Dynamic User Recruitment with Truthful Pricing for Mobile CrowdSensing

Wenbin Liu[†], Yongjian Yang[†], En Wang[†,*], and Jie Wu[‡]
[†]College of Computer Science and Technology, Jilin University, P.R. China
[‡]Department of Computer and Information Sciences, Temple University, Philadelphia, USA
Email: liuwb16@mails.jlu.edu.cn, yyj@jlu.edu.cn, wangen0310@126.com, jiewu@temple.edu

Abstract*—Mobile CrowdSensing (MCS) is a promising paradigm that recruits users to cooperatively perform various sensing tasks. In most realistic scenarios, users dynamically participate in MCS, and hence, we should recruit them in an online manner. In general, we prefer to recruit a user who can make the maximum contribution at the least cost, especially when the recruitment budget is limited. The existing strategies usually formulate the user recruitment as the budgeted optimal stopping problem, while we argue that not only the budget but also the time constraints can greatly influence the recruitment performance. For example, if we have less remaining budget but plenty of time, we should recruit users with more patience. In this paper, we propose a dynamic user recruitment strategy with truthful pricing to address the online recruitment problem under the budget and time constraints. To deal with the two constraints, we first estimate the number of users to be recruited and then recruit them in segments. Furthermore, to correct estimation errors and utilize newly obtained information, we dynamically re-adjust the recruiting strategy and also prove that the proposed strategy achieves a competitive ratio of $(1-1/e)^2/7$. Finally, a reverse auction-based online pricing mechanism is lightly built into the proposed user recruitment strategy, which achieves truthfulness and individual rationality. Extensive experiments on three real-world data sets validate the proposed online user recruitment strategy, which can effectively improve the number of completed tasks under the budget and time constraints.

***Index Terms*—Mobile CrowdSensing, online user recruitment, submodular secretary problem, truthful pricing.**

I. Introduction

With the increasing popularity of portable devices, Mobile CrowdSensing (MCS) [1] has recently become a promising paradigm for recruiting users to cooperatively perform various sensing tasks [2]–[4], such as the monitoring of environment, traffic, and urban infrastructure. In most cases, we should provide rewards for the recruited users, in order to cover the sensing costs and encourage user participation [5]–[7]. However, due to the budget constraint, we have to select some effective users, which raises the fundamental user recruitment problem in MCS.

Most of the existing user recruitment strategies are conducted offline [8]–[10]. As shown in Fig. 1 (upper part), the offline method recruits users from a pre-determined pool at the beginning of the MCS campaign. However, in most realistic scenarios, users may dynamically participate in MCS and we should recruit them online. Fig. 1 (lower part) shows such

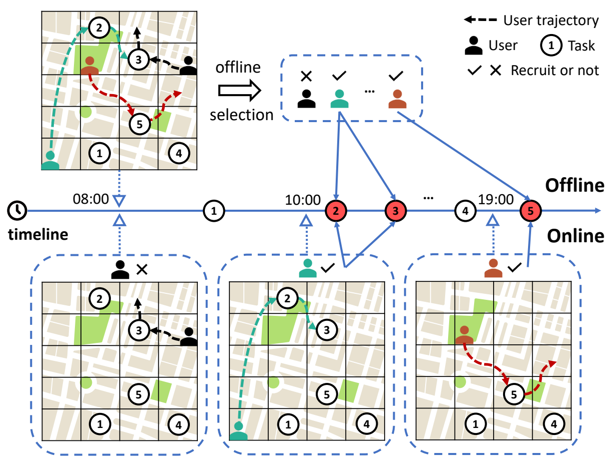


Fig. 1: Offline vs. online user recruitment in MCS.

an online scenario, where tasks are distributed in different areas and users unconsciously move among the sensing areas to perform the tasks. We prefer to recruit a user who can make the maximum contribution at the least cost. However, in the online cases, users' costs and contributions are invisible until they participate in MCS. Hence, deciding whether to recruit the current user is more challenging in the online scenario, especially when the total recruitment budget is limited. To this end, some existing works formulate the online user recruitment as the optimal stopping problem and utilize the dynamic programming [11] or secretary problem [12] to deal with it, but they ignore the budget constraint. Some researchers further consider the budget constraint [13]–[15], but ignore the influence of the remaining time on user recruitment. We study a similar online scenario, but consider not only the budget but also the time constraints. The two constraints seem to be independent but jointly affect the online user recruitment. For example, when there is little time left, we prefer to recruit all participating users, in order to use up the budget as soon as possible. Similarly, if we have less remaining budget but plenty of time, we should recruit users with more patience. Hence, how to address *the budget and time constraints* in online user recruitment is the first challenge. Moreover, the dynamical participation introduces a lot of uncertainty, especially regarding the user's mobility, cost and participating rate. Therefore, the second challenge is how we can *dynamically re-adjust* our online user recruitment strategy

* En Wang is the corresponding author.

along with the online recruiting process. Finally, in order to encourage user participation and also avoid being deceived, we should determine a *truthful price* for each recruited user in this online manner.

To deal with the budget and time constraints, we assume that the distribution of all users' participating time is periodic and can be learned from historical data. Thus, we first estimate the number of all participating users according to the time constraint. Then, taking the budget constraint into consideration, we further estimate the number of users to be recruited. In this way, the online user recruitment problem can be naturally interpreted as 'recruit k out of all participating users to maximize the total contributions', which is actually a classic k-secretaries problem. Note that the tasks only need to be completed once. After a task has been completed by the previous recruited users, it becomes invisible to the later users, and hence, the recruited users are actually expected to have diminishing contributions. Thus, we further extend the k-secretaries problem with a submodular contribution function [16]. We approximately divide the real participating users into k equally-sized segments and try to recruit the best user in each segment, while for different segments, we recruit users according to their submodular marginal contributions.

Note that we could not accurately calculate a user's contribution (estimated by the user's coverage of tasks based on his uncertain mobility) and cost (randomly claimed from an independent cost distribution). Moreover, although the participation is assumed to be periodical, its period still needs to be learned through historical data. All these uncertain factors make the estimated numbers not always precise. In order to correct the errors in estimation and make use of the new information obtained during this online process, we further present a dynamic user recruitment strategy. The basic idea is to conduct a re-estimation according to the remaining budget and time after recruiting a new user, which is actually a dynamic iteration of estimation and online user recruitment in the above k-segments strategy. Furthermore, we prove that the proposed dynamic online recruitment strategy achieves a competitive ratio of $(1-1/e)^2/7$.

Finally, we also conduct a reverse auction-based pricing mechanism, as a supplement to the dynamic user recruitment strategy. This pricing mechanism determines the payment for each recruited user and also satisfies the budget and time constraints. We can easily build this mechanism into our online user recruitment strategies without much extra computation. In addition, this online pricing mechanism is proved to achieve truthfulness [17] and individual rationality.

In summary, this paper makes the following contributions:

- *Dynamic Online User Recruitment*: We study the online user recruitment problem with the budget and time constraints. To deal with the two constraints, we first estimate the number of users to be recruited and then propose a segmented online user recruitment strategy. Furthermore, a dynamic re-estimation is presented to correct the estimation errors and utilize the newly obtained information, where the competitive ratio is proved to be $(1-1/e)^2/7$.
- *Reverse Auction-based Online Pricing*: We present a reverse auction-based pricing mechanism, which can be built into the online user recruitment strategy without much extra computation. Meanwhile, this mechanism achieves truthfulness and individual rationality.
- *Extensive Evaluation*: We conduct an extensive evaluation based on three real-world data sets. The results verify the effectiveness of our strategy on improving the number of completed tasks under the budget and time constraints.

II. Related Work

Mobile CrowdSensing is a promising paradigm, which allows us to recruit users carrying portable devices, in order to cooperatively perform various sensing tasks [18]–[20]. Considering the sensing costs, Karaliopoulos *et al.* [21], Zhang *et al.* [22], Song *et al.* [23], and Wang *et al.* [24] study the user recruitment problem to achieve the goal of the MCS campaigns and minimize the total costs. Similarly, Liu *et al.* [8] and Wang *et al.* [9] propose the prediction-based algorithms to recruit the effective users, in order to complete more tasks under a budget constraint. However, most of the existing user recruitment strategies are conducted offline and cannot deal with the users' dynamic participation, which is actually a more realistic online scenario.

Recently, the user recruitment problem has been studied for the online scenarios. Wang *et al.* [25] study the location-aware and location diversity based online MCS but focus on the task assignment. Li *et al.* [26] propose a dynamic user selection algorithm but divide the online recruiting process into many time slots and greedily recruit users for each time slot in an offline manner. Yang *et al.* [12] present a prediction-based online user selection framework, however, they only recruit a pre-determined number of users and ignore the variable costs of users under the budget constraint. Zhao *et al.* [13], Gao *et al.* [14], and Li *et al.* [15] further consider the budget constraint in online incentive mechanisms and user selection. These methods divide the total budget into some stages and recruit users until the sub-budget in each stage is exhausted, however, they haven't dealt with the total budget and ignore the influence of the remaining time of the MCS campaigns.

For the secretary recruitment, the classic secretary problem is to recruit only one best user from all participating users in an online manner [27]. As a variant, Preater [28] studies that the more than one user may be recruited in the secretary problem. Considering the submodular utility function, Bateni *et al.* [16] propose the submodular k-secretaries problem, where they divided the participating users into the fixed k equally-sized segments and select the best user from each segment.

III. System Model and Problem Formulation

A. System Model

We first discuss the system model of online user recruitment under the budget and time constraints in this section and the main notations are listed in Table I. We consider a practical online scenario of MCS, where a crowd of rational users move around and participate in the MCS campaign in real

TABLE I: Main notations

Notation	Meaning
u, c, $[A^b, A^e]$	User, cost, and active time of users.
s, l, $[T^b, T^e]$	Task, location, and duration time of MCS.
p, B	Payment and budget.
S, U, μ	Set of tasks, users and recruited users.
m, n, k	Number of tasks, users and recruited users.
$Z_u(l_i, l_j, T)$, $Q_u(l_i, l_j, T)$	Probability that u moves from l_i to l_j within time T, and just at time T.
$P(u_i, s_j)$	Probability that u_i will complete s_j.
$E(\mu, s_j)$	Expected probability that μ will complete s_j.

time to perform the sensing tasks. Users are denoted as $U \triangleq \{u_1, u_2, ..., u_n\}$, each with an active time (working time) $[A_i^b, A_i^e]$ and sensing cost c_i, which indicates that user u_i will work from A_i^b to A_i^e with cost c_i[1]. Tasks are denoted as $S \triangleq \{s_1, s_2, ..., s_m\}$ each with a location l_j, which indicates that a user u_i moving to location l_j within his active time $[A_i^b, A_i^e]$ can perform the task s_j. Under the online scenarios, users are participating in real time and we decide whether to recruit them immediately, with the payments $p_i \geq c_i$ under a limited budget B. Then, the recruited users, denoted as μ with the set cardinality $k = |\mu|$, perform the sensing tasks within the duration time of the MCS campaign $[T^b, T^e]$.

We assume that all tasks are equal in quality and only need to be completed once[2]. We consider that the tasks are uniformly distributed and the active time of users is far less than the total time of the MCS campaign, otherwise the users participating later have great disadvantages and we would better recruit the earlier users. Actually, this setting is reasonable for most practical purposes, since users won't work for a long time for the MCS campaigns. Similarly, users won't wait for the recruitment decisions for a long time, and thus we need to decide whether to recruit them immediately, without knowing the future. After receiving the decisions, users will leave and their next participation will be seen as the new ones.

B. Mobility Prediction

From the opportunistic perspective, when a recruited user reaches the location of one task within his active time, we consider that the task can be completed successfully. We estimate the user's contribution according to his mobility prediction[3]. For the prediction, as shown in Fig. 2, we divide the full map into some grids. Tasks are distributed in the grids and users reaching one grid can complete the tasks in this grid. Then, we use a modified Semi-Markov Process Model [8], [9], [12] to predict the time-dependent transition probabilities between the grids as the user's mobility prediction. In order to further reduce the great amount of calculation, for each gird, we only consider its transitions between the nearby grids, *i.e.*, up, down, left, right and itself. The time-dependent semi-Markov kernel $Z_u(l_i, l_j, T)$, *i.e.*, the probability that user u will move from his current grid l_i to his next grid l_j within time T, is defined by Eq. (1).

[1]Resource consumption, risk compensation and other costs.

[2]Different settings can be easily added at the contribution function of users.

[3]Other measures of contribution could be modified easily, in order to judge whether the user is good or not.

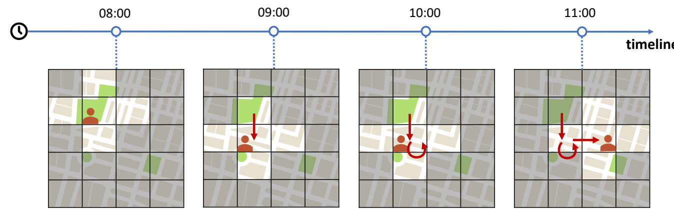


Fig. 2: An example of mobility prediction model.

$$Z_u(l_i, l_j, T) = Z(L_u^{n+1} = l_j, t_u^{n+1} - t_u^n \leq T | L_u^n = l_i), \quad (1)$$

where L_u indicates the user's moving sequence of grids and t_u is the arrival time. Note that the user's next grid is associated with his current grid and we can derive the probability Z from the statistical results of the user's history records. Then, we obtain another kernel $Q_u(l_i, l_j, T)$, *i.e.*, the probability that user u will move from the grid l_i to l_j just at the time T, denoted by Eq. (2).

$$Q_u(l_i, l_j, T) = \begin{cases} \Sigma_{l_k}^{L_u} \Sigma_{t=1}^{T} (Z_u(l_i, l_k, t) - Z_u(l_i, l_k, t-1)) \cdot \\ Q_u(l_k, l_j, T-t), \qquad l_i \neq l_j \\ 1 - \Sigma_{l_k, l_k \neq l_i}^{L_u} (Z_u(l_i, l_k, T) - \\ \Sigma_{t=1}^{T} (Z_u(l_i, l_k, t) - Z_u(l_i, l_k, t-1)) \cdot \\ Q_u(l_k, l_i, T-t)), \qquad l_i = l_j \end{cases} \quad (2)$$

where $Q_u(l_i, l_i, 0) = 1$ and $Q_u(l_i, l_j, 0) = 0$, if $l_i \neq l_j$. Specifically, when $l_i \neq l_j$, we consider the relay state transitions as $l_i \rightarrow l_k \rightarrow l_j$ and calculate the total probability. When $l_i = l_j$, we further consider the probability that users stay at the same grid. With the $Q_u(l_i, l_j, T)$ from mobility prediction, we obtain the probability that user u_i can complete task s_j, and finally calculate the expected contribution of the recruited user set for each task, as follows:

$$P(u_i, s_j) = 1 - \Pi_{t=A_i^b}^{A_i^e} \big(1 - Q_{u_i}(l_{u_i}, l_{s_j}, t)\big), \quad (3)$$

$$E(s_j, \mu) = 1 - \Pi_{u_i \in \mu} \big(1 - P(u_i, s_j)\big). \quad (4)$$

C. Problem Formulation

Problem [Online User Recruitment under the Budget and Time Constraints]: *Given a set of MCS tasks, with a limited budget and the duration time of the MCS campaign, we recruit a set of sequential participating users who move around to perform sensing tasks, with the objective of maximizing the expected number of completed tasks*:

$$\text{maximize} \quad \Sigma_{s_j \in S} E(s_j, \mu) \quad (5)$$

$$\text{subject to} \quad \mu \subseteq U, \ \Sigma_{u_i \in \mu} p_i \leq B, \ T^b \leq t \leq T^e \quad (6)$$

A running example shown in Fig. 3 provides an intuitive interpretation of our online user recruitment problem. Considering that there are three users moving around the 5×4 grids. They will participate in the MCS campaign in real time, and we can only recruit two of them under the budget and time constraints. At 8:00, user 1 participates and we predict that he will reach the location of task 3 within his active time. However, user 1 may perform task 3 but cannot perform

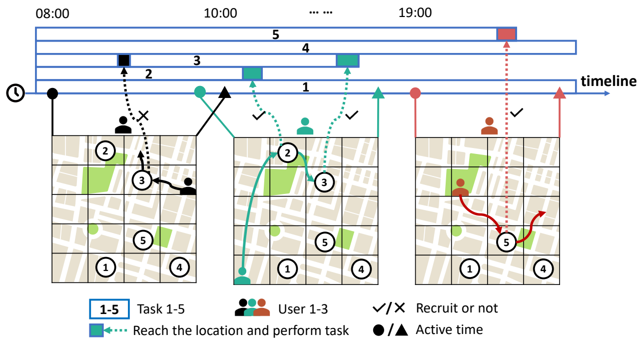


Fig. 3: An example of online user recruitment in MCS.

any other tasks. His contribution seems relatively less and we decide to keep waiting since we have enough time. When user 2 connects to server, we find that he can perform tasks 2 and 3, which contributes a lot and thus we recruit him. After user 2 completes tasks 2 and 3, we then drop these tasks. When user 3 connects to server, although he may perform only one task but we have less remaining time, and hence, we recruit him. Finally, we recruit users 2 and 3 in an online manner, and they move around the sensing areas and complete the most tasks under the budget and time constraints.

IV. Online User Recruitment under Budget and Time Constraints

A. Problem Hardness

Before prescribing an online strategy, we first prove that the online user recruitment problem under the budget and time constraints is NP-hard, as shown in the following theorem.

Theorem 1. *The online user recruitment problem under the budget and time constraints is NP-hard.*

Proof. Without loss of generality, we ignore the mobility prediction but consider the pre-determined traces. Here, the completed tasks by user u_i are denoted as S_{u_i} and the total completed tasks by the recruited user set μ is $\cup_{u_i \in \mu} S_{u_i}$. Further considering a special case that all users cost equally, $i.e.$, under a budget B and the user cost c, we could recruit $k = \lfloor B/c \rfloor$ users at most, which is indeed a classic NP problem, *Max k-cover* [29]: given a collection of task sets $\{S_{u_1}, S_{u_2}, ..., S_{u_n}\}$, each will cover several tasks $S_{u_i} = \{s_{i1}, s_{i2}, ...\}$, then the objective is to select k sub-collections to cover the most tasks. That is to say, the special case is NP-hard. Consequently, further considering the budget and time constraints, the online user recruitment problem is NP-hard. The theorem holds. □

In this paper, we use a mobility prediction model to estimate the users' coverage of tasks as their contribution. To simplify the notation, we use $f(\mu) = \Sigma_{s_j \in S} E(s_j, \mu)$ as the predicted contribution of the recruited user set μ, which has the following property:

Theorem 2. *1) $f(\emptyset) = 0$; 2) $f(\mu)$ is non-decreasing; 3) $f(\mu)$ is submodular.*

Proof. 1) $\mu = \emptyset$ means that we don't recruit users and thus no tasks can be completed, $i.e.$, $E(s_j, \emptyset) = 0$ for each $s_j \in S$, according to Eq. (4). Therefore, $f(\emptyset) = \Sigma_{s_j \in S} E(s_j, \emptyset) = 0$.

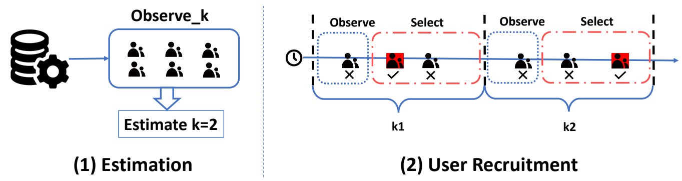


Fig. 4: Segmented online user recruitment strategy.

2) Consider two user subsets μ_1 and μ_2, and $\mu_1 \subseteq \mu_2$. According to Eq. (4), we obtain the following inequation:

$$\begin{aligned} & E(s_j, \mu_1) - E(s_j, \mu_2) \\ = & \Pi_{u_i \in \mu_2}(1 - P(u_i, s_j)) - \Pi_{u_i \in \mu_1}(1 - P(u_i, s_j)) \qquad (7) \\ = & \Pi_{u_i \in \mu_1}(1 - P(u_i, s_j)) \cdot (\Pi_{u_i \in \mu_2 \backslash \mu_1}(1 - P(u_i, s_j)) - 1) \leq 0. \end{aligned}$$

Since $P(u_i, s_j)$ is the probability that user u_i can complete task s_j, according to Eq. (4), we have $0 \leq P(u_i, s_j) \leq 1$. Thus, $f(\mu_1) - f(\mu_2) = \Sigma_{s_j \in S}(E(s_j, \mu_1) - E(s_j, \mu_2)) \leq 0$ and $f(\mu)$ is non-decreasing. 3) Similar to 2), we consider an arbitrary user $u_k \in U \backslash \mu_2$, and obtain the following inequation:

$$\begin{aligned} & (f(\mu_1 \cup \{u_k\}) - f(\mu_1)) - (f(\mu_2 \cup \{u_k\}) - f(\mu_2)) \\ = & \Sigma_{s_j \in S}(E(s_j, \mu_1 \cup \{u_k\}) - E(s_j, \mu_1)) - \\ & \Sigma_{s_j \in S}(E(s_j, \mu_2 \cup \{u_k\}) - E(s_j, \mu_2)) \\ = & \Sigma_{s_j \in S}(\Pi_{u_i \in \mu_1}(1 - P(u_i, s_j)) \cdot P(u_k, s_j) - \\ & \Pi_{u_i \in \mu_2}(1 - P(u_i, s_j)) \cdot P(u_k, s_j)) \\ = & \Sigma_{s_j \in S} \Pi_{u_i \in \mu_1}(1 - P(u_i, s_j)) \cdot P(u_k, s_j) \cdot \\ & (1 - \Pi_{u_i \in \mu_2 \backslash \mu_1}(1 - P(u_i, s_j))) \geq 0. \qquad (8) \end{aligned}$$

As discussed above, we know that $0 \leq P(u_i, s_j) \leq 1$, and thus $f(\mu_1 \cup \{u_k\}) - f(\mu_1) \geq f(\mu_2 \cup \{u_k\}) - f(\mu_2)$, which holds the submodular property of $f(\mu)$. □

B. Segmented Online User Recruitment Strategy

In online scenarios, all the users participate in real time and they form a sequence according to their participating time. We should make an immediate decision on whether to recruit the current participating user according to his predicted contribution and cost, without knowing the future users. Further considering the budget and time constraints, the online user recruitment problem becomes more challenging. In order to deal with the two constraints, we present a segmented online user recruitment strategy, which first estimates the number of users to be recruited and then segmentally recruits them in an online manner, as shown in Fig. 4.

1) Estimation via Submodular Maximization with Knapsack Constraint: In the online user recruitment problem, the biggest difficulty is the unknown future information, especially when we need to deal with the constraints and online recruiting simultaneously. In order to reduce the difficulty, we first make an assumption to deal with the budget and time constraints before the online recruiting, as shown in Assumption 1.

Assumption 1. *The distribution of user participating time is periodical and the users have an independent cost distribution.*

In many scenarios where humans are involved, Assumption 1 is common and reasonable, such as people's check-in records of an App, the numbers of which within the same periods in

Algorithm 1 Greedy Offline *Estimation()*

Input: S, B, $T = [t^s, t^e]$, U_h

1: Estimate n' and conduct $U' = \{u'_1, u'_2, ..., u'_n\}$ from U_h according to the time T;
2: Initialize $\mu = \emptyset$;
3: **while** $U' \neq \emptyset$ **do**
4:     Calculate $\delta_{u_i} = \frac{f(\mu \cup \{u_i\}) - f(\mu)}{c_i}, \forall u_i \in U'$;
5:     Find $u^* = \arg\max_{u_i \in U'} \delta_{u_i}$;
6:     **if** $\sum_{u_j \in \mu} c_j + c_i \leq B$ **then** $\mu \leftarrow \mu \cup u_i$;
7:     $U' \leftarrow U' \setminus u^*$;
return n', $k = |\mu|$

Algorithm 2 Online User Recruitment *Segmented()*

Input: S, B, $U = \{u_1, u_2, ..., u_n\}$, n', k, $\mu = \emptyset$

1: Initialize $l = \lfloor n'/k \rfloor$, $l_{ob} = \lfloor l/e \rfloor$ and $\varepsilon = 0$;
2: **while** $U \neq \emptyset$ **do**
3:     Wait for the next user u_i's coming, $U \leftarrow U \setminus u_i$;
4:     **if** $i > n'$ **and** $\sum_{u_j \in \mu} c_j + c_i \leq B$ **then** $\mu \leftarrow \mu \cup \{u_i\}$;
5:     **else**
6:         Initialize *segmentID* $= i/l$;
7:         Calculate $\delta_{u_i} = (f(\mu \cup \{u_i\}) - f(\mu))/c_i$;
8:         **if** $i \leq$ *segmentID* $* l + l_{ob}$ **then**
9:             $\varepsilon = max\{\varepsilon, \delta_{u_i}\}$; ▷ `Observe`
10:         **else if** $i >$ *segmentID* $* l + l_{ob}$ **and** $\delta_{u_i} \geq \varepsilon$ **then**
11:             **if** $\sum_{u_j \in \mu} c_j + c_i \leq B$ **then**
12:                 $\mu \leftarrow \mu \cup \{u_i\}$; ▷ `Recruit`
13:                 $i =$ *segmentID* $* l + 1$ and $\varepsilon = 0$;
14:                 **Continue**; ▷ **Break** at Algorithm 3
return μ

weekdays are roughly the same and independent from people. Under this assumption, we can explore the historical data to estimate the number of all participating users and the number of users to be recruited, in order to satisfy the budget and time constraints in the online user recruitment. The basic idea is to learn the periodical distribution to estimate the number of participating users, according to the remaining time of the MCS campaign. Then, we can conduct a simulated user set from historical data, which can actually be seen as a replacement of real participating users. With the simulated user set, we consider the user recruitment problem under the budget constraint as the submodular maximization with knapsack constraint [30] and propose a greedy offline estimation algorithm to estimate the number of users to be recruited.

The algorithm is summarized in Algorithm 1. We first estimate the number of participating users n' and conduct a simulated user set U' from historical data U_h, according to the time T (line 1). From the simulated user set, we greedily select the users who contribute most in an offline manner (line 3) and recruit the users under the budget constraint (lines 4-5). Finally, we obtain the estimated n' and k, which actually reflect the budget and time constraints and directly influence the online user recruitment. In this way, with the help of Assumption 1, we estimate the number of participating users and the number of users to be recruited from historical data before the real online recruiting process, in order to approximately deal with the budget and time constraints first.

2) User Recruitment via Submodular k-Secretaries Problem: With the estimated numbers n' and k, we formulate the online user recruitment problem as a variant of the famous secretary problem, $i.e.$, the submodular k-secretaries problem [16], without considering the budget and time constraints but focusing on the online recruiting process.

The classic secretary problem is to recruit the best one out of n participating users, where users are participating in sequence and the recruitment decisions should be made immediately. As a variant, the submodular k-secretaries problem presented by Bateni $et\ al.$ [16] further considers the multiple recruited users with submodular utility function, which provides an appropriate solution for our online user recruitment problem with a submodular contribution function. Without considering the budget and time constraints, our online user recruitment can be naturally formulated as a submodular k-secretaries problem, interpreted as 'recruit k out of n users to maximize the expected number of completed tasks'. Then, we propose a k-segments online user recruitment algorithm to deal with the problem, as summarized in Algorithm 2.

According to the estimated n' and k, we approximately partition the real participating users into k equally-sized segments and select the best user from each segment. If the real number of participating users n is larger than our estimated n', we will continue to recruit them until the budget is exhausted (line 3). For each segment, under the budget constraint, we observe the first l_{ob} users, record the largest contribution/cost ratio as a threshold ε (line 9) and select the next one who has a larger δ_{u_i} than ε (line 10). Finally, we obtain the recruited user set μ in an online manner from the real participating users.

C. Dynamic Online User Recruitment Strategy

By exploiting the historical data, our proposed segmented strategy can deal with the online user recruitment problem under the budget and time constraints well. However, due to the users' dynamical participation, we could not accurately calculate a user's contribution and cost. Moreover, although the participation is assumed to be periodical, its period still needs to be learned through historical data. All these uncertain factors make the estimated numbers not always precise. In order to correct the errors in estimation and make use of the new information obtained during the online recruiting process, we further propose a dynamic user recruitment strategy extended from the Algorithms 1 and 2, $i.e.$, the *Estimation()* and *Segmented()* mentioned above.

The basic idea of the dynamic user recruitment strategy is to conduct a re-estimation after recruiting a new user, which actually can be seen as a dynamic iteration of *Estimation()* and *Segmented()*. Fig. 5 illustrates a straightforward example, where the dynamic strategy first estimates the numbers $n' = 6$ and $k = 2$ from historical data, and partitions the online users into 2 equal-sized segments for online recruiting (Round 1).

Algorithm 3 Dynamic Online User Recruitment

Input: S, B, $T = [t^s, t^e]$, $U\{u_1, u_2, ..., u_n\}$

1: $\mu = \emptyset$, $B_{rest} = B$, $S_{rest} = S$, $U_{rest} = U$, $T_{rest} = [t^s, t^e]$;
2: **while** $B_{rest} > 0$ **and** $U \neq \emptyset$ **do**
3: $\quad n', k \leftarrow$ *Estimation(*S_{rest}, B_{rest}, T_{rest}*)*;
4: $\quad\quad \triangleright$ **If** $k = 0$ or $n > n'$, recruit users under B_{rest}
5: $\quad \mu \leftarrow$ *Segmented(*S_{rest}, B_{rest}, U_{rest}, n', k, μ*)*;
6: $\quad\quad \triangleright$ **Break** after a new user has been recruited
7: $\quad$ Update S_{rest}, U_{rest}, B_{rest}, T_{rest};
return μ

After one user has been recruited, the dynamic strategy then re-estimates the remaining numbers of participating users $n' = 4$ and recruited users $k = 2$, according to the currently obtained information, $i.e.$, the remaining tasks, users, budget and time. The formal dynamic strategy is provided in Algorithm 3. We iteratively run *Estimation()* and *Segmented()* for each recruitment (lines 3-6), until the budget is exhausted or the MCS campaign is finished (line 2). Specifically, *Segmented()* will **break** after one user has been recruited (line 5 in Algorithm 3 and line 12 in Algorithm 2), and then we update the current information for re-estimation. For the special case, $e.g.$, $k = 0$ or $n > n'$, we will recruit the remaining users who can satisfy the budget constraint. In this way, our proposed dynamic strategy can make use of the newly obtained information and correct the estimation errors constantly, and better solve the online user recruitment under the budget and time constraints.

D. Competitive Ratio for the Algorithms

We first give the competitive ratio of the segmented online user recruitment strategy in the following theorem.

Theorem 3. *The segmented online user recruitment strategy approximately achieves a competitive ratio of* $\frac{(1-1/e)^2}{7}$.

Proof. 1) We first relax the online user recruitment to an offline scenario, where all of the users and tasks are pre-determined and we select a user set in a totally offline manner. As proved in Theorem 2, our objective function $f(\mu)$ is non-decreasing and submodular. Then, the user selection problem is formulated as a variant of submodular maximization problem with a knapsack constraint, and our offline greedy algorithm can achieve a $(1 - 1/e)$ approximation of the optimal value[4], denoted as $f(\mu\_U) \geq (1 - 1/e)f(OPT\_U)$, where OPT_U is the optimal user set in U and μ_U is the greedily selected user set with cardinality $k = |\mu\_U|$. 2) Similarly, we consider the online user recruitment problem as a submodular k-secretaries problem, where $f(\mu)$ is proved as a non-decreasing submodular function and the users are participating in real time. Under the online scenarios, our segmented user recruitment strategy can be proved to achieve an expected competitive ratio of $(1-1/e)/7$ [16], denoted as $E\{f(\mu)\} \geq (1-1/e)/7 \cdot f(OPT\_k\_U)$, where OPT_k_U is the optimal set in U with cardinality k. 3) As discussed in 1) and 2), OPT_k_U is the optimal set and μ_U is the greedy result under the same cardinality k, and OPT_U is the global optimal set without cardinality constraints. Thus, we obtain the following inequation:

$$E\{f(\mu)\} \geq \frac{1-1/e}{7} f(OPT\_k\_U) \geq \frac{1-1/e}{7} f(\mu\_U) \geq \frac{(1-1/e)^2}{7} f(OPT\_U). \quad (9)$$

4) In the online scenario, we cannot exactly obtain the users who will participate in the MCS campaign in advance. Under Assumption 1, we conduct the simulated user set U' as a replacement of the real user set U, and greedily select μ_U' from U' to estimate μ_U. Therefore, we have $f(\mu\_U') \approx f(\mu\_U)$ and $E\{f(\mu)\}$ approximately achieves $(1-1/e)^2/7$ competitive ratio of the optimal $f(OPT\_U)$. $\square$

[4]The offline submodular maximization problem with knapsack constraint in *Estimation()* is NP-hard and has been proved that the greedy algorithm achieves a $(1 - 1/e)$ approximation combined with partial enumeration [30].

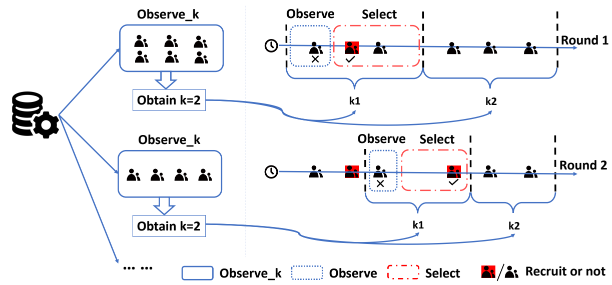


Fig. 5: Dynamic online user recruitment strategy.

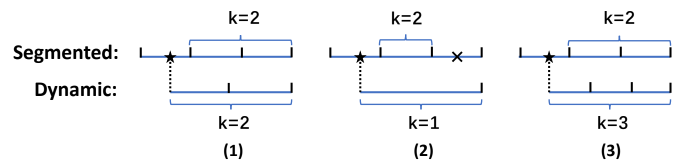


Fig. 6: Different dynamic and segmented cases.

Actually, the dynamic user recruitment strategy is an extension of the segmented strategy, which can correct the errors and make use of new information during the online recruiting process. Thus, the dynamic strategy can outperform the segmented strategy in expectation. The proof is simple and we provide some intuitive examples in Fig. 6: after one user has been recruited, if the estimated k in the dynamic strategy is the same as the one in the segmented strategy, the dynamic strategy has more participating users, since it doesn't need to skip over some users to the next segment like the segmented strategy. Thus, the dynamic strategy will expectedly outperform the segmented strategy. Similarly, if the estimated k in the dynamic strategy is different with the one in the segmented strategy, it means that the previous recruited users in segmented strategy cost too much/little. However, the segmented strategy cannot correct the errors and make use of new information in time, which leads to a worse performance.

V. Reverse Auction-based Truthful Pricing for Online User Recruitment

A. Reverse Auction-based Online Pricing Mechanism

In general, the organizers and users in MCS are rational and selfish. From the user side, the organizer should provide proper

Algorithm 4 Reverse Auction-based Pricing

Input: S, B, $U = \{u_1, u_2, ..., u_n\}$, n', k, $\mu = \emptyset$

In *Segmented(), u_i is coming*:

1: **if** $i > n'$ **and** $\sum_{u_j \in \mu} p_j + c_i \leq B$ **then**
2: Recruit u_i with pricing c_i;
3: **else if** $i >$ *segmentID* $* l + l_{ob}$ **and** $\delta_{u_i} \geq \varepsilon$ **then**
4: $p_i = c_i \cdot \delta_{u_i} / \varepsilon$;
5: **if** $\sum_{u_j \in \mu} p_j + p_i \leq B$ **then**
6: Recruit u_i with pricing p_i;

rewards for the recruited users to cover the sensing costs and encourage user participation. From the organizer side, the pricing mechanism also needs to ensure that users bid their costs truthfully, in order to pay less and earn more. Recently, the reverse auction has been used for pricing in MCS to simultaneously satisfy the truthfulness and individual rationality [5], [31], where users bid first according to their costs and then the organizers determine the payments. However, the existing mechanisms determine the prices for the recruited users by ordering them according to their contributions and costs in an offline manner, which can hardly be used in online recruitment, especially considering the budget and time constraints.

In our proposed dynamic and segmented strategies, the user recruitment in segments can actually be seen as an online ordering of users, and thus a reverse auction-based pricing mechanism can be easily modified, as summarized in Algorithm 4. We first deal with the special cases, $i.e.$, the real number of participating users n is larger than our estimated n', where we will recruit the extra users and only pay their claimed costs until B is exhausted (lines 1-2). For the user recruitment in each segment, we use the claimed costs observed from the first l_{ob} users and determine a price for the recruited user (lines 3-4), denoted as $p_i = c_i \cdot \delta_{u_i} / \varepsilon$. Note that the total payments (instead of costs) of recruited users are constrained by B, and thus we only recruit the users we can afford (lines 5-6). In this way, the pricing mechanism has been lightly built into the online user recruitment strategy without much extra computation, and the truthfulness and individual rationality will be proved next.

B. Truthfulness and Individual Rationality

Theorem 4. *The reverse auction-based online pricing mechanism is truthful.*

Proof. 1) We first prove that our proposed online user recruitment strategy is *bid-monotone*. Suppose that a user u_i has been recruited by the online strategy, we obtain $\delta_{u_i} = \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c_i$ and $\delta_{u_i} \geq \varepsilon$ according to Algorithms 2 and 4. If u_i bids a smaller cost $c'_i < c_i$, we have $\big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c'_i > \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c_i \geq \varepsilon$. With a smaller c'_i, the user's contribution/cost ratio is larger than the threshold ε. Thus, u_i will still be recruited by the proposed online strategy, which holds the bid-monotone. 2) Then, we prove that the payment determined by our proposed online pricing mechanism is the *critical value*, $i.e.$, if the recruited user u_i bids a larger cost c_i than the determined payment p_i, he won't be recruited, and otherwise, u_i will be recruited. According to Algorithms 2 and 4, we have $p_i = c_i \cdot \delta_{u_i} / \varepsilon = \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/\varepsilon$. Assume that user u_i bids a larger cost $c'_i > p_i$, and we obtain $\big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c'_i < \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/p_i = \varepsilon$. Since the user's contribution/cost ratio is less than ε, user u_i won't be recruited by our proposed strategy. Similarly, assume that u_i bids a smaller $c'_i \leq p_i$, u_i will still be recruited, $i.e.$, $\big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c'_i \geq \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/p_i = \varepsilon$. 3) Finally, with the proved *bid-monotone* and *critical value* in 1) and 2), according to Myerson theorem [17], the reverse auction-based online pricing mechanism is truthful. $\square$

Theorem 5. *The reverse auction-based online pricing mechanism achieves individual rationality.*

Proof. The individual rationality of the online pricing mechanism means that the reward that each recruited user gets should be no less than the cost. According to Algorithms 2 and 4, assume that u_i is one of the recruited users, we have $\delta_{u_i} \geq \varepsilon$ and $p_i = c_i \cdot \delta_{u_i} / \varepsilon$. Thus, we obtain $p_i / c_i = \delta_{u_i} / \varepsilon \geq 1$ and $p_i \geq c_i$. In the special cases, $i.e.$, the real n is larger than the estimated n', the rewards of the extra users are equal to their claimed costs. Therefore, the individual rationality of the online pricing mechanism is proved. $\square$

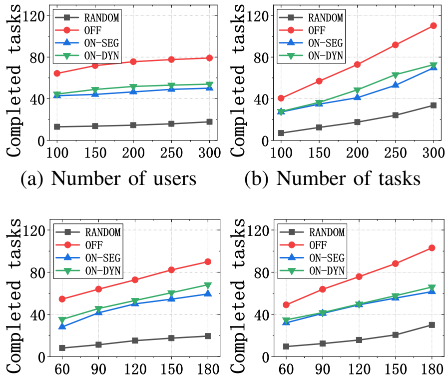
VI. Performance Evaluation

A. Data sets & Settings

The three real-world data sets are used for the evaluation:

- *Feeder* [32] contains four kinds of data, $i.e.$, the cellphone CDR data, smartcard data, taxicab GPS data, and bus GPS data collected from Shenzhen, China. We select 300 taxi traces as the participating users, each of which has the continuous GPS records collected from the same period of time, $i.e.$, 8:00-18:00, for two days.
- *Shanghai* contains the GPS data collected from taxis and trucks in Shanghai, China. Similar to *Feeder*, we select 310 traces as the participating users, each of which was collected from 8:00 to 18:00 in two days. Note that nearly half of them were collected from trucks, which have the more regular mobilities than the other trajectories.
- *GeoLife* [33], [34] was collected from phones carried by 182 users, which record a broad range of users' outdoor movements. It contains 17,000+ trajectories and has a total duration of 50,000+ hours, from which we select 727 traces with the same period of time. Compared with the other data sets, *GeoLife* has the fine-grained trajectories but users may stay at some places for a long time.

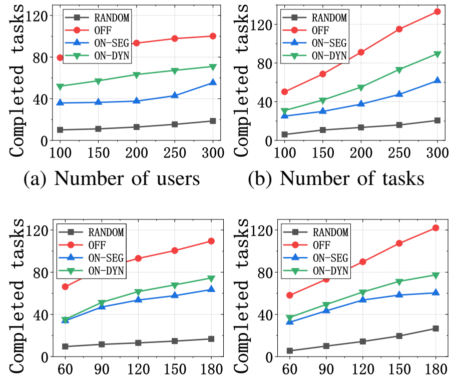
For mobility prediction, we split the urban area of *Feeder*, *Shanghai* and *GeoLife* into 15*10 grids, each with the size of $2km$*$2km$, as shown in Fig. 10. For the selected traces, we use the first day's data to train the mobility prediction model and conduct the historical data. Then, the MCS campaign begins at 8:00 of the second day. The tasks will be generated in grids and the users will participate in the MCS campaign in real time, with the uniform costs and active time. If we recruit one user, he will perform the tasks in the grids he will pass by within his active time.



(a) Number of users (b) Number of tasks

(c) Active time (min) (d) Duration time (min)

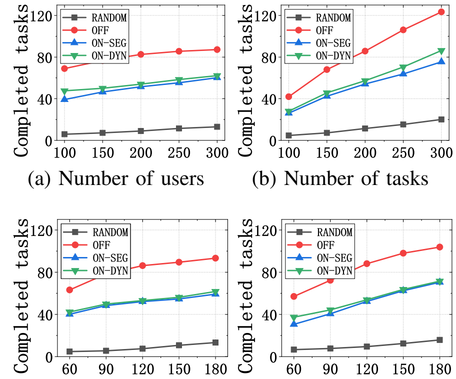
Fig. 7: Main results of *Feeder*.



(a) Number of users (b) Number of tasks

(c) Active time (min) (d) Duration time (min)

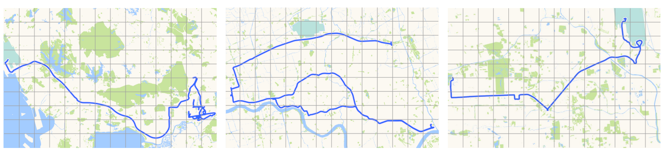
Fig. 8: Main results of *Shanghai*



(a) Number of users (b) Number of tasks

(c) Active time (min) (d) Duration time (min)

Fig. 9: Main results of *GeoLife*.



(a) *Feeder* (b) *Shanghai* (c) *GeoLife*

Fig. 10: An example of trajectories and grids.

B. Comparison Algorithms & Metrics

The online user recruitment problem with the budget and time constraints is quite different from the existing works, so we mainly compare our proposed online user strategies (referred as "ON-SEG" for segmented strategy and "ON-DYN" for dynamic strategy) with the following algorithms:

- RANDOM, which randomly recruits users from all participating users until the budget is exhausted.
- OFF, which greedily recruits the users who have the largest contribution/cost ratio in an offline manner, *i.e.*, $\arg\max_{u_i \in U} \big(f(\mu \cup \{u_i\}) - f(\mu)\big)/c_i$.
- OPT, which exhaustively recruits the optimal user set under the budget and time constraints.

Obviously, OPT costs a lot in the submodular user recruitment problem, and we implement it to verify our bound. In most cases, OFF and RANDOM can be seen as the upper and lower bound of our proposed strategies.

We use the following metrics to evaluate the compared algorithms: 1) Number of completed tasks, which is the main metric to evaluate our user recruitment strategy. 2) Consumed budget, which limits the number of users to be recruited and reflects the effectiveness. 3) Overpayment ratio, which shows the effectiveness of our online pricing mechanism, defined as the total payment/cost ratio, *i.e.*, $\sum_{u_i \in \mu}(p_i - c_i)/\sum_{u_i \in \mu} c_i$. In addition, the competitive ratio, truthfulness, and individual rationality are also presented in the following subsection.

C. Evaluation Results

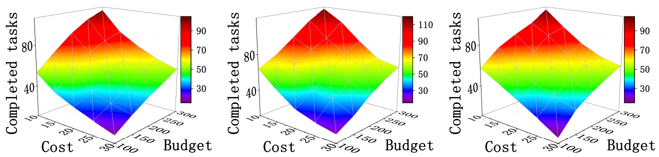
1) Completed tasks: We first illustrate the results in terms of the main metric, *i.e.*, the number of completed tasks, as shown in Figs. 7, 8, and 9. In order to provide a comprehensive evaluation, we change the number of participating users (from 100 to 300), the number of tasks (from 100 to 300), and the average active time of users (from 60 to 180 minutes) respectively, while keeping the others fixed. Moreover, we also evaluate the extra time constraint, *i.e.*, the duration time for tasks, which means that tasks should be completed within their duration time. The tasks will be generated with the uniform duration time and we change the average duration time of tasks from 60 to 180 minutes. Besides, the budget is set to 200 units and the average cost of users is 20 units. The results over three data sets show the similar tendencies and our proposed online user recruitment strategies can achieve a good performance.

Specifically, ON-SEG and ON-DYN outperform RANDOM and achieve high competitive ratios of OFF. Note that ON-DYN always complete more tasks than ON-SEG, since ON-DYN can correct the estimation errors and make use of the newly obtained information. Moreover, comparing the subfigures (a) and (b) of Figs. 7, 8, and 9, we find that the growth rates over the number of users are lower than tasks. The reason is that we have already recruited the effective users to perform the tasks, and thus the more users cannot improve the performance significantly. In addition, compared with *GeoLife*, ON-DYN performs better than ON-SEG in *Feeder* and *Shanghai*, since the traces in these two data sets have the stronger mobility and our dynamic strategies can make the adjustments in time.

2) Budget: We then evaluate the budget constraint with variable costs of users. We set the other variables fixed, then change the budget from 100 to 300 units and change the average cost of users from 10 to 30 units. As shown in Fig. 11, the lower budget and cost lead to a smaller number of completed tasks, since we have to recruit fewer users, and vice versa. Furthermore, we set the average cost to 20 units and illustrate the consumed budget over three data sets, as shown in Fig. 12. Obviously, the OFF and RANDOM consume more budget, since they recruit users in the offline manner until the budgets are exhausted. Note that ON-DYN always consumes more budget than ON-SEG, which shows that our dynamic strategy can make better use of the limited budget and conduct adjustments in time. These observations match the theoretical analysis and prove the effectiveness of our proposed strategy.

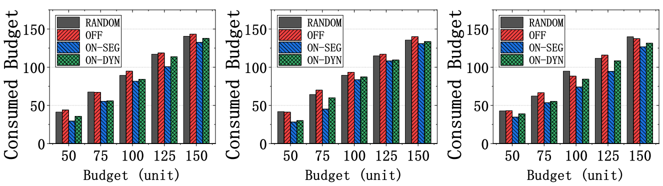
TABLE II: Completed tasks and competitive ratio.

	Feeder				*Shanghai*				*GeoLife*			
Budget	ON-SEG	ON-DYN	OPT	Ratio-DYN	ON-SEG	ON-DYN	OPT	Ratio-DYN	ON-SEG	ON-DYN	OPT	Ratio-DYN
100	22.9	30.5	71.3	0.4277	28.2	31.1	71.8	0.3927	31.6	37.1	73.2	0.4316
150	37.5	38.725	90.5	0.4279	38.5	43.85	94.4	0.4078	43.7	46.1	89.9	0.4860
200	46.4	48.75	115.3	0.4228	49.8	58.475	110.3	0.4514	55.7	56.675	102.8	0.5418
250	52.8	59.4	137	0.4335	53.7	73.125	127.1	0.4225	63.96	65.375	117.3	0.5452
300	67.5	69.6	160.5	0.4336	67.5	84	142	0.4753	70.1	72.4	123.9	0.5657



(a) *Feeder* (b) *Shanghai* (c) *GeoLife*

Fig. 11: Budget and cost of *Feeder*, *Shanghai*, and *GeoLife*.

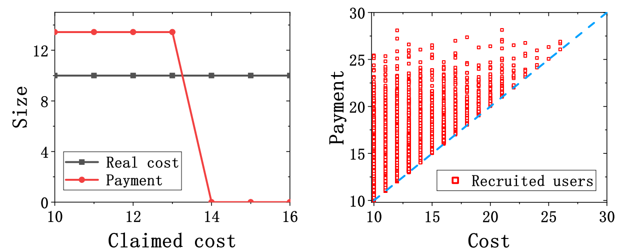


(a) *Feeder* (b) *Shanghai* (c) *GeoLife*

Fig. 12: Consumed budget of *Feeder*, *Shanghai*, and *GeoLife*.

TABLE III: Overpayment ratio.

	Budget				
	100	150	200	250	300
Feeder	0.2201	0.3080	0.3863	0.3954	0.3966
Shanghai	0.2229	0.3045	0.3812	0.3945	0.3942
GeoLife	0.2195	0.3046	0.3801	0.3920	0.3972



(a) Truthfulness (b) Individual rationality

Fig. 13: Truthfulness and individual rationality.

We also illustrate the competitive ratio of our proposed strategies in Table II. Under different budget constraints, our ON-DYN can achieve a 40%-50% competitive ratio of the optimal results, which is far higher than $(1 - 1/e)^2/7$ proven in Section IV.D. With the increase in budget, our ON-DYN even achieves a better competitive ratio, since we can recruit more effective users and the results are closer to OPT.

3) Pricing: Finally, we evaluate the performance of the online pricing mechanism lightly built into our online strategy. We first illustrate the overpayment ratio of the pricing mechanism in Table III. With the budget increase, we find that the pricing mechanism achieves a higher overpayment ratio. On the one hand, the larger budget allows us to pay more. On the other hand, under the larger budget, we will recruit more users, which means that the number of users in each segment decreases and we may use some worse observed threshold (ε in Algorithm 2) to set the payment ($p_i = c_i \cdot \delta_{u_i}/\varepsilon$ in Algorithm 4). Therefore, if we have more participating users (or some similar historical data), the online pricing mechanism will achieve a better performance.

In addition, we also provide two examples to verify the truthfulness and individual rationality, as shown in Fig. 13. To verify the truthfulness, we randomly select a recruited user, change his claimed cost and illustrate his obtained payment in Fig. 13 (a). The real cost of this user is 10 units, and the payment determined by the online pricing mechanism is 13.44 units. When the user bids a smaller cost than the determined payment, he will be recruited and obtain his payoff $13.44 - 10 = 3.44$. If the user claims a larger cost, he won't be recruited. Note that the payment is determined according to the users' claimed costs, thus they won't claim the lower costs, which is actually known as the individual rationality. As shown in Fig. 13 (b), we compare the payments with the users' claimed costs under the budget $B = 200$ units and the average cost $c = 20$ units. We can see that each payment is larger than the related cost, which verifies the individual rationality.

VII. Conclusion

In this paper, we investigate the online user recruitment problem under the budget and time constraints in MCS, where users participate in real time and we decide whether to recruit them immediately when they are arriving. To deal with the budget and time constraints, we first estimate the number of users to be recruited and then segmentally recruit users in an online manner. In order to correct estimation errors and utilize newly obtained information, we further present a dynamic re-estimation after recruiting every new user, and finally a truthful pricing mechanism is lightly built into the dynamic user recruitment strategy. Extensive evaluations on three real-world data sets have verified the effectiveness of our proposed strategy. In the future work, we would like to explore the privacy protection mechanism in our online user recruitment solutions for MCS.

Acknowledgment

This work is supported in part by NSFC under Grant No. 61772230, 61972450, 61702215, and Changchun Science and Technology Development Project No.18DY005, and NSF grants CNS 1824440, CNS 1828363, CNS 1757533, CNS 1618398, CNS 1651947, and CNS 1564128.

REFERENCES

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] D. Zhang, L. Wang, H. Xiong, and B. Guo, "4w1h in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 42–48, 2014.

[3] Z. Liu, S. Jiang, P. Zhou, and M. Li, "A participatory urban traffic monitoring system: the power of bus riders," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2851–2864, 2017.

[4] H. Aly, A. Basalamah, and M. Youssef, "Automatic rich map semantics identification through smartphone-based crowd-sensing," *IEEE Transactions on Mobile Computing*, no. 10, pp. 2712–2725, 2017.

[5] D. Yang, G. Xue, X. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1732–1744, Jun. 2016.

[6] K. Han, H. Huang, and J. Luo, "Quality-aware pricing for mobile crowdsensing," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1728–1741, Aug 2018.

[7] M. Karaliopoulos, I. Koutsopoulos, and L. Spiliopoulos, "Optimal user choice engineering in mobile crowdsensing with bounded rational users," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1054–1062.

[8] W. Liu, Y. Yang, E. Wang, Z. Han, and X. Wang, "Prediction based user selection in time-sensitive mobile crowdsensing," in *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2017, pp. 1–9.

[9] 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.

[10] N. Wang and J. Wu, "Cost-efficient heterogeneous worker recruitment under coverage requirement in spatial crowdsourcing," *IEEE Transactions on Big Data*, pp. 1–1, 2019.

[11] L. Pu, X. Chen, J. Xu, and X. Fu, "Crowd foraging: A qos-oriented self-organized mobile crowdsourcing framework over opportunistic networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 848–862, April 2017.

[12] Y. Yang, W. Liu, E. Wang, and J. Wu, "A prediction-based user selection framework for heterogeneous mobile crowdsensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[13] D. Zhao, X. Li, and H. Ma, "Budget-feasible online incentive mechanisms for crowdsourcing tasks truthfully," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 647–661, April 2016.

[14] H. Gao, C. H. Liu, J. Tang, D. Yang, P. Hui, and W. Wang, "Online quality-aware incentive mechanism for mobile crowd sensing with extra bonus," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[15] J. Li, J. Wu, and Y. Zhu, "Selecting optimal mobile users for long-term environmental monitoring by crowdsourcing," in *Proc. of the IEEE/ACM International Symposium on Quality of Service (IWQoS 2019)*, 2019.

[16] M. Bateni, M. Hajiaghayi, and M. Zadimoghaddam, "Submodular secretary problem and extensions," *ACM Transactions on Algorithms (TALG)*, vol. 9, no. 4, p. 32, 2013.

[17] R. Myerson, "Optimal auction design," *Math Oper Res*, vol. 6, 01 1981.

[18] J. Liu, H. Shen, H. S. Narman, W. Chung, and Z. Lin, "A survey of mobile crowdsensing techniques: A critical component for the internet of things," *ACM Trans. Cyber-Phys. Syst.*, vol. 2, no. 3, pp. 18:1–18:26, Jun. 2018.

[19] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions and opportunities," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[20] Y. Liu, L. Kong, and G. Chen, "Data-oriented mobile crowdsensing: A comprehensive survey," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[21] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, "User recruitment for mobile crowdsensing over opportunistic networks," in *IEEE INFOCOM 2015 - IEEE Conference on Computer Communications*, 2015.

[22] F. Zhang, B. Jin, H. Liu, Y. Leung, and X. Chu, "Minimum-cost recruitment of mobile crowdsensing in cellular networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–7.

[23] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, "Qoi-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4618–4632, 2014.

[24] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, "Social-network-assisted worker recruitment in mobile crowd sensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1661–1673, July 2019.

[25] X. Wang, R. Jia, X. Tian, and X. Gan, "Dynamic task assignment in crowdsensing with location awareness and location diversity," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 2420–2428.

[26] H. Li, T. Li, W. Wang, and Y. Wang, "Dynamic participant selection for large-scale mobile crowd sensing," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2018.

[27] T. S. Ferguson, "Who solved the secretary problem?" *Statistical Science*, vol. 4, no. 3, pp. 282–289, 1989.

[28] J. Preater, "On multiple choice secretary problems," *Mathematics of Operations Research*, vol. 19, no. 3, pp. 597–602, 1994.

[29] U. Feige, "A threshold of ln n for approximating set cover," *Journal of the Acm*, vol. 45, no. 4, pp. 634–652, 1999.

[30] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Oper. Res. Lett.*, vol. 32, no. 1, pp. 41–43, Jan. 2004.

[31] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Transactions on Mobile Computing*, vol. 17, no. 12, pp. 2982–2997, Dec 2018.

[32] D. Zhang, J. Zhao, F. Zhang, R. Jiang, and T. He, "Feeder: Supporting last-mile transit with extreme-scale urban infrastructure data," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, ser. IPSN '15. New York, NY, USA: ACM, 2015, pp. 226–237.

[33] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th International Conference on Ubiquitous Computing*, ser. UbiComp '08. New York, NY, USA: ACM, 2008, pp. 312–321.

[34] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 791–800.