Dynamic Online User Recruitment With (Non-) Submodular Utility in Mobile CrowdSensing

Wenbin Liu, *Member, IEEE*, Yongjian Yang, En Wang, *Member, IEEE*, Hengzhi Wang, Zihe Wang, and Jie Wu, *Fellow, IEEE*

***Abstract*— Mobile CrowdSensing (MCS) has recently become a powerful paradigm that recruits users to cooperatively perform various tasks. In many realistic settings, users participate in real time and we have to recruit them in an online manner. The existing works usually formulate the online recruitment problem as a budgeted optimal stopping problem with submodular user utility, while we first argue that not only the budget but also the time constraints can jointly influence the recruitment performance. For example, if we have less budget but plenty of time, we should recruit users with more patience. Second, considering the user's cooperative willingness, its contribution may be diminishing or even irregular. Hence, we also need to address not only submodular cases but also their non-submodular utility. In this paper, we study the online user recruitment problem with (non-)submodular utility 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. Moreover, we extend the segmented strategy with a non-submodular utility, which has the submodularity ratio γ and the competitive ratio $\gamma^2(1-e^{-1})/7$. Furthermore, to correct estimation errors and utilize newly obtained information, we dynamically re-adjust the segmented strategy and also prove that the dynamic strategy achieves a competitive ratio of $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$. 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 under the (non-) submodular utility and two constraints.**

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

Manuscript received June 21, 2020; revised October 26, 2020 and January 30, 2021; accepted May 11, 2021; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor J. Liu. Date of publication June 24, 2021; date of current version October 15, 2021. This work was supported in part by the National Natural Science Foundations of China under Grant 61972450 and Grant 62072209; in part by the Natural Science Foundations of Jilin Province under Grant 20190201022JC; in part by the Changchun Science and Technology Development Project under Grant 18DY005; in part by the Jilin Province Young Talents Lifting Project under Grant 3D4196993421; and in part by the NSF under Grant CNS 1824440, Grant CNS 1828363, Grant CNS 1757533, Grant CNS 1629746, Grant CNS 1651947, and Grant CNS 1564128. A preliminary version of this work appeared in the IEEE International Conference on Computer Communications (INFOCOM 2020) [1].
(Corresponding author: En Wang.)

Wenbin Liu, Yongjian Yang, En Wang, and Hengzhi Wang are with the College of Computer Science and Technology, Jilin University, Changchun 130012, China (e-mail: liuwb16@mails.jlu.edu.cn; yyj@jlu.edu.cn; wangen@jlu.edu.cn; wanghz17@mails.jlu.edu.cn).

Zihe Wang is with the Gaoling School of Artificial Intelligence, Renmin University of China, Beijing 100872, China (e-mail: wzh5858588@163.com).

Jie Wu is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: jiewu@temple.edu).
This article has supplementary downloadable material available at https://doi.org/10.1109/TNET.2021.3083955, provided by the authors.

Digital Object Identifier 10.1109/TNET.2021.3083955

I. INTRODUCTION

WITH the great popularization of mobile devices and wireless communications, Mobile CrowdSensing (MCS) [2] has rapidly become a powerful paradigm, which recruits a number of mobile users to cooperatively perform various urban sensing and computing tasks, e.g., traffic estimation, indoor-outdoor localization, and map semantics identification. In general, the MCS applications should provide proper rewards for the recruited users in order to cover their costs and encourage the participation [3]. However, considering the limited budget, we usually cannot afford all of the users, but have to recruit those who can complete the tasks more effectively, which raises the fundamental user recruitment problem in MCS [4], [5].

Many existing works on user recruitment are conducted offline [6], [7], which recruits users with full information at the beginning of the MCS campaign. However, in many realistic settings, the users participate in the MCS campaign in real time and we have to recruit them in an online manner with partial information. Fig. 1 provides an illustrative example of the online scenario, where users will participate at any time and cooperatively perform the tasks with different locations. In this case, each user's cost and contribution are invisible before their participation, which makes the online user recruitment more challenging.

To deal with such online scenarios, some existing works formulate the online user recruitment as the optimal stopping problem and propose some effective algorithms based on secretary problem [8] or dynamic programming [9], but they simply use a limited number of recruited users as the budget constraint. Some researchers further consider the various costs of users [10]–[12], but they mainly focus on the budget constraint while ignoring the influence of the remaining time of the MCS campaign. In this paper, we argue that these two constraints seem to be independent but jointly affect the online user recruitment, which should be considered simultaneously. As an example, if there is less time left, we would better recruit all of the participating users, in order to use up the remaining budget as soon as possible. Similarly, when we have less remaining budget but plenty of time, we should recruit users with more patience. Therefore, how to deal with *the budget and time constraints* is the first challenge in online user recruitment.

Moreover, most of the existing works characterize the contribution by submodular utility, e.g., the probability of performing tasks according to the mobility, as shown in Fig. 1 (upper part), which reflects the diminishing marginal returns. However, in reality, the contribution of a user set is more complex and sometimes even non-submodular. Fig. 1 (lower part)

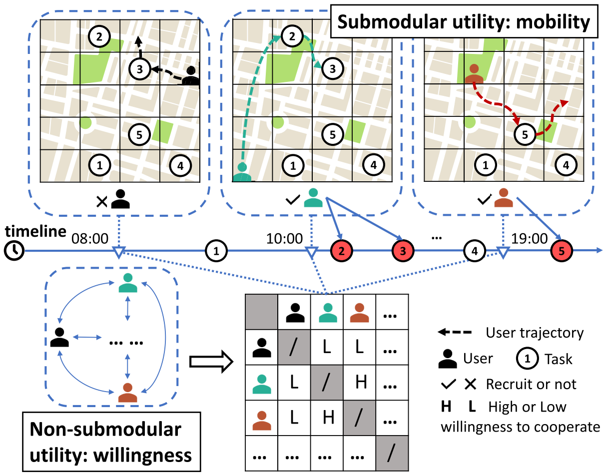


Fig. 1. Online user recruitment with (non-) submodular utility in MCS.

gives an example of the willingness of cooperation to show the non-submodular utility: the first user is not familiar with the others, and thus he may be unwilling to cooperate with them (e.g., because of privacy leaks existed in tracking tasks [13]), which leads to the irregular contribution. Obviously, under the online scenarios with unknown future information, we can hardly estimate *the submodular and in particular non-submodular utility*, which is the second challenge for online user recruitment.

Furthermore, both the users' dynamical participation and their (non-) submodular utility introduce a lot of uncertainty. As shown in Fig. 1, we can hardly accurately calculate the user's contribution (estimated by his mobility and cooperative willingness) and cost (randomly claimed from an independent cost distribution) during the online recruiting process. Moreover, even if we assume that users will participate periodically, the period still needs to be learned from historical data and may also be inaccurate. All these uncertain factors make the estimated results not always precise. Therefore, the third challenge is how we can *dynamically re-adjust* our recruitment strategy along with the online recruiting process.

Additionally, as a supplement to the online recruitment strategy, after deciding whether to recruit one user, we should also provide proper reward within a payment budget for covering the user's sensing cost. Meanwhile, we also need to encourage user participation and avoid being deceived. Therefore, we should determine a *truthful price* for each recruited user in this online manner, which is the fourth challenge.

In this paper, to address the budget and time constraints, we test the distribution of users' participating time according to the experiments and find a periodic pattern, which can be learned from the historical data. Thus, we first predict the total number of participating users according to the time constraint. Then, taking the budget constraint into account, we further estimate the number of users to be recruited. In this way, we formulate the online user recruitment problem under two constraints as a classic multiple-choice secretary problem [14]. Moreover, considering the (non-) submodular utility, i.e., the diminishing or even irregular contribution mentioned in the example of Fig. 1, we extend the multiple-choice secretary problem with a submodular or non-submodular contribution function [15]. Then, we approximately divide all participating users into some equally-sized segments and try to recruit the best user in each segment. For different segments, we recruit users according to their submodular or non-submodular marginal contributions. Especially, for the non-submodular utility with a submodularity ratio γ, we prove that our local-optimal selection leads to a global near-optimal solution, with the competitive ratio $\gamma^2(1-e^{-1})/7$.

Furthermore, in order to correct the errors in estimation and make use of the new information obtained during the recruitment 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 recruitment in the above segmented online process. Also, we prove that the proposed dynamic online recruitment strategy achieves a competitive ratio of $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$ under a (non-) submodular utility. Finally, we conduct a reverse auction-based pricing mechanism, which can be easily built into our online user recruitment strategies without much extra computation. This online pricing mechanism is proved to achieve truthfulness and individual rationality.

In summary, this paper has the following contributions:

- **Online User Recruitment**: We study the online user 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 propose a segmented online user recruitment strategy. Moreover, considering the users' diminishing or even irregular contribution, we extend the online user recruitment strategy from a submodular case to the non-submodular one, where the competitive ratio is proved to be $\gamma^2(1-e^{-1})/7$.
- **Dynamic Re-adjust**: We dynamically re-adjust the segmented online user recruitment strategy, in order to correct the estimation errors and utilize the newly obtained information, where the competitive ratio is proved to be $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$.
- **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 extensive evaluations based on three real-world data sets. The results verify the effectiveness of our strategy on improving the number of completed tasks under the (non-) submodular utility and two constraints.

This paper is organized as follows. After reviewing the related works in Section II, we introduce the system model and formulate the research problem in Section III. Then, the dynamic online user recruitment strategy is proposed in Section IV, followed by the theoretical analysis in Section V. In Section VI, we introduce the online pricing mechanism. Finally, we evaluate the performance in Section VII and conclude this paper in Section VIII.

II. Related Work

A. User Recruitment

Mobile CrowdSensing is a promising paradigm, which allows us to recruit users carrying portable devices, in order to cooperatively perform various sensing tasks. Considering the sensing costs, Karaliopoulos *et al.* [4], Zhang *et al.* [16], and Song *et al.* [17] study the user recruitment problem to achieve the goal of the MCS campaigns and minimize the total costs. Similarly, Liu *et al.* [6] and Wang *et al.* [7] propose the prediction-based algorithms to recruit the effective users, in order to complete more tasks under a budget constraint. Some researchers further consider the location dependent tasks [18], social media users [19], the multiple tasks [20] and diverse users' factors [21] to measure the completion quality and overall utility for task allocation, which is very similar to user recruitment but focuses on assigning tasks to users. However, most of the existing 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.* [22] study the location-aware and location diversity based online MCS but focus on the task assignment. Li *et al.* [23] 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.* [8] 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.* [10], Gao *et al.* [11], and Li *et al.* [12] 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, but they haven't dealt with the total budget and ignore the influence of the remaining time of the MCS campaigns.

B. (Non-) Submodular Utility

The user recruitment problem can be naturally formulated as the subset selection problem with the (non-) submodular utility under linear cost constraints [4]. For the submodular utility, the generalized greedy algorithm is to iteratively select the user with the largest ratio of the marginal contribution and cost with the $(1-e^{-1})/2$-approximation guarantee [24], which is further improved to $(1-e^{-1/2})$ [25]. Combined with partial enumeration, this generalized greedy algorithm has been proved to achieve a $(1-e^{-1})$ approximation ratio [26]. For the non-submodular utility, by using the submodularity ratio γ, which reflects how close the utility function is to being submodular, Qian *et al.* [27] derive that the generalized greedy algorithm obtains a $\gamma(1-e^{-\gamma})/2$-approximation guarantee and also propose an anytime randomized iterative approach to cost more time for better solutions. Bian *et al.* [28] also combine the curvature α and submodularity ratio γ to derive a tight approximation guarantee of $(1-e^{-\alpha\gamma})/\alpha$ for cardinality constrained maximization.

C. Secretary Problem

For the secretary recruitment, the classic secretary problem is to recruit only one best user from all participating users in an online manner [29]. As a variant, Preater [14] studies that more than one user may be recruited in the secretary problem. Considering the submodular utility function, Bateni *et al.* [15] 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. In this paper, we further consider the non-submodular utility function for online user recruitment problem by using the submodularity ratio γ.

III. System Model and Problem Formulation

A. System Model

We consider a practical online scenario of MCS, where a crowd of rational users move around and participate in the MCS campaign in real time to perform the sensing tasks. Users are denoted as $U \triangleq \{u_1, u_2, \ldots, u_n\}$, each with active (working) time slots $A_i \triangleq \{t_i^b, \ldots, t_i^e\}$ and sensing cost c_i, which indicates that user u_i will work in A_i with cost c_i.[1] Tasks are denoted as $S \triangleq \{s_1, s_2, \ldots, s_m\}$ each with a location $l_j \in L$, and we consider that a user u_i moving to location l_j within his active time slots A_i can perform the task s_j. Under the online scenarios, a user u_i participates in real time with a bid price b_i, and we decide whether to recruit him immediately, with the payment p_i under a budget constraint $\sum_{i=1}^n p_i \leq B$. In general, we hope that users will bid truthfully, i.e., $b_i \approx c_i$, and we should provide proper rewards to cover their costs and encourage the participation, i.e., $p_i \geq b_i$. Then, the recruited users, denoted as μ with the set cardinality $k = |\mu|$, perform the sensing tasks within the duration time slots of the MCS campaign $T \triangleq \{t^b, \ldots, t^e\}$.[2] The main notations used in this paper are listed in Table I.

TABLE I

Main Notations

Notation	Meaning
u, c, $\{A^b, ..., A^e\}$	User, cost, and active time slots.
s, l, $\{T^b, ..., T^e\}$	Task, location, and duration time slots.
b, p, B	Bid price, payment and budget.
S, U, μ	Set of tasks, users and recruited users.
m, n, k	Number of tasks, users and recruited users.
$f(\mu)$, $g(\mu)$	Submodular and non-submodular utility functions.
$Z_u(l_i, l_j, t)$, $Q_u(l_i, l_j, t)$	Probability that u moves from l_i to l_j within t time slots, and just at the t-th time slot.
$P/\hat{P}(u_i, s_j)$	Probability that u_i will complete s_j (with the cooperative willingness).
$E/\hat{E}(\mu, s_j)$	Probability that μ will complete s_j (with the cooperative willingness).
w_{ij}	Cooperative willingness between u_i and u_j.
$W(u_i, \mu)$	Average cooperative willingness among u_i and other users in $\mu \setminus u_i$.

To further reduce the complexity, we assume that all tasks are equal in quality and only need to be completed once Also, 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, in most cases, 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 information. After receiving the decisions, users will leave and their next participation will be seen as the new ones.

B. Mobility Prediction

For the user recruitment, we would like to select the best user set that cooperatively contributes the most on MCS campaign. To reflect the diminishing marginal rewards of the newly recruited users, most of the existing works characterize

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

[2] Note that we consider a discrete model to deal with the users' irregular participation, and the continuous model can be easily modified.

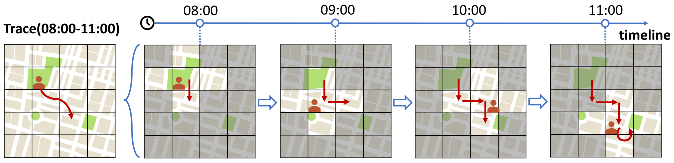


Fig. 2. An example of mobility prediction model.

the users' contribution by submodular utility. As a typical example, we can utilize the users' uncertain mobilities to estimate their contributions (the submodular property will be proved in the next section). From such an 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 [30]. In another word, we only use the mobility prediction to estimate the user's contribution while ignoring his skills, device, and quality of sensed data, in order to reduce the complexity.[3]

Specifically, as shown in Fig. 2, we divide the full map into some grids [31]. 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 [6]–[8] to predict the time-dependent transition probabilities between the grids as the user's mobility prediction.[4] Then, the time-dependent semi-Markov 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 t-th time slot, is defined by Eq. (1).

$$Q_u(l_i, l_j, t) = \begin{cases} \Sigma_{l_k}^{L} \Sigma_{t'=1}^{t} (Z_u(l_i, l_k, t') - Z_u(l_i, l_k, t'-1)) \cdot \\ Q_u(l_k, l_j, t-t'), \quad l_i \neq l_j \\ 1 - \Sigma_{l_k, l_k \neq l_i}^{L} (Z_u(l_i, l_k, t) - \\ \Sigma_{t'=1}^{t} (Z_u(l_i, l_k, t') - Z_u(l_i, l_k, t'-1)) \cdot \\ Q_u(l_k, l_i, t-t')), \quad l_i = l_j \end{cases} \tag{1}$$

where $Z_u(l_i, l_j, t)$ denotes the probability that user u will move from his current grid l_i to his next grid l_j within t time slots and can be derived from the statistical records. Specifically, we consider the relay state transitions as $l_i \rightarrow l_k \rightarrow l_j$ when $l_i \neq l_j$, and also calculate the probability that users stay at the same grid when $l_i = l_j$. In order to further reduce the great amount of calculation, as shown in Fig. 2, we only calculate the transitions between the nearby grids, $i.e.$, up, down, left, right and itself. Based on the Q function, 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:

$$P(u_i, s_j, T) = 1 - \Pi_{t \in A_i \wedge t \in T} \big(1 - Q_{u_i}(l_{u_i}, l_{s_j}, t)\big), \tag{2}$$

$$E(s_j, \mu, T) = 1 - \Pi_{u_i \in \mu} \big(1 - P(u_i, s_j, T)\big). \tag{3}$$

C. Cooperative Willingness

Besides the mobilities, in reality, there are also many factors that may influence the users' contribution on MCS campaign, such as the users' abilities, preferences, and willingness. These factors make the users' contribution more complex that the utility function may even be non-submodular. We offer a real and specific application scenario, which exploits the relays of mobile users to help transport package, without influencing their daily routes too much.[5] Under this scenario, a user has to meet the next user to transport packages at a pre-determined time and location. Thus, their privacy information has to be exposed to each other. Obviously, due to the privacy concerns, users may not want to cooperate with unfamiliar people but prefer to cooperate with their friends, which leads to a more complex utility function (i.e., a non-submodular function).

Specifically, we define the pairwise willingness to cooperate between users u_i and u_j as w_{ij}, where $0 < w_{ij} < 1$ denotes the probability that they would like to perform tasks. Then, we simply calculate the average cooperative willingness $W(u_i, \mu)$ among u_i and other users in $\mu \backslash u_i$:

$$W(u_i, \mu) = \frac{\sum_{u_j \in \mu \backslash u_i} w_{ij}}{|\mu| - 1}. \tag{4}$$

Note that the average willingness W is actually a discrete and time-varying function in our online user recruitment problem, since the recruited (and active) users are changed in an online manner. In other words, for the t-th time slot, the users' cooperative willingness should be calculated according to the current μ, denoted as $W(u_i, \mu_t)$. Combined with the Q function in mobility prediction, we then obtain the new probability that user u_i can complete task s_j under the cooperative willingness, and get the new contribution:

$$\hat{P}(u_i, s_j, \mu, T) = 1 - \Pi_{t \in A_i \wedge t \in T} \times \big(1 - W(u_i, \mu_t) Q_{u_i}(l_{u_i}, l_{s_j}, t)\big), \tag{5}$$

$$\hat{E}(s_j, \mu, T) = 1 - \Pi_{u_i \in \mu} \big(1 - \hat{P}(u_i, s_j, \mu, T)\big). \tag{6}$$

D. Problem Formulation

To summarise, in this paper, we consider a practical online MCS with dynamically participating users. There is a budget B for the whole MCS campaign rather than per task, which is used to recruit users and pay for their participation. Note that tasks are weighted equally and only need to be completed once. We also assume that users entering a grid can complete the tasks in it, ignoring the differentiation of users' skills, devices, and task completion qualities. Moreover, we further consider the cooperative willingness attributed to each user depending on which users are recruited together. Then, we describe our research problem formally.

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 cooperatively perform tasks, with (non-) submodular utility and the objective of maximizing the number of completed tasks*:

$$\text{maximize} \quad \Sigma_{s_j \in S} \hat{E}(s_j, \mu, T) \tag{7}$$

$$\text{subject to} \quad \mu \subseteq U, \; \Sigma_{u_i \in \mu} p_i \leq B \tag{8}$$

A running example shown in Fig. 3 provides an intuitive interpretation of our online user recruitment problem. Consider 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 tasks 2 and 3 within his active time. However, user 1 is not familiar with others and she doesn't want to

[3]Actually, such assumptions or conditions can be regarded as the probabilistic effects on completing tasks, which can be simply modified and added into the probabilistic utility function.

[4]Other existing mobility prediction models can be easily modified in our utility function.

[5]It is a more general application scenario from CrowdDeliver [32].

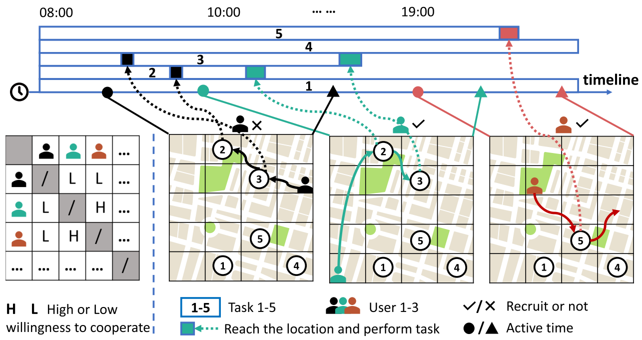


Fig. 3. An example of online user recruitment in MCS.

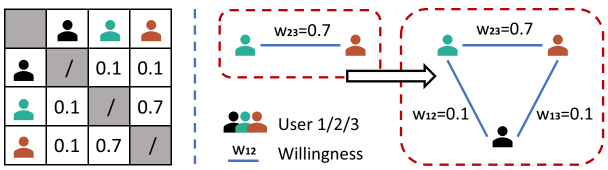


Fig. 4. An example of non-submodular utility.

cooperatively perform these tasks. Hence, her contribution seems relatively less and we decide to keep waiting since we have enough time. When user 2 connects to server, we find that she would like to cooperate with user 3 to perform tasks 2 and 3, which contributes a lot and thus we recruit her. When user 3 connects to server, although she may perform only one task but we have less remaining time, and hence, we recruit her. Finally, we recruit users 2 and 3 in an online manner to cooperatively complete the most tasks under the budget and time constraints.

IV. Online User Recruitment Under Budget and Time Constraints

A. Problem Hardness and (Non-) Submodular Utility

Before prescribing the online strategies, we first prove that the online user recruitment problem is NP-hard.

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

proof. Without loss of generality, we ignore the constraints and the online process, but consider a simple MCS scenario that we can recruit k users to complete tasks, which is indeed a classic NP problem, *Max k-cover* [33]: given a collection of task sets $\{S_{u_1}, S_{u_2}, \ldots, S_{u_n}\}$, each of which covers several tasks $S_{u_i} = \{s_{i_1}, s_{i_2}, \ldots\}$ completed by the user u_i, then the objective is to select k sub-collections to cover the most tasks. Thus, the special case is NP-hard. Consequently, further considering the constraints and the online process, our problem is NP-hard. □

Moreover, considering the users' diminishing and even irregular contribution, in this paper, we use a submodular function $f()$ and a non-submodular function $g()$ to measure the recruited users' utility. Specifically, we define $f(\mu) = \Sigma_{s_j \in \mathcal{S}} E(s_j, \mu)$ according to Eq. 3, as the submodular predicted contribution of the recruited user set μ. Similarly, we have $g(\mu) = \Sigma_{s_j \in \mathcal{S}} \hat{E}(s_j, \mu)$ according to Eq. 6, as the non-submodular utility of μ with the following property:

***Proposition** 1: 1) $g(\emptyset) = 0$; 2) $g(\mu)$ is non-submodular.*

proof. 1) If we haven't recruited any users, then no tasks will be performed. Hence, we have $g(\emptyset) = \Sigma_{s_j \in \mathcal{S}} \hat{E}(s_j, \emptyset) = 0$ according to Eq. 6. 2) As shown in Fig. 4, we provide an intuitive example to prove the non-submodular property of $g(\mu)$. Without loss of generality, we consider that all of the three users have the pre-determined traces that each of them can cover 2 tasks without overlapping. The willingness between two of them are listed on the left part, where $w_{12} = w_{13} = 0.1$ and $w_{23} = 0.7$. According to Eq. 6, we obtain

$$\begin{aligned}
&\big(g(\{u_2, u_3\}) - g(\{u_3\})\big) - \big(g(\{u_2, u_1, u_3\}) - g(\{u_1, u_3\})\big) \\
&\quad = \Sigma_{s_j \in \mathcal{S}} \hat{E}(s_j, \mu_{23} = \{u_2, u_3\}) - \ldots \\
&\quad = 2 \cdot W(u_2, \mu_{23}) Q_{u_2} + 2 \cdot W(u_3, \mu_{23}) Q_{u_3} - \ldots \\
&\quad = (2.8 - 2) - (1.8 - 0.4) = 0.8 - 1.4 < 0. \qquad (9)
\end{aligned}$$

Moreover, assume that we already have recruited users u_2 and u_3 and then add u_1 to them. Then we have $g(\{u_1\} \cup \{u_2, u_3\}) = 1.8 < 2.8 = g(\{u_2, u_3\})$. Thus, $g()$ is non-submodular and non-monotonic. □

In addition, in real recruitment, we always recruit users if and only if they have the positive gains, i.e., $g(u_i \cup \{\mu\}) - g(\{\mu\}) > 0$. Hence, we actually consider the non-decreasing functions and not additions that decrease the overall utility.

B. Segmented Online User Recruitment Strategy

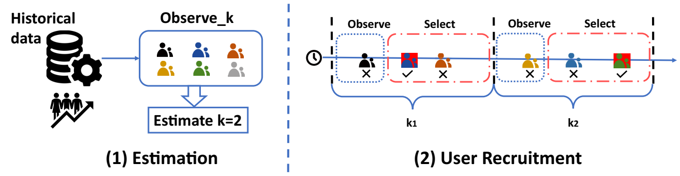


Fig. 5. Segmented online user recruitment strategy.

In online scenarios, to deal with the budget and time constraints, we present a segmented 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. 5.

1) Estimation via (Non-) 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 (non-) submodular utilities simultaneously. To reduce the difficulty, we first consider the payments are equal to users' bids and costs, i.e., $p_i = b_i = c_i$, and also make an assumption to deal with the budget and time constraints before the online recruiting:

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

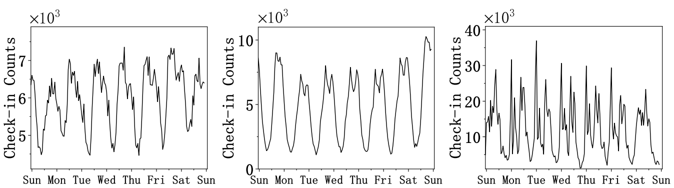
In many scenarios where humans are involved, Assumption 1 is common and reasonable, such as people's check-in records of *Gowalla*, *Brightkite*, and *GeoLife*[6] in Fig. 6, the counts of which within the same periods in weekdays are roughly the same and independent from people. Under this assumption, 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

[6] *Gowalla* and *Brightkite* [34] are two famous location-based social networking service providers where users share their locations by checking-in, and *GeoLife* [35], used in our evaluation section, can be seen as a MCS application. The user activities in these apps have a lot in common, e.g, the users have to reach some location to check-in or perform sensing tasks. Thus, we use them to verify Assumption 1.

Algorithm 1 Greedy Offline *Estimation()*

Input: S, B, $T = \{t^s, \dots, t^e\}$, U_h

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



(a) *Gowalla* (b) *Brightkite* (c) *GeoLife*

Fig. 6. Check-in records of *Gowalla*, *Brightkite*, and *GeoLife*.

construct 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 (non-) submodular maximization with knapsack constraint [24]–[27], and propose a greedy offline algorithm to estimate the number of users to be recruited, which is summarized in Algorithm 1.

We first estimate the number of participating users n' and construct a simulated user set U' from historical data U_h, according to the time T (line 1). From the simulated user set, according to the (non-) submodular utility function, we greedily select the users who have the largest ratio of marginal revenues and payments in an offline manner and recruit them under the budget constraint (lines 4-5). Finally, we obtain the estimated n' and k, which actually represents the budget and time constraints with the (non-) submodular utility. 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) Segmented 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 (non-) submodular k-secretaries problem [15], not 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 $et\ al.$ [15] further considers the multiple recruited users with submodular utility functions. Similarly, we further extend the k-secretaries problem with a non-submodular utility function, which also provides an appropriate solution for our online user recruitment problem with non-submodular contribution functions.

Algorithm 2 Online User Recruitment *Segmented()*

Input: S, B, $U = \{u_1, u_2, \dots, u_n\}$, 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** $\sum_{u_j \in \mu} p_j + p_i \leq B$ **then** $\mu \leftarrow \mu \cup \{u_i\}$;
5: **else**
6: Initialize *segmentID* $= i/l$;
7: Calculate $\delta_{u_i} = (f(\mu \cup \{u_i\}) - f(\mu))/p_i$ $(f \rightleftharpoons g)$;
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** $\sum_{u_j \in \mu} p_j + p_i \leq B$ **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

Without considering the budget and time constraints, our online user recruitment can be naturally formulated as a (non-) submodular k-secretaries problem, interpreted as 'recruit k out of n users to maximize the expected number of completed tasks according to a (non-) submodular utility function'. Then, we propose a k-segments online user recruitment algorithm, as summarized in Algorithm 2. With the estimated n' and k, we first approximately partition the real participating users into k equally-sized segments and select the best user from each segment according to the (non-) submodular utility functions. 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 4). For each segment, under the budget constraint, with the natural constant e, we observe the first $l_{ob} = \lfloor 1/e \cdot l \rfloor$ users, record the largest contribution/payment 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 users.

Note that users are split into segments according to arrival order. Actually, from the view of unknown future users and their arrival time, the users can be seen as being assigned to segments randomly, which can be dealt with well by the secretary problem. As shown in Fig. 7, we provide some intuitive examples and conduct some experiments to explain the segmented strategy, especially on the indexing order. Specifically, we index the users in descending, ascending, random, and original orders respectively, divide them into segments in order, and compare their performances in terms of completed tasks. Since the secretary problem observes the first some users and then select the next larger one, the descending case can recruit no users and thus completes no tasks. Similarly, the ascending case will recruit the first user after observing, thus the recruited users can complete relatively few tasks, while the random and original cases can achieve higher and similar performances. Therefore, our segmented strategy can deal with the online user recruitment well.

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 and (non-) submodular

Algorithm 3 Dynamic Online User Recruitment

Input: S, B, $T=\{t^s,\dots,t^e\}$, $U\{u_1,u_2,\dots,u_n\}$
1: $\mu=\emptyset$, $S_{rest}=S$, $B_{rest}=B$, $T_{rest}=T$, $U_{rest}=U$;
2: **while** $B_{rest}>0$ **and** $U\neq\emptyset$ **do**
3: $n',k\leftarrow$ *Estimation(*S_{rest}*,* B_{rest}*,* T_{rest}*)*;
4: ▷ **If** $k=0$ or $n>n'$, recruit users under B_{rest}
5: $\mu\leftarrow$ *Segmented(*S_{rest}*,* B_{rest}*,* U_{rest}*,* n'*,* k*,* μ*)*;
6: ▷ **Break** after a new user has been recruited
7: Update S_{rest}, B_{rest}, T_{rest}, U_{rest};
return μ

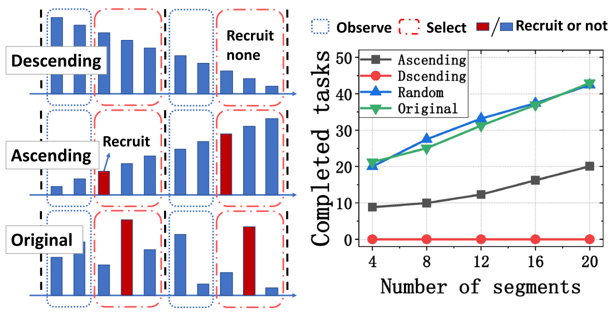


Fig. 7. Segmented online user recruitment strategy with different indexing orders.

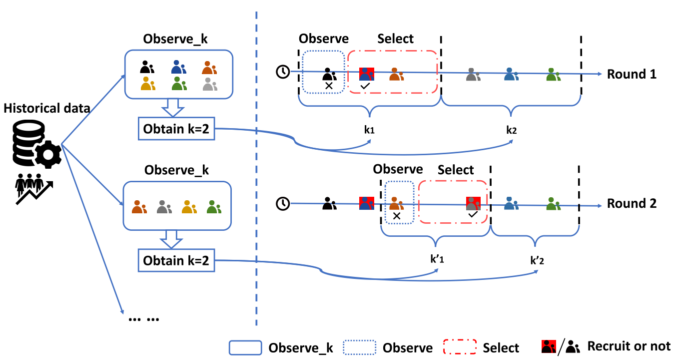


Fig. 8. Dynamic online user recruitment strategy.

utilities, 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. 8 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). 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.

V. Theoretical Analysis

A. *Analysis on* Estimation() *Algorithm*

Before analyzing the online user recruitment strategies, we first analyze the Algorithms 1, i.e., the *Estimation()*. We 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 mentioned above, our objective function $f(\mu)$ is non-decreasing and submodular. Then, the user recruitment problem is formulated as a variant of submodular maximization problem with linear costs, which is NP-hard and our offline greedy algorithm can achieve a $(1-e^{-1/2})$ approximation [25] of the optimal value, denoted as $f(\mu)\geq(1-e^{-1/2})\cdot f(OPT)$, where OPT is the optimal user set and μ is the greedily selected user set with cardinality $k=|\mu|$.

For the *Estimation()* with non-submodular utility, we introduce the submodularity ratio in Definition 1, which characterizes how close a non-submodular utility function $g(\mu)$ is to being submodular.

***Definition** 1: The submodularity ratio $\gamma\in(0,1)$ of a non-negative and non-submodular set function $g()$ is defined as*

$$[g(\mu_1\cup u)-g(\mu_1)]\geq\gamma\cdot[g(\mu_2\cup u)-g(\mu_2)], \qquad (10)$$

where $\forall\mu_1\subseteq\mu_2, u\notin\mu_2$.

Same with the above analysis for submodular utility, the *Estimation()* can be formulated as a variant of non-submodular maximization problem with a knapsack constraint. By using the submodularity ratio γ of $g()$, our offline greedy algorithm with non-submodular utility can achieve the $(1-e^{-\gamma/2})$ approximation ratio,[7] i.e., $g(\mu)\geq(1-e^{-\gamma/2})\cdot g(OPT)$. Before starting the proof, we first present some useful properties of the non-submodular functions.[8]

***Lemma** 1: Given the submodularity ratio γ of $g()$, we have* $g(\mu_2)-g(\mu_1)\leq\frac{1}{\gamma}\sum_{u\in\mu_2\setminus\mu_1}[g(\mu_1\cup u)-g(\mu_1)],\forall\mu_1\subseteq\mu_2$.

Then, we prove that there is an improvement towards the optimal solution in each selection.

***Lemma** 2: For each selection u_i, we have* $g(\mu_{i+1})-g(\mu_i)\geq\frac{c_{u_i}}{B}(g(OPT)-g(\mu_i))$.

***Theorem** 2: The offline greedy selection with non-submodular utility can achieve an approximation ratio of* $(1-e^{-\gamma/2})$.

proof. When $k=1$, the theorem is true since we will automatically select the best one. When $k>1$, according to

[7]Combined with partial enumeration, a modified greedy algorithm can achieve a $(1-1/e)$ approximation ratio [26]. However, it is too computationally expensive for real world applications (the computation cost is $O(n^5)$ in general).

[8]The proofs of Lemmas 1, 2, and 3 are standard, which are included in the appendix for the sake of completeness.

Lemma 2, we have

$$\begin{aligned}g(\mu) &= g(\mu_{k-1}) + [g(\mu) - g(\mu_{k-1})]\\ &\geq g(\mu_{k-1}) + \gamma \cdot \frac{c_{u_k}}{B} \cdot [g(OPT) - g(\mu_{k-1})]\\ &= (1 - \gamma \cdot \frac{c_{u_k}}{B})g(\mu_{k-1}) + \gamma \cdot \frac{c_{u_k}}{B} \cdot g(OPT)\\ &\geq [1 - \prod_{i=1}^{k}(1 - \gamma \cdot \frac{c_{u_i}}{B})]g(OPT). \end{aligned} \tag{11}$$

Similar with the Theorem 1 in [25], we consider three cases:
Case 1: If $g(v^*) > 1/2 \cdot g(OPT)$, we achieve $g(\mu^*) \geq g(v^*) > 1/2 \cdot g(OPT)$ according to line 8 in Algorithm 1, where $v^* = \arg\max_{u_i \in U'} \frac{f(\{u_i\})}{p_i}$ and $\mu^* = \arg\max_{\mu^* \in \{\{v^*\}, \mu\}} f(\mu^*)$.
Case 2: If $g(v^*) \leq 1/2 \cdot g(OPT)$ and $\sum_{u \in \mu} c_u \leq 1/2 \cdot B$, we have $c_{u'} > 1/2 \cdot B, \forall u' \in U \setminus \mu$ and $|OPT \setminus \mu| \leq 1$. Thus, we have $g(OPT \setminus \mu) \leq 1/2 \cdot g(OPT)$. Due to the non-submodularity of $g()$, we have $g(OPT \cap \mu) - g(\emptyset) \geq \gamma[g(OPT) - g(OPT \setminus \mu)]$. Thus, we have $g(\mu) \geq g(OPT \cap \mu) \geq \gamma/2 \cdot g(OPT)$.
Case 3: If $g(v^*) \leq 1/2 \cdot g(OPT)$ and $\sum_{u \in \mu} c_u > 1/2 \cdot B$, according to Eq. 11, we have

$$\begin{aligned}g(\mu) &\geq [1 - \prod_{i=1}^{k}(1 - \gamma \cdot \frac{c_{u_i}}{2\sum_{u \in \mu} c_u})]g(OPT)\\ &\geq [1 - (1 - \frac{\gamma}{2k})^k]g(OPT) \geq (1 - e^{-\gamma/2})g(OPT),\end{aligned} \tag{12}$$

where the second inequality uses the fact that $1 - \prod_{i=1}^{n}(1 - \frac{c_i}{\sum_{i=1}^{n} c_i})$ achieves the minimum value when $c_1 = \ldots = c_n$.

In all cases, we have $g(\mu^*) \geq \min\{1/2, \gamma/2, 1 - e^{-\gamma/2}\}$. For $0 < \gamma < 1$, we have $1/2 > \gamma/2 > 1 - e^{-\gamma/2}$, and thus we obtain $g(\mu^*) \geq (1 - e^{-\gamma/2}) \cdot g(OPT)$. In fact, as shown in Eq. 10, $g()$ becomes submodular when $\gamma = 1$ and its approximation ratio is equal to the one of $f()$. □

B. Analysis on Segmented() Strategy

Then, we analyze the Algorithms 2, i.e., the *Segmented()*. For the submodular utility, 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 scenario, with the determined k, our segmented user recruitment strategy can be proved to achieve an expected competitive ratio of $(1 - e^{-1})/7$ [15], denoted as $E\{f(\mu)\} \geq (1 - e^{-1})/7 \cdot f(OPT\_k)$, where OPT_k is the optimal set under the cardinality k.

Similarly, we use the submodularity ratio γ in Definition 1 to deal with the non-submodular utility function $g()$, where the competitive ratio of our *Segmented()* is presented in the following theorem.

***Theorem** 3: With the determined k, the Segmented() algorithm achieves an expected competitive ratio of $\gamma^2(1 - e^{-1})/7$.*

The basic idea of the following analysis is to use the submodularity ratio to approximate the non-submodular $g()$ to a submodular function, and then further calculate its competitive ratio. Recall that the optimal user set is OPT with the cardinality k. Considering that the user set of the i-th segment is U_i, $1 \leq i \leq k$, we define that $\nu = \bigcup_{i=1}^{k}\{U_i \cap OPT\}$. We first show that selecting one user from each segment means that we can really select many optimal users, as follows:

***Lemma** 3: The expected $|\nu|$ is at least $k(1 - e^{-1})$.*

Then, we show that partial optimal users can ensure a good approximation of the optimal solution:

***Lemma** 4: For a random subset ν of OPT, the expected value of $g(\nu)$ is at least $\gamma \cdot \frac{|\nu|}{k} \cdot g(OPT)$.*

proof. Let $OPT = \{u_1, u_2, \ldots, u_k\}$, $G_i = g(\{u_1, u_2, \ldots, u_i\})$, $G_0 = 0$. Define $D_i = G_i - G_{i-1} = g(\{u_1, u_2, \ldots, u_i\}) - g(\{u_1, u_2, \ldots, u_{i-1}\})$. Let $(u_1^*, u_2^*, u_3^*, \ldots, u_i^*)$ denote the cyclic permutation of $(u_1, u_2, u_3, \ldots, u_i)$, where $u_1^* = u_i, u_2^* = u_1, u_3^* = u_2, \ldots, u_i^* = u_{i-1}$. Note that G_i is equal to the expectation of $g(\{u_2^*, u_3^*, \ldots, u_{i+1}^*\})$ since $\{u_2^*, u_3^*, \ldots, u_{i+1}^*\}$ is equal to $\{u_1, u_2, \ldots, u_i\}$. What's more, G_i is also equal to the expectation of $g(\{u_1^*, u_2^*, \ldots, u_i^*\})$ because the sequence $(u_1^*, u_2^*, \ldots, u_i^*)$ has the same distribution as that of $(u_1, u_2, \ldots, u_i)$. Therefore, $D_i = g(\{u_1^*, u_2^*, \ldots, u_i^*\}) - g(\{u_2^*, u_3^*, \ldots, u_i^*\})$. Similarly, $D_{i+1} = g(\{u_1^*, u_2^*, \ldots, u_{i+1}^*\}) - g(\{u_2^*, u_3^*, \ldots, u_{i+1}^*\})$. According to Definition 1, we can get $D_i \geq \gamma \cdot D_j, \forall j > i$.

Afterwards, we have $g(OPT) = G_k = D_1 + D_2 + \ldots + D_k$ and $g(\nu) = G_{|\nu|} = D_1 + D_2 + \ldots + D_{|\nu|}$. Considering the submodularity ratio γ, we can get

$$\begin{aligned}k \cdot g(\nu) &= k(D_1 + D_2 + \ldots + D_{|\nu|})\\ &\geq \gamma(D_1 + D_2 + \ldots + D_{|\nu|})\\ &\quad + \gamma(D_2 + D_3 + \ldots + D_{|\nu|+1})\\ &\quad + \ldots + \gamma(D_k + D_1 + \ldots + D_{|\nu|-1})\\ &= \gamma \cdot |\nu| \cdot (D_1 + D_2 + \ldots + D_k)\\ &= \gamma \cdot |\nu| \cdot g(OPT).\end{aligned} \tag{13}$$

Therefore, $g(\nu) \geq \gamma \cdot \frac{|\nu|}{k} \cdot g(OPT)$. □

Define μ_k as the selected user set after k segments, we then prove that the local non-submodular optimization in each segment can lead to a global approximation.

***Lemma** 5: The expected $g(\mu_k)$ is at least $\gamma^2 \cdot \frac{|\nu|}{7k} \cdot g(OPT)$.*

proof. Recall $\nu = \{u_1, u_2, \ldots, u_{|\nu|}\}$ and $u_i \in U_{h_i} \cap OPT$ for $1 \leq i \leq |\nu|$ and $1 \leq h_i \leq k$. Define $\triangle_j := g(\mu_j) - g(\mu_{j-1})$ as the gain of the j-th segment. With probability $1/e$, we can choose the best u_j for the j-th segment, which maximizes $g(\mu_j)$ with the fixed μ_{j-1}, thus

$$E(\triangle_{h_i}) \geq E[g(\mu_{h_{i-1}} \cup u_i) - g(\mu_{h_{i-1}})]/e. \tag{14}$$

In order to prove Lemma 5 by contradiction, we first assume that $E[g(\mu_k)] < \gamma^2 \cdot \frac{|\mu_k|}{7k} \cdot g(OPT)$. Define $\psi = \{u_i, u_{i+1}, \ldots, u_{|\nu|}\}$. By Lemma 1 and monotonicity of $g(\mu)$,[9]

$$\begin{aligned}g(\psi) &\leq g(\psi \cup \mu_{h_{i-1}}) \leq g(\mu_{h_{i-1}})\\ &\quad + \frac{1}{\gamma}\sum_{j=i}^{|\nu|}[g(\mu_{h_{i-1}} \cup u_j) - g(\mu_{h_{i-1}})],\\ E[g(\psi)] &\leq E[g(\mu_{h_{i-1}})]\\ &\quad + \frac{1}{\gamma}\sum_{j=i}^{|\nu|}E[g(\mu_{h_{i-1}} \cup u_j) - g(\mu_{h_{i-1}})].\end{aligned} \tag{15}$$

Note that users are coming in random order with uniform distribution of utilities, we consider that $\forall u_i \in \psi$ is the same

[9] We won't recruit the users with no marginal revenue. Thus, $g(\mu)$ can actually be seen as non-decreasing.

in expectation, thus

$$E[g(\psi)] \leq E[g(\mu_{h_{i-1}})] + \frac{1}{\gamma}(|\nu| - i + 1)E[g(\mu_{h_{i-1}} \cup u_i) - g(\mu_{h_{i-1}})]. \quad (16)$$

Considering Eqs. 14 and 16, we have

$$E(\triangle_{h_i}) \geq E[g(\mu_{h_{i-1}} \cup u_i) - g(\mu_{h_{i-1}})]/e \geq \gamma \frac{E[g(\psi)] - E[g(\mu_{h_{i-1}})]}{e(|\nu| - i + 1)}. \quad (17)$$

According to Lemma 4, we have $E[g(\psi)] \geq \gamma \cdot \frac{|\psi|}{k} \cdot f(OPT) = \gamma \cdot \frac{|\nu|-i+1}{k} \cdot f(OPT)$. Since we assume that $E[g(\mu_k)] < \gamma^2 \cdot \frac{|\nu|}{7k} \cdot g(OPT)$, and $g()$ is monotone, we have

$$E(\triangle_{h_i}) \geq \gamma \frac{E[g(\psi)] - E[g(\mu_{h_{i-1}})]}{e(|\nu| - i + 1)} \geq \frac{\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^3}{7ek(|\nu| - i + 1)} g(OPT). \quad (18)$$

To eliminate i, we add Eq. 18 for $1 \leq i \leq \lceil |\nu|/2 \rceil$ and obtain

$$\sum_{i=1}^{\lceil |\nu|/2 \rceil} E(\triangle_{h_i}) \geq \lceil \frac{|\nu|}{2} \rceil \frac{\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^3}{7ek} g(OPT) \sum_{i=1}^{\lceil |\nu|/2 \rceil} \frac{1}{|\nu| - i + 1}. \quad (19)$$

Since $\sum_{j=a}^{b} 1/j \leq \ln b/(a+1), 1 < a \leq b$, we have

$$\sum_{i=1}^{\lceil |\nu|/2 \rceil} E(\triangle_{h_i}) \geq \lceil \frac{|\nu|}{2} \rceil \frac{\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^3}{7ek} g(OPT) \ln \frac{|\nu|}{\lfloor |\nu|/2 \rfloor}. \quad (20)$$

Similarly, we add Eq. 18 for $1 \leq i \leq \lfloor |\nu|/2 \rfloor$, and then obtain

$$\sum_{i=1}^{\lfloor |\nu|/2 \rfloor} E(\triangle_{h_i}) \geq \lfloor \frac{|\nu|}{2} \rfloor \frac{\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^3}{7ek} g(OPT) \ln \frac{|\nu|}{\lceil |\nu|/2 \rceil}. \quad (21)$$

Add Eqs. 20 and 21, since $\frac{|\nu|^2}{\lfloor |\nu|/2 \rfloor \lceil |\nu|/2 \rceil} < 4.5$, we have

$$\begin{aligned} 2E[g(\mu_k)] &\geq \sum_{1}^{\lceil |\nu|/2 \rceil} E(\triangle_{h_i}) + \sum_{1}^{\lfloor |\nu|/2 \rfloor} E(\triangle_{h_i}) \\ &\geq \frac{|\nu|\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^2}{7ek} g(OPT) \ln \frac{|\nu|^2}{\lfloor |\nu|/2 \rfloor \lceil |\nu|/2 \rceil} \\ &\geq \frac{|\nu|\gamma^2}{ek} g(OPT) - \frac{|\nu|\gamma^3}{7ek} g(OPT) \cdot \ln 4.5 \\ &= \frac{|\nu|\gamma^2}{k} g(OPT)(\frac{1}{e} - \frac{\ln 4.5}{7e}\gamma) \geq \frac{|\nu|\gamma^2}{k} g(OPT) \cdot \frac{2}{7}, \end{aligned} \quad (22)$$

which contradicts $E[g(\mu_k)] < \gamma^2 \cdot \frac{|\nu|}{7k} \cdot g(OPT)$, hence we obtain $E[g(\mu_k)] \geq \gamma^2 \cdot \frac{|\nu|}{7k} \cdot g(OPT)$. □

Finally, we prove Theorem 3, i.e., *Segmented()* algorithm achieves an expected competitive ratio of $\gamma^2(1 - e^{-1})/7$.
proof.

Recall $\nu = \bigcup_{i=1}^{k} \{U_i \cap OPT\}$. According to Lemma 3, the expected value of $|\nu| \geq (1 - e^{-1})k$, i.e., $\sum_{i=1}^{k} Pr[|\nu| = i] \cdot |\nu| \geq (1 - e^{-1})k$. According to Lemma 5, we also know that $g(\mu_k) \geq \gamma^2 \cdot \frac{|\nu|}{7k} \cdot g(OPT)$, i.e., $\sum_{v \in \mathcal{V}} Pr[g(\mu) = v||\mu| = |\nu|] \cdot v \geq \gamma^2 \cdot \frac{m}{7k} \cdot g(OPT)$, where $\mathcal{V}$ denotes the set of values that our algorithm can get. Therefore, we obtain

$$\begin{aligned} E[g(\mu_k)] &= \sum_{i=1}^{k} E[g(\mu_k)||\nu| = i] Pr[|\nu| = i] \\ &\geq \sum_{i=1}^{k} \gamma^2 \cdot \frac{i}{7k} \cdot g(OPT) \cdot Pr[|\nu| = i] \\ &= \frac{\gamma^2}{7k} \cdot g(OPT) \cdot E[|\nu|] \geq \frac{\gamma^2(1 - e^{-1})}{7} \cdot g(OPT). \end{aligned} \quad (23)$$

Actually, when the submodularity ratio $\gamma = 1$, the competitive ratio becomes $1 - e^{-1}/7$, which is the same as the one of submodular function $f()$. □

C. Analysis on Online User Recruitment Strategies

With the analysis on *Estimation()* and *Segmented()*, we then give the competitive ratio of the segmented online user recruitment strategy with (non-) submodular utility:

Theorem 4: *The segmented online user recruitment strategy achieves an approximation ratio of* $\gamma^2(1 - e^{-1})(1 - e^{-\gamma/2})/7$.

Proof.

1) As discussed above, OPT_k is the optimal set under the cardinality k, and OPT is the global optimal set without cardinality constraints. Since μ is the greedy recruited user set with the same cardinality k, we obtain the following inequality:

$$\begin{aligned} E\{g(\mu)\} &\geq \frac{\gamma^2(1 - e^{-1})}{7} g(OPT\_k) \\ &\geq \frac{\gamma^2(1 - e^{-1})}{7} g(\mu) \\ &\geq \frac{\gamma^2(1-e^{-1})(1-e^{-\gamma/2})}{7} g(OPT). \end{aligned} \quad (24)$$

2) In the online scenario, we cannot exactly obtain the users who will participate in the MCS campaign in advance. Under Assumption 1, we construct the simulated user set U' as a replacement of the real user set U, and greedily select μ_U' from U' to estimate μ. Therefore, we have $g(\mu\_U') \approx g(\mu)$ and $E\{g(\mu)\}$ achieves an approximation ratio of $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$. □

Actually, the dynamic online user recruitment is an extension of the above 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 that we provide some intuitive examples in Fig. 9: after one user has been recruited, if the estimated k in the dynamic strategy is the same as in the segmented strategy, the dynamic strategy has more participating users than the segmented strategy, since it needs to skip over some users to the next segment and thus the dynamic strategy will expectedly outperform the segmented strategy. Similarly, if the estimated ks are different, it means

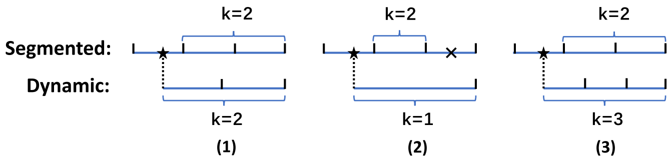


Fig. 9. Different dynamic and segmented cases.

that the previously recruited users cost too much/little, and the segmented strategy cannot correct the errors and make use of new information in time, which leads to a worse performance.

Finally, we give a brief analysis on the computation complexity. In *Estimation()*, the total number of users is n, thus the loop in lines 3-7 of Algorithm 1 is performed at most n times. Considering that the complexity of the main step in line 4 is $O(n)$, we obtain that the complexity of *Estimation()* is $O(n^2)$. Similarly, the loops in *Segmented()* can be performed at most n times, thus it achieves a complexity $O(n)$. Consequently, as the dynamic iterations of *Estimation()* and *Segmented()* for at most n times, the computation complexity of our dynamic strategy is $O(n^3)$.

VI. Reverse Auction-Based Truthful Pricing for Online User Recruitment

In general, the organizers and users in MCS are rational and selfish. From the user side, the organizer should provide proper 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[10] in MCS to simultaneously satisfy the truthfulness and individual rationality [3], [36], [37], where users bid first according to their costs and then the organizers determine the true payments. However, the existing mechanisms determine the prices for the 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 constraint.

In our proposed dynamic and segmented strategies, the user recruitment in segments can actually be seen as the ordering of users, and thus a reverse auction-based pricing mechanism can be easily modified, as summarized in Algorithm 4. The basic idea is to pay a recruited user according to the other users' bids, and the payment equals the weighted cost bid by the second "best" user. Specifically, we first deal with the special case, $i.e.$, the real number of participating users is larger than our estimated n', where we will recruit the first extra user who we can afford and pay him all of the remaining budget[11] (line 1-2), in order to ensure the truthfulness. For the user recruitment in each segment, we use the bids observed from the first l_{ob} users and determine a price for the recruited user (line 3-4), denoted as $p_i = b_i \cdot \delta_{u_i}/\varepsilon$. Note that the total payments (instead of bids or costs) of recruited users are constrained by B, and thus we only recruit the users we can afford (line 5-6). In this way, the pricing mechanism has been skillfully added into the online user recruitment strategy

[10]Note that the pricing is almost the same as the existing incentive mechanisms, while it pays more attention on determining the rewards but not encouraging the user participation.

[11]The remaining budget is usually very small, since our proposed dynamic strategy will re-adjust after recruiting a new user.

Algorithm 4 Reverse Auction-Based Pricing

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

In *Segmented(), u_i is coming*:

1: **if** $i > n'$ **and** $\sum_{u_j \in \mu} p_j + b_i \leq B$ **then**

2: Recruit u_i with pricing $B - \sum_{u_j \in \mu} p_j$;

3: **else if** $i >$ *segmentID* $* l + l_{ob}$ **and** $\delta_{u_i} \geq \varepsilon$ **then**

4: $p_i = b_i \cdot \delta_{u_i}/\varepsilon$;

5: **if** $\sum_{u_j \in \mu} p_j + p_i \leq B$ **then**

6: Recruit u_i with pricing p_i;

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

Fig. 10. An example of trajectories and grids.

without much extra computation, and the truthfulness and individual rationality will be proved in the appendix.

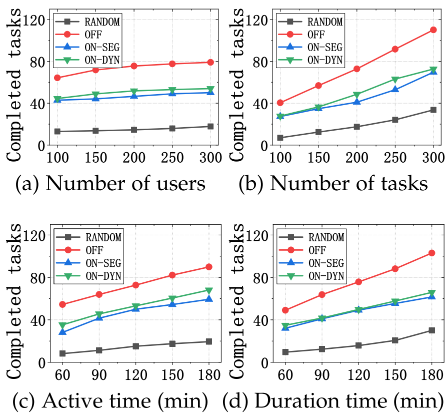
VII. Performance Evaluation

A. Data Sets & Settings

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

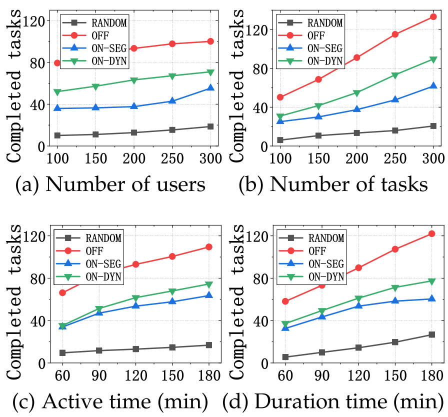
- *Feeder* [39] contains four kinds of data, $i.e.$, the cellphone CDR data, smartcard data, taxicab GPS data, and bus GPS data collected from Shenzhen, China. We select 300 taxi traces as the participating users, each of whom has continuous GPS records collected from the same periods of time, 8:00-18:00.
- *Shanghai* contains the GPS data collected from taxis and trucks in Shanghai, China. Similar to *Feeder*, we select 310 traces as users. Note that nearly half of them were collected from trucks, which have the more regular mobilities.
- *GeoLife* [35] was collected from phones carried by 182 users, which recorded a broad range of users' outdoor movements. It contains 17000+ trajectories and has a total duration of 50000+ hours, from which we select 727 traces. Compared with *Feeder* and *Shanghai*, *GeoLife* has the fine-grained trajectories but users may stay at the same place 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 5-hour data to train the mobility prediction model and construct the historical data for user recruitment strategies. Then, the MCS campaign begins at 13:00 and participating users move according to the traces. For the cooperative willingness, we randomly generate the matrices of relations between users. The tasks will be generated in grids with the uniform duration time. The users will participate in the MCS campaign in real time, with the uniform costs and active time (working time). If we recruit one user, he will perform the tasks in the grids he will pass by during his active time (working time).



(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

Fig. 11. Main results of *Feeder* with submodular utility.



(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

Fig. 12. Main results of *Shanghai* with submodular utility.

B. Comparison Algorithms & Metrics

We mainly compare our proposed online user strategies (referred to as "ON-SEG" for segmented strategy and "ON-DYN" for dynamic strategy) with the following algorithms:

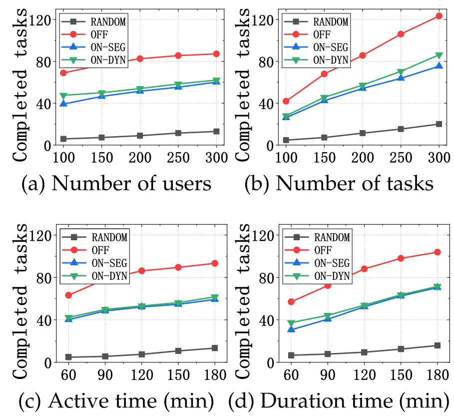
- RANDOM, which randomly recruits users from all participating users until the budget is exhausted.
- OFF, which greedily recruits the users who have the largest contribution/cost ratio, *i.e.*, $\arg\max_{u_i \in U} \frac{f(\mu \cup \{u_i\}) - f(\mu)}{c_i}$, in an offline manner.
- OPT, which exhaustively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

Obviously, OPT costs a lot in the submodular user recruitment problem, and we implement it to verify our bound in Section IV.D. 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 recruited users 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.*, $\sum_{u_i \in \mu}(p_i - c_i) / \sum_{u_i \in \mu} c_i$.

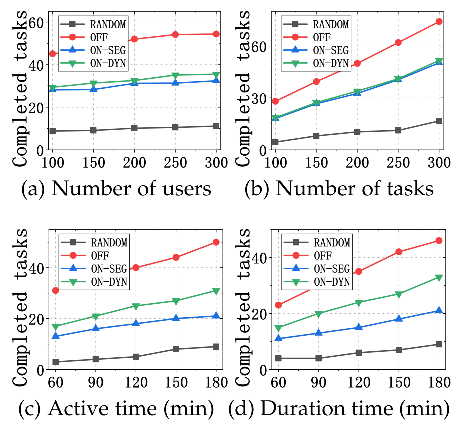
C. Evaluation Results

Completed tasks: We first illustrate the results in terms of the main metric, *i.e.*, the number of completed tasks, as shown in Figs. 11, 12, and 13 with submodular utility, and



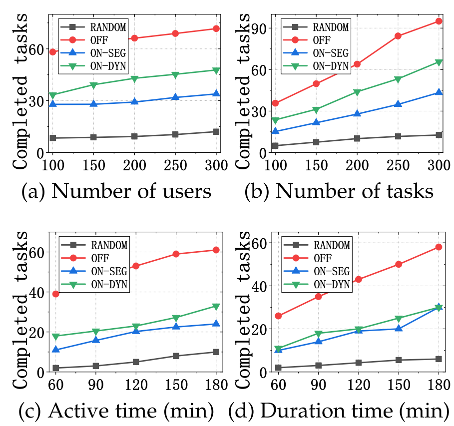
(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

Fig. 13. Main results of *GeoLife* with submodular utility.



(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

Fig. 14. Main results of *Feeder* with non-submodular utility.



(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

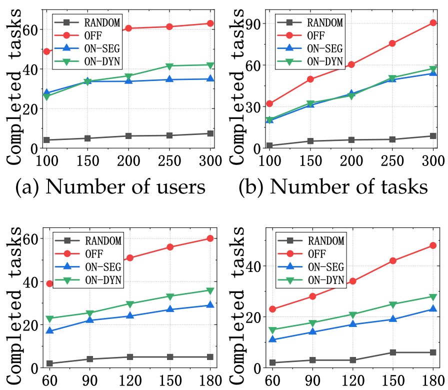
Fig. 15. Main results of *Shanghai* with non-submodular utility.

Figs. 14, 15, and 16 with non-submodular utility. In order to provide a comprehensive evaluation, we change the number of participating users, the number of tasks, the average active time of users, and the average duration time of tasks respectively, while keeping the others fixed (i.e., the number of users and tasks is 200, and the average active time and duration time is 120). We set the budget to 200 units and the average cost of users is 20. The results over three data sets have the similar tendencies and show that 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

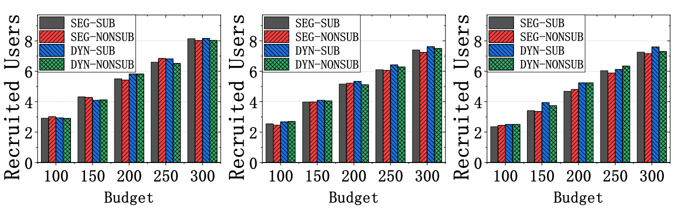
TABLE II
COMPLETED TASKS AND COMPETITIVE RATIO

	Feeder with submodular utility				*Shanghai* with submodular utility				*GeoLife* with submodular utility			
Budget	ON-SEG	ON-DYN	OPT	Ratio	ON-SEG	ON-DYN	OPT	Ratio	ON-SEG	ON-DYN	OPT	Ratio
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
	Feeder with non-submodular utility				*Shanghai* with non-submodular utility				*GeoLife* with non-submodular utility			
Budget	ON-SEG	ON-DYN	OPT	Ratio	ON-SEG	ON-DYN	OPT	Ratio	ON-SEG	ON-DYN	OPT	Ratio
100	14.5	14.6	51.1	0.2857	10.4	18.9	66.3	0.2850	17.9	20	63.1	0.3169
150	26	27.8	69.7	0.3988	20.1	40.2	88.4	0.4547	30.2	32.1	81.3	0.3948
200	39.2	40.3	85.4	0.4718	28.4	50.5	106.3	0.4750	42.4	44.5	99.2	0.4485
250	49.4	50.5	100.8	0.5009	38.6	61.1	120	0.5091	54	57.9	110.1	0.5258
300	57.2	58.5	113.2	0.5167	46.1	70.8	131.1	0.5400	62.9	66.2	122.5	0.5404



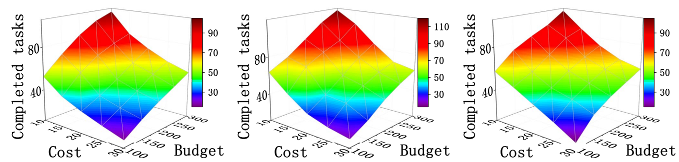
(a) Number of users (b) Number of tasks
(c) Active time (min) (d) Duration time (min)

Fig. 16. Main results of *GeoLife* with non-submodular utility.



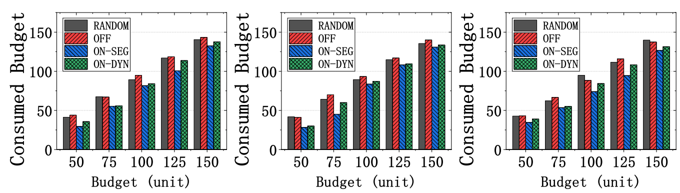
(a) *Feeder* (b) *Shanghai* (c) *GeoLife*

Fig. 17. Recruited users of *Feeder*, *Shanghai*, and *GeoLife*.



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

Fig. 18. Budget and cost of *Feeder*, *Shanghai*, and *GeoLife*.



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

Fig. 19. Consumed budget of *Feeder*, *Shanghai*, and *GeoLife*.

ON-DYN can correct the estimation errors and make use of the newly obtained information. Moreover, comparing the subfigure (a) and (b) of Figs. 9-14, we find that the growth rates over users are lower than tasks. The reason is that we have already recruited the effective users to perform tasks, and thus more users cannot improve the performances significantly. In addition, compared with *Feeder* and *GeoLife*, our strategies perform better in *Shanghai*, since the traces are collected from trucks in Shanghai City, which has a stronger regularity so that our dynamic strategies can achieve more accurate predictions and make the timely adjustments.

Moreover, the algorithms with submodular function complete more tasks than the ones with non-submodular function, since the non-submodular case further considers the cooperative willingness, which leads to the decline of the probability of completing tasks. We also compare the numbers of recruited users in SEG and DYN in two cases. As shown in Fig. 17, the numbers are very close, which shows that the worse performances are not caused by users but the non-submodular function. Actually, the trends of the submodular and non-submodular results are very similar, which also shows that our proposed algorithms work well with the non-submodular utility.

Budget: We then consider the main constraints in this paper, $i.e.$, the budget and cost. We set other variables fixed, then change the budget from 100 to 300 and change the average cost of users from 10 to 30. As shown in Fig. 18, 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 and illustrate the consumed budget over three data sets, as shown in Fig. 19. Obviously, the OFF and RANDOM consume more budget, since they recruit users in the offline manner, until their budget is 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 timely adjustments.

We also illustrate the competitive ratio of our proposed strategies in Table II with (non-) submodular utility. Under different budget constraints, our ON-DYN can achieve a 30%-50% competitive ratio of the optimal results, which is far much higher than $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$ proved in Section IV.D. With the increase in budget, our ON-DYN even achieves a better competitive ratio, since we can recruit more

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

effective users and the results are close to OPT. Moreover, Table II also shows the trends that more users can complete more tasks, but the ratio of task completion is relatively low due to the large ranges of movements.

Pricing: Finally, we evaluate the performance of the online pricing mechanism embedded into our online strategy. We illustrate the overpayment ratio of the pricing mechanism in Table III. With the increase of budget, 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).

VIII. 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 two constraints jointly, we first estimate the number of recruited users and then recruit users in segments. Moreover, using the mobility prediction and cooperative willingness as examples, we extend the segmented strategy with a general (non-) submodular utility function, and prove that the competitive ratio is $\gamma^2(1-e^{-1})/7$ (where γ is the submodularity ratio). In order to correct estimation errors and utilize newly obtained information, we further present a dynamic re-estimation after recruiting every new user, which achieves a competitive ratio of $\gamma^2(1-e^{-1})(1-e^{-\gamma/2})/7$. Finally, we conduct a truthful pricing mechanism embedded into the dynamic strategy. Extensive evaluations on three real-world data sets have verified the effectiveness of our proposed strategies.

REFERENCES

[1] W. Liu, Y. Yang, E. Wang, and J. Wu, “Dynamic user recruitment with truthful pricing for mobile CrowdSensing,” in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Jul. 2020, pp. 1–10.

[2] R. Ganti, F. Ye, and H. Lei, “Mobile crowdsensing: Current state and future challenges,” *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011.

[3] 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.

[4] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, “User recruitment for mobile crowdsensing over opportunistic networks,” in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*. IEEE, 2015, pp. 2254–2262.

[5] M. Karaliopoulos, I. Koutsopoulos, and L. Spiliopoulos, “Optimal user choice engineering in mobile crowdsensing with bounded rational users,” in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2019, pp. 1054–1062.

[6] W. Liu, Y. Yang, E. Wang, Z. Han, and X. Wang, “Prediction based user selection in time-sensitive mobile crowdsensing,” in *Proc. 14th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2017, pp. 1–9.

[7] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, “An efficient prediction-based user recruitment for mobile crowdsensing,” *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 16–28, Jan. 2018.

[8] Y. Yang, W. Liu, E. Wang, and J. Wu, “A prediction-based user selection framework for heterogeneous mobile CrowdSensing,” *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2460–2473, Nov. 2019.

[9] L. Pu, X. Chen, J. Xu, and X. Fu, “Crowd foraging: A QoS-oriented self-organized mobile crowdsourcing framework over opportunistic networks,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 848–862, Apr. 2017.

[10] D. Zhao, X.-Y. Li, and H. Ma, “Budget-feasible online incentive mechanisms for crowdsourcing tasks truthfully,” *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 647–661, Apr. 2016.

[11] 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 Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2589–2603, Nov. 2019.

[12] J. Li, J. Wu, and Y. Zhu, “Selecting optimal mobile users for long-term environmental monitoring by crowdsourcing,” in *Proc. IEEE/ACM 27th Int. Symp. Qual. Service (IWQoS)*. IEEE, 2019, pp. 1–10.

[13] H. Chen, B. Guo, Z. Yu, and Q. Han, “CrowdTracking: Real-time vehicle tracking through mobile crowdsensing,” *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7570–7583, Oct. 2019.

[14] J. Preater, “On multiple choice secretary problems,” *Math. Oper. Res.*, vol. 19, no. 3, pp. 597–602, Aug. 1994.

[15] M. Bateni, M. Hajiaghayi, and M. Zadimoghaddam, “Submodular secretary problem and extensions,” *ACM Trans. Algorithms*, vol. 9, no. 4, p. 32, 2013.

[16] F. Zhang, B. Jin, H. Liu, Y.-W. Leung, and X. Chu, “Minimum-cost recruitment of mobile crowdsensing in cellular networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.

[17] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, “QoI-aware multitask-oriented dynamic participant selection with budget constraints,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4618–4632, Nov. 2014.

[18] E. Wang *et al.*, “Deep learning-enabled sparse industrial CrowdSensing and prediction,” *IEEE Trans. Ind. Informat.*, early access, Oct. 6, 2020, doi: 10.1109/TII.2020.3028616.

[19] P. Micholia, M. Karaliopoulos, I. Koutsopoulos, L. M. Aiello, G. De Francisci Morales, and D. Quercia, “Incentivizing social media users for mobile crowdsourcing,” *Int. J. Hum.-Comput. Stud.*, vol. 102, pp. 4–13, Jun. 2017.

[20] J. Wang *et al.*, “Multi-task allocation in mobile crowd sensing with individual task quality assurance,” *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 2101–2113, Sep. 2018.

[21] J. Wang, F. Wang, Y. Wang, D. Zhang, B. Lim, and L. Wang, “Allocating heterogeneous tasks in participatory sensing with diverse participant-side factors,” *IEEE Trans. Mobile Comput.*, vol. 18, no. 9, pp. 1979–1991, Sep. 2019.

[22] X. Wang, R. Jia, X. Tian, and X. Gan, “Dynamic task assignment in crowdsensing with location awareness and location diversity,” in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, Apr. 2018, pp. 2420–2428.

[23] H. Li, T. Li, W. Wang, and Y. Wang, “Dynamic participant selection for large-scale mobile crowd sensing,” *IEEE Trans. Mobile Comput.*, vol. 18, no. 12, pp. 2842–2855, Dec. 2019.

[24] A. Krause and C. Guestrin, *A Note on the Budgeted Maximization of Submodular Functions*. Pittsburgh, PA, USA: Carnegie Mellon Univ., 2005.

[25] H. Lin and J. Bilmes, “Multi-document summarization via budgeted maximization of submodular functions,” in *Proc. Hum. Lang. Technol, Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 912–920.

[26] 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.

[27] C. Qian, J.-C. Shi, Y. Yu, and K. Tang, “On subset selection with general cost constraints,” in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2613–2619.

[28] A. A. Bian, J. M. Buhmann, A. Krause, and S. Tschiatschek, “Guarantees for greedy maximization of non-submodular functions with applications,” in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 498–507.

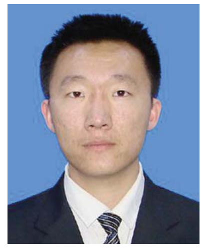
[29] T. S. Ferguson, “Who solved the secretary problem?” *Stat. Sci.*, vol. 4, no. 3, pp. 282–289, Aug. 1989.

[30] M. H. Cheung, F. Hou, J. Huang, and R. Southwell, “Distributed time-sensitive task selection in mobile crowdsensing,” *IEEE Trans. Mobile Comput.*, vol. 20, no. 6, pp. 2172–2185, Jun. 2021.

[31] D. Zhang, T. He, F. Zhang, and C. Xu, "Urban-scale human mobility modeling with multi-source urban network data," *IEEE/ACM Trans. Netw.*, vol. 26, no. 2, pp. 671–684, Apr. 2018.
[32] C. Chen *et al.*, "Crowddeliver: Planning city-wide package delivery paths leveraging the crowd of taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1478–1496, Jun. 2017.
[33] U. Feige, "A threshold of ln n for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, Jul. 1998.
[34] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2011, pp. 1082–1090.
[35] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proc. 10th Int. Conf. Ubiquitous Comput. (UbiComp)*. New York, NY, USA: ACM, 2008, pp. 312–321.
[36] M. Tang, H. Pang, S. Wang, L. Gao, J. Huang, and L. Sun, "Multi-dimensional auction mechanisms for crowdsourced mobile video streaming," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2062–2075, Oct. 2018.
[37] Y. Zhu, S. D. Fu, J. Liu, and Y. Cui, "Truthful online auction toward maximized instance utilization in the cloud," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2132–2145, Oct. 2018.
[38] R. Myerson, "Optimal auction design," *Math. Oper. Res.*, vol. 6, pp. 58–73, Feb. 1981.
[39] D. Zhang, J. Zhao, F. Zhang, R. Jiang, and T. He, "Feeder: Supporting last-mile transit with extreme-scale urban infrastructure data," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw.* New York, NY, USA: ACM, Apr. 2015, pp. 226–237.

Wenbin Liu (Member, IEEE) received the B.S. degree in physics and the Ph.D. degree in computer science and technology from Jilin University, China, in 2012 and 2020, respectively. He was a Visiting Ph.D. Student with the Department of Wireless Networks and Multimedia Services, Telecom SudParis/Institut Mines-Telecom, France. He is currently a Post-Doctoral Researcher in the Dingxin Scholar Program with the College of Computer Science and Technology. His research interests include mobile CrowdSensing, mobile computing, and ubiquitous computing.

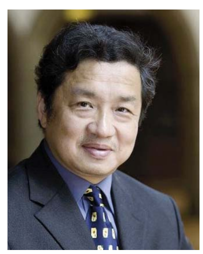
Yongjian Yang received the B.E. degree in automatization from the Jilin University of Technology, Changchun, Jilin, China, in 1983, the M.E. degree in computer communication from the Beijing University of Posts and Telecommunications, Beijing, China, in 1991, and the Ph.D. degree in software and theory of computer from Jilin University, Changchun, in 2005. He is currently a Professor and a Ph.D. Supervisor with Jilin University, the Director of the Key Laboratory under the Ministry of Information Industry, and the Standing Director of the Communication Academy. His research interests include network intelligence management, wireless mobile communication and services, and wireless mobile communication.



En Wang (Member, IEEE) received the B.E. degree in software engineering and the M.E. and Ph.D. degrees in computer science and technology from Jilin University, Changchun, in 2011, 2013, and 2016, respectively. He is currently an Associate Professor with the Department of Computer Science and Technology, Jilin University. His current research focuses on the efficient utilization of network resources, scheduling and drop strategy in terms of buffer-management, and energy-efficient communication between human-carried devices and mobile crowdsensing.

Hengzhi Wang received the bachelor's degree in software engineering from Jilin University, Changchun, China, in 2017, where he is currently pursuing the Ph.D. degree in computer science and technology. His current research focuses on mobile crowdsensing and privacy preserving in mobile computing.

Zihe Wang received the B.E. and Ph.D. degrees in computer science from Tsinghua University in 2011 and 2016, respectively. He was an Assistant Professor with the Shanghai University of Finance and Economics. He is currently an Assistant Professor with the Gaoling School of Artificial Intelligence, Renmin University of China. His research focuses on algorithms and mechanism design.



Jie Wu (Fellow, IEEE) served as the Chair for the Department of Computer and Information Sciences from 2009 to 2016 and an associate vice provost for international affairs from 2015 to 2017. He was a Program Director of the National Science Foundation and a Distinguished Professor with Florida Atlantic University. He serves as the Director for international affairs with the College of Science and Technology. He is currently the Director of the Center for Networked Computing and a Laura H. Carnell Professor with Temple University. He regularly publishes in scholarly journals, conference proceedings, and books. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. He was a recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award. He was the General Co-Chair of IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, ICPP 2016, and IEEE CNS 2016, and the Program Co-Chair of IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and the Chair of the IEEE Technical Committee on Distributed Processing (TCDP). He is a CCF Distinguished Speaker. He serves on several editorial boards, including IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON SERVICES COMPUTING, *Journal of Parallel and Distributed Computing*, and *Journal of Computer Science and Technology*.