l(s) - \tau\right] \|_F}{\eta\sqrt{|\mathscr{L}' \bigcup \{l\}|}}\big) \\ &\quad - \big(1 - \frac{\| \left[\eta_1(s), \eta_2(s), \ldots, \eta_{|\mathscr{L}'|}(s)\right] \|_F}{\eta\sqrt{|\mathscr{L}'|}}\big)\Big), \end{aligned} \tag{6}$$

where $\eta_l(s), l \in \mathscr{L}, s \in \mathscr{S}$ is the number of pieces of sensing data that have not been collected yet, and $\| \cdot \|_F$ is the Frobenius norm, which is mathematically used to measure the spatial length of a matrix, to quantify the difference between the required and attained values.

Let us make a simple example to explain the meaning of important grid, which employs the Shapley value method to measure. As shown in Fig. 2, suppose the whole sensing region is divided into four grids, i.e., Grid 1, Grid 2, Grid 3 and Grid 4, each of which needs 10 pieces of sensing data. We assume that the campaign has completed 12 stages, the amount of data that has not been collected yet for the next stage is $\eta_1(13) = 5, \eta_2(13) = 3, \eta_3(13) = 7, \eta_4(13) = 2$, respectively. Apparently, Grid 3 is the most important one among all of grids as there is not any piece of sensing data collected. The Shapley value is correspondingly the highest among the four grids, which is 0.55, calculated by (6). Grid 1 is the second important one, and the Shapley value is 0.47, the rest of Grid 2 and 4 are 0.27 and 0.15, respectively. The Shapley value indicates sensing data collection order which follows the principle of some crowd sensing applications.

Follow the example, we describe the whole process of the reward allocation method. Suppose the platform pays 100 units budget to collect total 40 pieces of sensing data from the four grids, then the basic offered reward is 2.5 units, i.e., $c^b = 2.5$. Each of stage is assumed to last 10 minutes, i.e., $T = 10$. At the beginning of Stage 13, the platform first has to decide the maximum offered reward. As the set of variables that a floating reward could be $\mathscr{V} = \{-1.5, -1, 0, 1, 1.5, 2\}$, according to the proposed method, the floating rewards of every location are supposed to be $c_1(13) = 1, c_2(13) = 2, c_3(13) = 1.5, c_4(13) = 0$, then the maximum offered reward of each location is $c_1^f(13) = 3.5$, $c_2^f(13) = 4.5$, $c_3^f(13) = 4$, $c_4^f(13) = 2.5$. As the rest of budget may be not enough for paying for collecting the rest amount of sensing data at the next Stage 13, the platform employs Shapley method to decide collection order. Based on the Shapley values calculated above, the order that platform recruits participants is Grid $3, 1, 2, 4$. The results of the maximum offered reward of each grid and collection order will be employed to the credible participant recruitment method.

Algorithm 1 Learning-based Reward Allocation Mechanism (L-RAM)

Input: Budget B, sensing requirement η, grid $\mathscr{L}$
Output: Maximum offered reward of every location $c_l^f(s)$ at stage s

1: $c^b = B/(L * \eta)$;
2: **for** $l = 1, \ldots, L$ **do**
3: Calculate the floating reward candidate by (4);
4: $c_l(s) = \arg\max_{v_k} f_l(k)$;
5: $c_l^f(s) = c^b + c_l(s)$;
6: **end for**
7: Calcualte Shapley value $\lambda_l(s)$ by (5);
8: Rank locations using $\lambda_l(s)$ in descending order;

The learning-based reward allocation mechanism is presented in Algorithm 1; the procedure of which is that, firstly the mechanism calculates the basic reward (Line 1). As the task needs to collect $L * \eta$ pieces of sensing data (L grids need to be sensed and η pieces of data are requested to be contributed at each grid) under budget B, the basic reward means the required amount of sensing data is met meanwhile the budget is exhausted. Then the mechanism calculates the floating reward for each grid at one stage, and the maximum offered reward (Line 2 - Line 6). Finally, the Shapley method is used to rank all of locations in descending order (Line 7 - Line 8).

B. Credible Participant Recruitment Method

As we mentioned in Section I, the recruited participants may not receive professional training on how to collect sensing data in a desired manner. Therefore, the platform needs a participant recruitment strategy to recruit credible participants, with the purpose of avoiding to receive unsatisfying service. The unsatisfying service is defined as not only receiving low quality sensing data, but also including a dishonest situation that a recruited participant does not contribute sensing data after he/she gains reward. Actually, the dishonest situation is a special case that a recruited participant contributes low quality sensing data, which is also useless for the platform. Therefore, in this paper, we employ low quality sensing data to represent these two types of unsatisfying service.

We denote the quality of data by $q_{i,n}$, which is contributed by a participant i in the nth time. And the quality of data contributed by a participant is modeled as a semi-Markov with discrete time, which means that the probability of a participant i contributing high quality data at the nth time depends on that of the $(n-1)$th time. Therefore, process $q_{i,n}$ is a standard discrete time Markov chain.

We define the semi-Markov kernel part in (7), where $W_i^{uh}(s)$ represents the probability that a participant i contributes high quality sensing data at a stage s while he/she contributed low quality sensing data last time. And T is the time period of every stage $s \in \mathcal{S}$, $t_i(s)$ is the lasting time that a participant contributes sensing data. Here we assume that a participant may not contribute sensing data at the very beginning of every stage, but he/she contributes sensing data before the end of stage.

$$W_i^{uh}(s) = P(q_{i,n} = h, t_i(s) \le T | q_{i,n-1} = u). \tag{7}$$

Next we define the probability that a participant i will contribute high quality sensing data at the nth time when he/she contributes unusable data at the $(n-1)$th time, before time unit T as $Z_i^{uh}(T)$, which is shown as:

$$\begin{aligned} Z_i^{uh}(T) &= P(t_i(s) \le T | q_{i,n} = h, q_{i,n-1} = u) \\ &= \sum_{x=1}^{T} P(t_i(s) = x | q_{i,n} = h, q_{i,n-1} = u). \end{aligned} \tag{8}$$

The probability that a participant i contributes high quality sensing data at the nth time, given he/she contributes unusable quality data at the $(n-1)$th time is shown as $P_i^{uh} = P(q_{i,n} = h | q_{i,n-1} = u) = num_i^{uh}/num_i^u$, where num_i^{uh} is the number of times data quality contributed from unusable to high quality, while num_i^u is the number of times unusable data contributed.

We rewrite (7) based on (8), which is shown as:

$$\begin{aligned} W_i^{uh}(s) &= P(q_{i,n} = h, t_i(s) \le T | q_{i,n-1} = u) \\ &= Z_i^{uh}(T) P_i^{uh}. \end{aligned} \tag{9}$$

Based on (9), the probability that a participant i contributes high quality sensing data at a stage s is calculated by $P_i^h(s) = \big(W_i^{uh}(s) + W_i^{hh}(s)\big) / \big(W_i^{uh}(s) + W_i^{uu}(s) + W_i^{hu}(s) + W_i^{hh}(s)\big)$.

We next describe the platform and selected participants' strategies which follow the dynamic two-round game theory, and discuss the conditions following Nash equilibrium. We first describe the strategies of a selected participant in the second round, who can employ a strategy to contribute high quality sensing data, or low quality sensing data. The expectation returns that he/she gains are denoted by $P_i^h(s)(c_i - \theta_i)$ when he/she wants to contribute high quality sensing data, or $(1 - P_i^h(s))c_i$ when he/she wants to contribute low quality sensing data, respectively. If the selected participant is required to contribute high quality sensing data, which has to satisfy $P_i^h(s)(c_i - \theta_i) \ge (1 - P_i^h(s))c_i$. Here we assume a participant is rational, that if the returns when he/she contribute high or low quality sensing data are equal, he/she will choose to contribute high quality sensing data in order to keep performing crowd sensing campaign and gaining rewards. On the other hand, the expected returns are denoted by $P_i^t(s)(g_i - c_i)$ when the platform recruits the participant, or $(1 - P_i^t(s))\tau_i$ when the platform recruits another participant.

Algorithm 2 Credible Participant Recruitment Mechanism (C-PRM)

Input: Budget B, uncollected data $\eta_l(s)$, sensing stage $\mathcal{S}$
Output: Recruited participants $\mathcal{I}_s$

```
1: for s=1,...,S do
2:     Calucation of the maximum offered reward in Algorithm 1;
3:     for Recruit participants from grids in the new order do
4:         if c_i^r ≤ c_l^f(s) && c_i^r ≤ B && η_l(s) > 0 && P_i^h(s) ≥ 0.5 then
5:             Calculate P_i^t(s) by (12);
6:             num = 1 with probability P_i^t(s);
7:             if num == 1 && data is contributed then
8:                 i → 𝓘_s;
9:                 η_l(s) = η_l(s) − 1;
10:                B = B − c_i^r;
11:            end if
12:        end if
13:    end for
14: end for
```

Let $\tau_i = \beta c_i$, $\beta \ge 0$ is the least return the platform wants to gain. If the participant wants to be recruited, it has to satisfy $P_i^t(s)(g_i - c_i) \ge (1 - P_i^t(s))\tau_i$. The two inequalities we discussed above satisfy the Nash equilibrium, which we have that:

$$\begin{aligned} \theta_i &\le \frac{c_i\big(2P_i^h(s) - 1\big)}{P_i^h(s)} \\ g_i &\ge \frac{\tau_i\big(1 - P_i^t(s)\big) + c_i P_i^t(s)}{P_i^t(s)}. \end{aligned} \tag{10}$$

Based on (10), we know that the lower boundary of (1) is:

$$\sum_{i=1}^{I}\sum_{s=1}^{S}\Big(\frac{\tau_i\big(1 - P_i^t(s)\big) + c_i P_i^t(s)}{P_i^t(s)} - \frac{c_i\big(2P_i^h(s) - 1\big)}{P_i^h(s)}\Big)x_i(s). \tag{11}$$

For every selected participant at a stage, we employ partial derivative function of (11) to calculate the maximum the minimum profit. And we employ the expected value to express $x_i(s)$, that $x_i(s) = P_i^h(s)(1 - P_i^h(s))$. Then we have that:

$$P_i^t(s) = \frac{\tau_i\big(2P_i^h(s) - 1\big)}{\tau_i\big(2P_i^h(s) - 1\big) + 2c_i\big(1 - P_i^h(s)\big)}. \tag{12}$$

From (10) we know that $P_i^h(S) \in [0.5, 1]$, and $\tau_i = \beta c_i$, then we have $\tau_i\big(2P_i^h(s) - 1\big) + 2c_i\big(1 - P_i^h(s)\big) > 0$ and $P_i^t(s) \ge 0$.

The basic work flow of LC-PRS is that, the reward allocation method provides the maximum offered rewards of each grid at the next stage to the credible participant recruitment method. After recruiting credible participants and collecting sensing data, the last method returns the data collection result and remaining budget to the former one. For recruiting participants, if the requested reward of a participant does not exceed the amount of maximum offered reward, he/she has probability to be recruited, which depends on the value of probability $P_i^t(s)$. The platform keeps recruiting participants till the amount of required sensing data is met or the budget is exhausted.

The detailed algorithm of participant recruitment mechanism is presented in Algorithm 2, correspondingly we describe the main processes as follows.

Step 1 of C-PRM: In the beginning of each stage, the platform first calculates the maximum offered reward and allocation order employing Algorithm 1 (see Line 2).

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JIOT.2020.2976778

Step 2 of C-PRM: The platform recruits participants based on the new reward allocation order. For participants whose probability value $P_i^h(s) \geq 0.5$, if his/her requested reward does not exceed the maximum offered reward and rest budget, and there is still several amount of sensing data needed to be collected, then the participant has a chance to be recruited (see Line 4). The platform selects the participant with probability $P_i^t(s)$, if he/she is recruited, the platform first collects the contributed sensing data and then pais the reward to him/her (see Line 5-11).

Step 3 of C-PRM: The crowd sensing campaign is ended when one of the three conditions is met: 1) the last sensing stage is finished, 2) the required amount of sensing data is met, 3) the budget is exhausted.

C. Desired Properties

In this paper, we aim to ensure that our worker selection policy has the following advantageous properties.

Proposition 1. *(Computational efficiency): The proposed mechanism is computationally efficient that it can be executed within polynomial time.*

Proof. We focus on the computational complexity at each stage. Since the mechanism first employs Algorithm 1 to calculate the maximum offered reward of every grid, which takes $\mathcal{O}(L)$ time in Line $2-6$ and $\mathcal{O}(2^L)$ time to calculate Shapley value in Line 7. Therefore, the mechanism takes $\mathcal{O}(2^L)$ time in Line 2 of Algorithm 2. Then the mechanism recruits participants from every grid which takes $\mathcal{O}(LI)$ in the worst condition in Line $3-13$ of Algorithm 2. Therefore, the computational complexity at each stage is bounded by $\mathcal{O}(2^L)$. □

Proposition 2. *(Selection efficiency): The proposed mechanism is selection efficient that its regret is bounded.*

Proof. The regret is defined as the difference of service quality:

$$regret(s) = \sum_{i=1}^{I} \Big(\frac{(g_i - \theta_i) z_i^*(s)}{c_i^r} - \frac{(g_i - \theta_i) z_i(s)}{c_i^r} \Big), \quad (13)$$

where $z_i^*(s)$ and $z_i(s)$ is the optimal and realistic selection of the platform, respectively. Let $\Delta_i = z_i^*(s) - z_i(s)$. And $\phi_i(s)$ denotes the number of times that a participant $i \in \mathcal{I}$ has been chosen before a stage s. $(g-\theta)/c$ is a constant which can be calculated by approximate methods. Then according to the study of Wu *et al*. [50], the expected total regret in the whole number of stages S is given by

$$\begin{aligned} \mathbb{E}\Big[\sum_{s=1}^{S} regret(s)\Big] &\leq \frac{SI(g-\theta)}{c} \mathbb{E}\Big[\sum_{s=1}^{S}\sum_{i=1}^{I}\Big(z_i^*(s) - z_i(s)\Big)\Big] \\ &= \frac{SI(g-\theta)}{c} \sum_{i=2}^{SI} \Delta_i \mathbb{E}[\sum_{s=1}^{S} \phi_i(s)] \\ &\leq \mathcal{O}\Big(\Big[\sum_{i=2}^{SI} \frac{1}{\Delta_i^2}\Big]^2 \log S\Big). \end{aligned} \quad (14)$$

From (14), we can conclude that when the number of stages S increases, cumulative regret's growth rate becomes lower and lower (because the derivative of $\log y$ is $\frac{1}{y}$), which reflects the learning procedure of Thompson sampling. □

Proposition 3. *(Individual Rationality): The proposed mechanism is individually rational.*

Proof. As a selected participant $i \in \mathcal{I}$ can gain his/her requested reward c_i^r, which proofs the proposed mechanism is individually rational. □

Proposition 4. *The proposed mechanism is truthful.*

Proof. Depending on the relationship of supply and demand, which is discussed in Section IV. The platform decides the upper bound of price which considers the desired amount of sensing data. If the requested reward c_i^r of a participant $i \in \mathcal{I}$ is higher than that of maximum offered reward $c_l^f(s)$, he/she will not be recruited. If the participant does not contribute sensing data after being recruited, he/she will not gain reward as the platform only pays recruited participants when their sensing data are contributed. □

V. Performance Evaluation

With the purpose of evaluating the performance of the proposed method, we employ three sets of datasets to do two sets of separate simulation experiments. In this section, we first introduce simulation setup, then present and analyze the simulation results.

A. Setup

The first dataset includes taxi mobility traces collected in Rome, Italy. In the dataset, GPS coordinates of approximately 320 taxis are recorded over 30 consecutive days [51]. We employ the dataset as the participants' trajectories in a mobile crowd sensing campaign. Each trajectory is marked by a sequence of time-stamped GPS points that contain taxi driver ID, time stamp (date and time), and taxi drivers' position (latitude and longitude).

The second dataset is map offset correction data[1]. Map offset is a value that indicates the value gap between GPS coordinates in real world (i.e., accurate values) and those in a digital map. We use the data as "data quality" in our experiment, which is employed to evaluate quality prediction mechanism that is mentioned in Section IV-B.

For the first two sets of datasets, we adopt the following procedures to set up our simulation platform:

- As all traces are recorded in different parts of Rome. We find a region about $800 \times 500\text{m}^2$ which is shown as Fig. 3(a). We use this region as the simulation area for the considered data collection campaign, and Fig. 3(b) shows the GPS points inside the region.
- All traces in the considered region are recorded from 347 potential (candidate) participants, i.e., $I = 347$. Since these traces are recorded at different days, in our simulation we overlay them into one day.

[1](Baidu, Google) map latitude and longitude GPS offset correction. Available: https://www.programering.com/a/MTO1IzNwATg.html

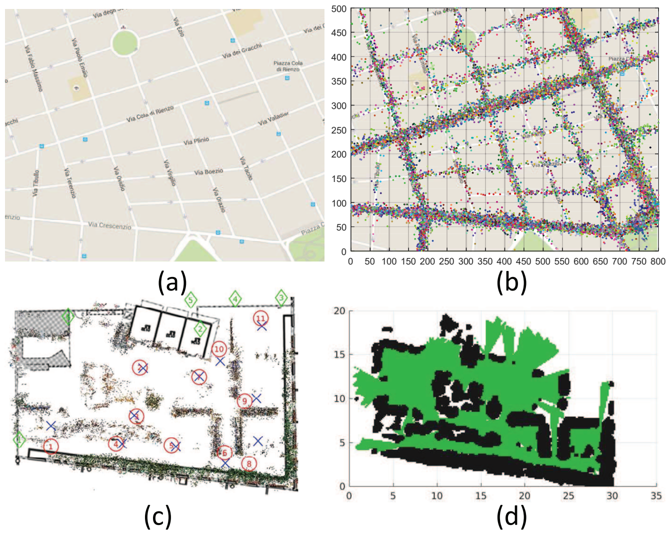

Fig. 3. Datasets employed in the simulation experiments, where (a) shows the sensing region. (b) shows the GPS points inside the region. (c) The library floor plan of Aalto University. (d) The sensing data collection result.

- The whole sensing region is divided into several grids, which is varied by 2, 5, 10, 15 and 20, respectively. The stages are varied from 40 to 60 from the beginning to deadline, with the increment of 5. And the number of piece of requested sensing data of every grid are 270. Budget is varied from 3000 to 4000 units with the increment of 200 units. The floating reward candidates are $V = [-5, 5]$, respectively. For the cold start, we set that the value of probability that the platform recruits a participant and a participant contributes high quality sensing data to 0.5, respectively. The requested rewards of participants are set as the uniformly distributed random numbers varied from 10 to 30 units.
- We employ the map offset values to indicate a participant's sensing data quality, which is denoted by $q_{i,n}, \forall i \in \mathscr{I}$. The map offset of use are nonlinear, in the range of $[300, 500]$ miles. We collect those in the same latitude into a set.

The third dataset is photos collected in a library of Aalto University, Finland (Fig. 3(c) and (d)). The size of the library area is around 350m^2. The photos are snapped by iPhone 7, Galaxy S7, and Nexus 5, respectively. In the dataset, the number of pieces of sensing data contributed by each participant at one time, together with the collection location are recorded. For quantifying the quality of data collected by each participant, as we mentioned in Section I, that low quality data is not equal to useless data, high quality photos for example are useless, if the photos do not contribute more mapping information. Therefore, we employ the ratio value of how many piece of sensing data are actually used and the total number of collected by him/her, which is in the range of $[0, 1]$. We extend the source dataset to 339 pieces of data. We first find out the minimum and maximum ratio values from the source data, then generate each of extended data randomly between them. We list the difference procedures compared with the first two sets of datasets:

- All traces in the considered region are recorded from 339 potential (candidate) participants, i.e., $I = 339$. Since the library is a small sensing regain, the whole sensing region is divided into 10 grids, i.e., $L = 10$.

To evaluate the performance of our proposed strategy (referred as "LC-PRS"), three other approaches are implemented and compared.

- We employ the upper confidence bound (UCB) method to replace the deep reinforcement learning method, which is a classical reinforcement learning method [50]. The UCB method is formulated as: $f_l(k) = X_l(s) + \gamma A_l(s), k \in [1, K]$, where $X_l(s)$ is the average profit that indicates participants has been selected when one floating reward has been chosen. $A_l(s)$ is an upper confidence bias added to the sample mean. Here we refer the UCB method as "UCB".
- We employ a greedy method to be the second compared approach. In our evaluation study, we sort participants by $P_i^h(s)$ in an diminishing order, and select the participants until the budget is exhausted or the required number of piece of sensing data is met (referred as "Greedy").
- The third compared approach select participants randomly at every stage, until the required number of piece of sensing data is met or the budget is exhausted (referred as "Random") [49].
- We employ the method as proposed in [50] to be the forth compared approach, which selects participants based on their service quality effectively in a reinforcement learning manner (referred as "MTSWS"). Here we employ the probability of contributing high quality sensing data $P_i^h(s)$ to be the value of service quality.
- For the final compared approach (referred as "Optimal"), we assume that all of selected participants contribute high quality sensing data, and employ exhaustive method to satisfy (1).

We evaluate the proposed LC-PRS strategy through numerical results in section V-B. We mainly compare the value of social welfare subject to the budget and the number of required data constraint, which is formulated as (1). Furthermore, as the platform needs to collect the high quality sensing data, we also compare the amount of collected high quality sensing data. Besides focusing on the impact of budget, we also consider the number of grids, participants and stages as parameters, all of which could affect the value of social welfare and the amount of collected high quality sensing data.

B. Simulation Results

We present the simulation results in Fig. 4 and 5. Here we set the requested reward of a participant randomly in the range of $[10, 30]$ units. The basic reward $c^b = B/(L * \eta)$, where $L * \eta$ is the number of pieces of requested sensing data. The maximum offered reward is composed of the basic reward and floating reward. The floating reward is a dynamic value, which is in the range of $[-4, 4]$ units. We normalize the map offset value in the range of $[0, 1]$, where 1 means the sensing data is utterly high quality and 0 is on the contrary. We set

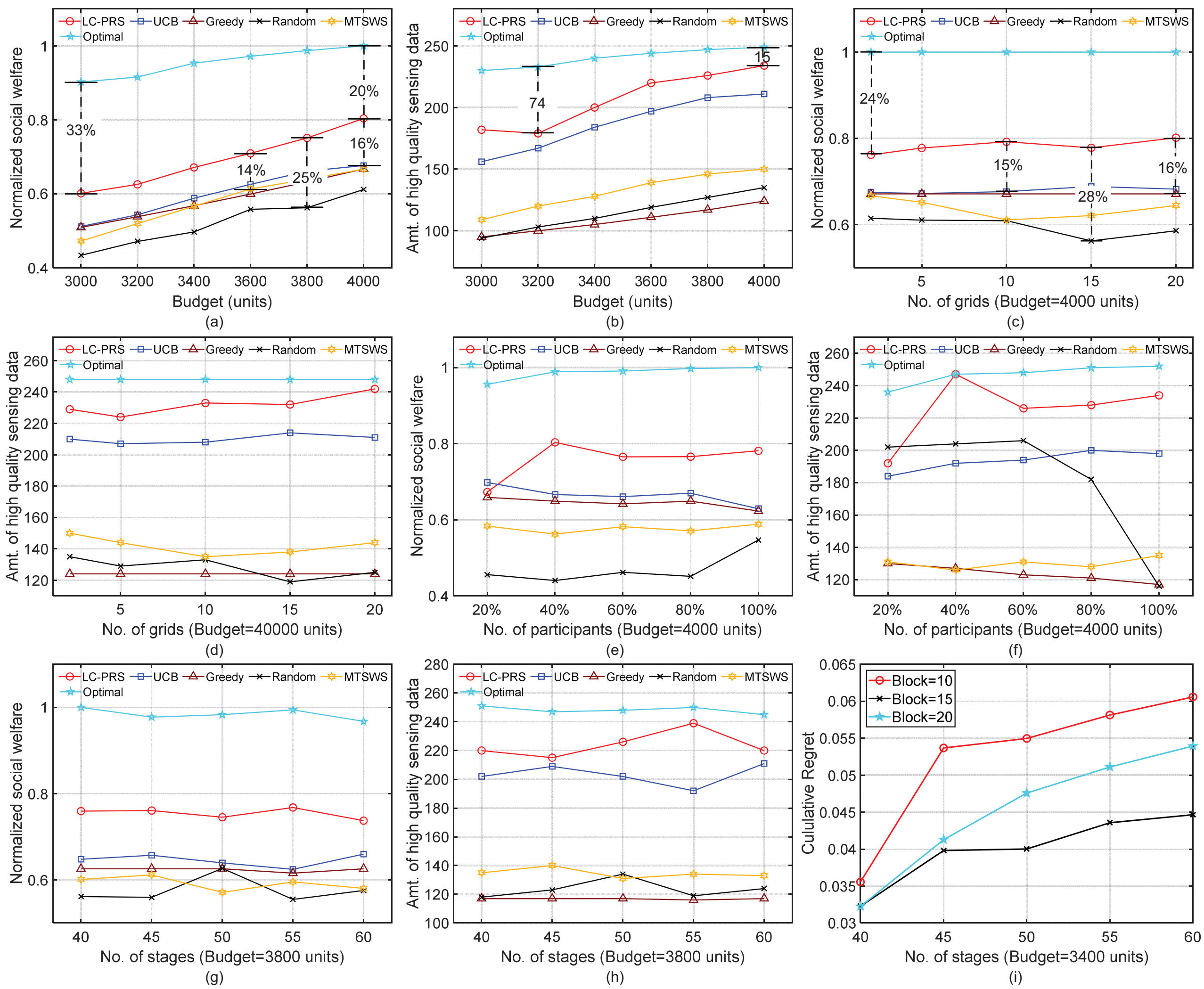

Fig. 4. Simulation results based on the first two datasets. (a-h) show the normalized social welfare and total number of piece of high quality sensing data, under various budget, number of grids, number of participants and number of stages, respectively. (i) shows the growth rates of regret.

the threshold value of high quality sensing data as 0.7. The cost of a participant i is set as $0.8 * c_i^r$.

Fig. 4(a) and (b) show the normalized social welfare, and total number of pieces of high quality sensing data under the setting of 3000, 3200,. . ., 4000 units. The whole region is divided into 15 grids, and the sensing stage is set to 45. Except the Optimal approach, LC-PRS always gains more amount of social welfare and high quality sensing data, compared with other methods. For example, the LC-PRS gains 14%, 16% and 25% more amount of social welfare than that of MTSWS, UCB and Random, when budget is 3600, 4000 and 3800, respectively. Compared with the Optimal that always gains the most amount of social welfare and high quality sensing data, the proposed approach only gains 33% and 20% less amount of social welfare compared with that of Optimal, when budget is 3000 and 4000, respectively. And LC-PRS collects 74 and 15 less amount of sensing data in the worst and best condition, when budget is 3200 and 4000, respectively. We observe that the value gap between the proposed method and Optimal narrows down, with the budget increasing.

Fig. 4(c) and (d) show the normalized social welfare, and total number of piece of high quality sensing data under the setting of 2, 5, 10, 15, 20 grids, where the budget is set to 4000 units, and the sensing stage $S = 60$. From the figures we can find that the proposed approach performs better than that of other four methods. For example, the proposed approach gains 15%, 16% and 28% more amount of social welfare than that of UCB, Greedy and Random, when the number of girds are 10, 20 and 15, respectively. And compared with the Optimal, the LC-PRS only gains 24% less amount of social welfare than that of Optimal in the worst condition, when the number of stages are 2. However, the LC-PRS gains 242 pieces of high quality sensing data when stages are 20 (Fig. 4(b)), and the optimal gains 248 pieces in the same stage, only two more than that of LC-PRS.

Fig. 4(e) and (f) show the results under different number of participants, where the whole region is divided into 15 grids and the sensing stage is set to 40. From Fig. 4(e) and (f), we observe that the proposed approach gains the most amount of social welfare and sensing data, compared with UCB and

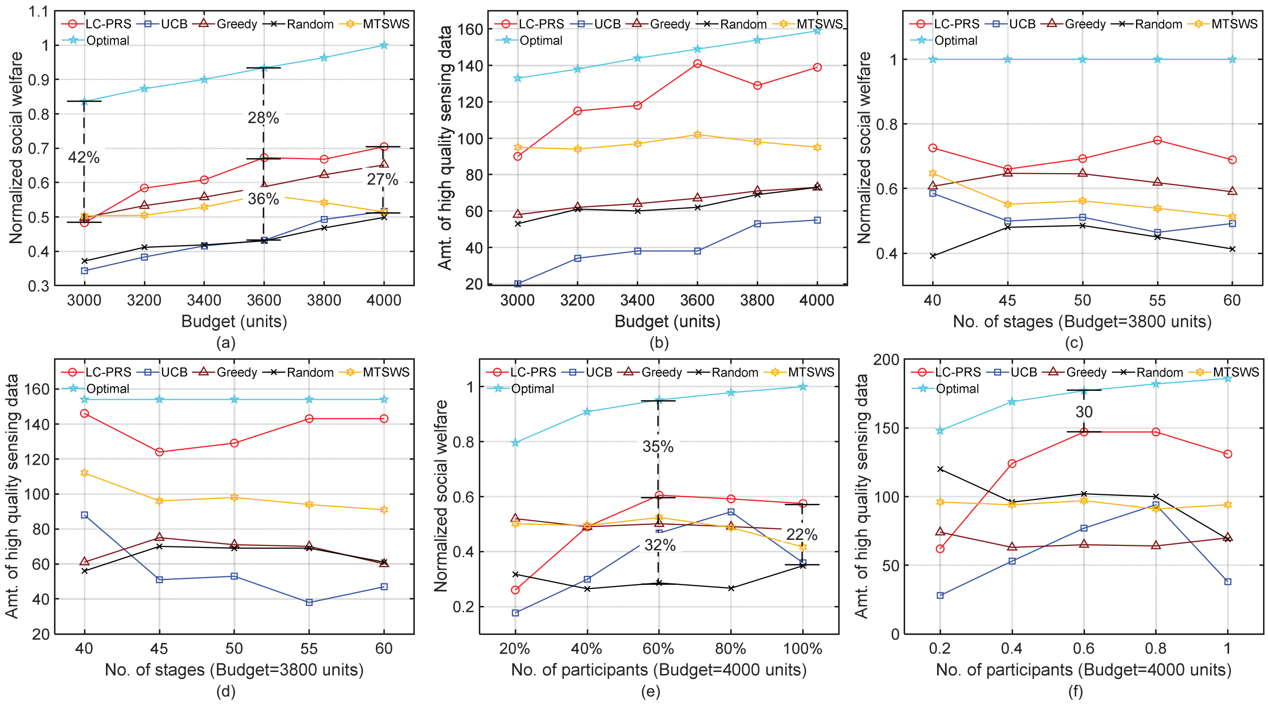

Fig. 5. Simulation results based on the last dataset. (a-f) show the normalized social welfare and total number of piece of high quality sensing data, under various budget, number of stages and number of grids, respectively.

Random, when the number of available participants are set to 40%, ..., 100%. Especially, when there are 40% number of available participants, the proposed method gains as the same amount of high quality sensing data as that of the Optimal.

Fig. 4(g) and (h) show the results under different number of stages, where budget is set to 3800 units, the whole region is divided into 15 stages and the sensing stage is set to 40. In addition, Fig. 4(i) shows the growth rates of regret become flat eventually, which verifies that the proposed approach is bounded.

From Fig. 5(a) and (b) we observe that the LC-PRS always gains more amount of social welfare and high quality sensing data than that of UCB and Random, under the setting of grids $L = 10$ and sensing stage $S = 50$. Compared with the Optimal, LC-PRS gains 28% and 42% less amount of social welfare under the best and worst results. However, the proposed method could gains only 8 pieces of high quality data less than that of Optimal, when budget is 3600 units. The proposed method also gains 27% and 36% more amount of social welfare than that of Greedy and UCB, when the budget is 4000 and 3600 units, respectively. Fig. 5(c) and (d) show the results of different number of stages under budget is 3800 units and grids is 10. While Fig. 5(e) and (f) is the results of different number of participant candidates when budget is 4000 units and sensing campaign is divided into 45 stages. We observe that though the Greedy, Random and MTSWS gains more amount of social welfare and high quality sensing data, the proposed method performs better than that of other four methods, in the most of time.

VI. Discussions

A. Practical Issue Related to Participant's Sensing Cost

In this paper, we assume there is no cost if a participant contributes low quality sensing data. Take the iPhone 11 Pro Max for example, which battery power is $3,969$ mAh[2]. It takes only 0.007 RMB to fully recharge a iPhone. As a task of MCS usually employs some sensors of a smartphone, the cost of which is even lower than that of recharging a smartphone. Compared with the requested reward of a participant, the cost of contributing low quality sensing data could be ignoring. However, if a participant wants to contribute high quality sensing data, he/she may perform several times before contributing the most satisfactory data, which normally will spends much more battery power and time than that of participant who usually contribute low quality data. Therefore, we take the cost of contributing high quality sensing data into account.

B. Practical Issue Related to Participant's Sensing Time

As we assume that a participant may not contribute sensing data at the very beginning of every stage, but he/she contributes sensing data before the end of stage. The assumption is based on the scenario that the platform first broadcasts a task at the beginning of every stage. Participants who have an interest in the task claim to participate at this stage. Finally the platform selects several participants and prepares for the next stage. The participants who could not perform the task during a stage time will not participate. However, if a participant perform a task but does not contribute sensing data, he/she

[2]https://9to5mac.com/2019/09/17/iphone-11-and-iphone-11-pro-battery-size/

will be seen as contributing low quality sensing data at this stage. And this maybe affect the chance of selecting him/her for next time.

VII. Conclusion and Future Work

In this paper, we proposed a learning-based credible participant recruitment strategy for mobile crowd sensing. We designed a participant recruitment algorithm, which consisted of two steps. For the first step, the platform calculated the maximum offered reward for every grid and participant recruitment order, based on the number of participants located in each grid and the rest required amount of sensing data. With the purpose of changing the maximum offered reward dynamically according with the varied number of participants, we divided the whole sensing campaign into several stages. After that, a semi-Markov model was employed to predict quality of data provided by each participant. Finally, we employed a two-round game theory to help the platform recruit credible participants. The result shown that the proposed algorithm gained most amount of social welfare and high quality sensing data, compared with other two methods.

As for the future, we plan to design a new participant recruitment strategy considering participants' mobility pattern. For the reward allocation step, there are two schemes to calculate the reward, one is that the platform first forecasts participants' trajectory and then decides the maximum offered rewards based on the predicted number of participants of every grids. Another scheme is the participant first claims his/her starting point and destination to the platform. Then the platform plans the route with considering the rarely sensed grid he/she could pass by. It is a challenge that the platform should offer a satisfied reward to the participant, with the purpose of compensating him/her for taking a long way around.

References

[1] P. M. Santos, J. G. Rodrigues, S. B. Cruz, T. Lourenço, P. M. dâĂŹOrey, Y. Luis, C. Rocha, S. Sousa, S. Crisóstomo, C. Queirós *et al.*, "Porto-livinglab: An iot-based sensing platform for smart cities," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 523–532, 2018.

[2] C. H. Liu, X. Ma, X. Gao, and J. Tang, "Distributed energy-efficient multi-uav navigation for long-term communication coverage by deep reinforcement learning," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[3] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M'hamed, "Sparse mobile crowdsensing: challenges and opportunities," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 161–167, 2016.

[4] H. Gao, C. H. Liu, W. Wang, J. Zhao, Z. Song, X. Su, J. Crowcroft, and K. K. Leung, "A survey of incentive mechanisms for participatory sensing," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 918–943, 2015.

[5] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," in *Proc. IEEE INFOCOM'15*. IEEE, 2015, pp. 2542–2550.

[6] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Computing Surveys*, vol. 48, no. 1, p. 7, 2015.

[7] H. Jin, L. Su, and K. Nahrstedt, "Theseus: Incentivizing truth discovery in mobile crowd sensing systems," in *Proc. ACM MobiHoc'17*, 2017, p. 1.

[8] J. Dong, M. Noreikis, Y. Xiao, and A. Ylä-Jääski, "Vinav: A vision-based indoor navigation system for smartphones," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1461–1475, 2018.

[9] S. Zhou, L. Xu, Y. Liu, L. Yang, and K. Li, "A distributed compressive data gathering framework for mobile crowdsensing," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[10] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, 2018.

[11] T. Li, T. Jung, H. Li, L. Cao, W. Wang, X.-Y. Li, and Y. Wang, "Scalable privacy-preserving participant selection in mobile crowd sensing," in *Proc. IEEE PerCom'17*. IEEE, 2017, pp. 59–68.

[12] B. Guo, Y. Liu, W. Wu, Z. Yu, and Q. Han, "Activecrowd: A framework for optimized multitask allocation in mobile crowdsensing systems," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 392–403, 2017.

[13] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, 2017.

[14] X. Wang, W. Wu, and D. Qi, "Mobility-aware participant recruitment for vehicle-based mobile crowdsensing," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4415–4426, 2018.

[15] Y. Yang, W. Liu, E. Wang, and J. Wu, "A prediction-based user selection framework for heterogeneous mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2460–2473, 2018.

[16] J. Nie, J. Luo, Z. Xiong, D. Niyato, P. Wang, and M. Guizani, "An incentive mechanism design for socially aware crowdsensing services with incomplete information," *IEEE Communications Magazine*, vol. 57, no. 4, pp. 74–80, 2019.

[17] C. Xu, Y. Si, L. Zhu, C. Zhang, K. Sharif, and C. Zhang, "Pay as how you behave: A truthful incentive mechanism for mobile crowdsensing," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 053–10 063, 2019.

[18] D. Li, S. Wang, J. Liu, H. Liu, and S. Wen, "Crowdsensing from the perspective of behavioral economics: An incentive mechanism based on mental accounting," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9123–9139, 2019.

[19] D. Li, L. Yang, J. Liu, and H. Liu, "Considering decoy effect and fairness preference: An incentive mechanism for crowdsensing," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[20] X. Chen, M. Liu, Y. Zhou, Z. Li, S. Chen, and X. He, "A truthful incentive mechanism for online recruitment in mobile crowd sensing system," *Sensors*, vol. 17, no. 1, p. 79, 2017.

[21] G. Yang, S. He, Z. Shi, and J. Chen, "Promoting cooperation by the social incentive mechanism in mobile crowdsensing," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 86–92, 2017.

[22] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. K. Das, and T. Melodia, "Quality of information in mobile crowdsensing: Survey and research challenges," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 4, p. 34, 2017.

[23] J. Wang, Y. Wang, D. Zhang, F. Wang, H. Xiong, C. Chen, Q. Lv, and Z. Qiu, "Multi-task allocation in mobile crowd sensing with individual task quality assurance," *IEEE Transactions on Mobile Computing*, 2018.

[24] K. Ota, M. Dong, J. Gui, and A. Liu, "Quoin: Incentive mechanisms for crowd sensing networks," *IEEE Network*, vol. 32, no. 2, pp. 114–119, 2018.

[25] Z. Wang, R. Tan, J. Hu, J. Zhao, Q. Wang, F. Xia, and X. Niu, "Heterogeneous incentive mechanism for time-sensitive and location-dependent crowdsensing networks with random arrivals," *Computer Networks*, vol. 131, pp. 96–109, 2018.

[26] D. Peng, F. Wu, and G. Chen, "Data quality guided incentive mechanism design for crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 307–319, 2018.

[27] X. Gong and N. Shroff, "Incentivizing truthful data quality for quality-aware mobile data crowdsourcing," in *Proc. ACM MobiHoc'18*. ACM, 2018, pp. 161–170.

[28] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[29] D. Tao, S. Zhong, and H. Luo, "Staged incentive and punishment mechanism for mobile crowd sensing," *Sensors*, vol. 18, no. 7, p. 2391, 2018.

[30] K. Yi, R. Du, L. Liu, Q. Chen, and K. Gao, "Fast participant recruitment algorithm for large-scale vehicle-based mobile crowd sensing," *Pervasive and Mobile Computing*, vol. 38, pp. 188–199, 2017.

[31] Y. Chen, P. Lv, D. Guo, T. Zhou, and M. Xu, "Trajectory segment selection with limited budget in mobile crowd sensing," *Pervasive and Mobile Computing*, vol. 40, pp. 123–138, 2017.

[32] C. Zhao, S. Yang, P. Yan, Q. Yang, X. Yang, and J. McCann, "Data quality guarantee for credible caching device selection in mobile crowd-

sensing systems," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 58–64, 2018.

[33] D. Wu, S. Si, S. Wu, and R. Wang, "Dynamic trust relationships aware data privacy protection in mobile crowd-sensing," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2958–2970, 2018.

[34] J. Wang, J. Tang, G. Xue, and D. Yang, "Towards energy-efficient task scheduling on smartphones in mobile crowd sensing systems," *Computer Networks*, vol. 115, pp. 100–109, 2017.

[35] J. Wang, Y. Wang, D. Zhang, and S. Helal, "Energy saving techniques in mobile crowd sensing: Current state and future opportunities," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 164–169, 2018.

[36] W. Gong, B. Zhang, and C. Li, "Task assignment in mobile crowd-sensing: Present and future directions," *IEEE Network*, no. 99, pp. 1–8, 2018.

[37] R. Zhou, Z. Li, and C. Wu, "A truthful online mechanism for location-aware tasks in mobile crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 8, pp. 1737–1749, 2018.

[38] R. Azzam, R. Mizouni, H. Otrok, S. Singh, and A. Ouali, "A stability-based group recruitment system for continuous mobile crowd sensing," *Computer Communications*, vol. 119, pp. 1–14, 2018.

[39] J. An, D. Liang, X. Gui, H. Yang, R. Gui, and X. He, "Crowdsensing quality control and grading evaluation based on a two-consensus blockchain," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 4711–4718, 2018.

[40] F. Restuccia, P. Ferraro, S. Silvestri, S. K. Das, and G. L. Re, "Incentme: effective mechanism design to stimulate crowdsensing participants with uncertain mobility," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1571–1584, 2018.

[41] S. Xu, X. Chen, X. Pi, C. Joe-Wong, P. Zhang, and H. Y. Noh, "ilocus: Incentivizing vehicle mobility to optimize sensing distribution in crowd sensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[42] X. Yan, F. Ye, Y. Yang, X. Deng, and D. Wang, "Incentive facilitation for peer data exchange in crowdsensing," *IEEE Transactions on Cloud Computing*, pp. 1–1, 2019.

[43] G. Fan, H. Jin, Q. Liu, W. Qin, X. Gan, H. Long, L. Fu, and X. Wang, "Joint scheduling and incentive mechanism for spatio-temporal vehicular crowd sensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[44] M. Pouryazdan, B. Kantarci, T. Soyata, L. Foschini, and H. Song, "Quantifying user reputation scores, data trustworthiness, and user incentives in mobile crowd-sensing," *IEEE Access*, vol. 5, pp. 1382–1397, 2017.

[45] H. Jin, B. He, L. Su, K. Nahrstedt, and X. Wang, "Data-driven pricing for sensing effort elicitation in mobile crowd sensing systems," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 2208–2221, 2019.

[46] Z. Zhou, J. Feng, B. Gu, B. Ai, S. Mumtaz, J. Rodriguez, and M. Guizani, "When mobile crowd sensing meets uav: Energy-efficient task assignment and route planning," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5526–5538, 2018.

[47] G. Sun, Y. Wang, X. Ding, and R. Hu, "Cost-fair task allocation in mobile crowd sensing with probabilistic users," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[48] H. Wang, D. Zhao, H. Ma, and L. Ding, "Mb-gvns: Memetic based bidirectional general variable neighborhood search for time-sensitive task allocation in mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2019.

[49] W. Liu, Y. Yang, E. Wang, and J. Wu, "User recruitment for enhancing data inference accuracy in sparse mobile crowdsensing," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[50] Y. Wu, F. Li, L. Ma, Y. Xie, T. Li, and Y. Wang, "A context-aware multiarmed bandit incentive mechanism for mobile crowd sensing systems," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7648–7658, 2019.

[51] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD dataset roma/taxi (v. 2014-07-17)," Downloaded from http://crawdad.org/roma/taxi/20140717, Jul. 2014.