ProSC+: Profit-driven Online Participant Selection in Compressive Mobile Crowdsensing

Yueyue Chen*, Deke Guo†, Ming Xu*
*College of Computer, National University of Defense Technology, Changsha, China
†Science and Technology on Information Systems Engineering Lab, National University of Defense Technology, China
Email: {yueyuechen, dekeguo, xuming}@nudt.edu.cn

***Abstract*—A mobile crowdsensing (MCS) platform motivates to employ participants from the crowd to complete sensing tasks. A crucial problem is to maximize the profit of the platform, i.e., the charge of a sensing task minus the payments to participants that execute the task. Recently, the appearance of data reconstruction method makes it possible to improve the platform's profit with a limited amount of sensing results in Compressive MCS (CMCS). However, It is of great challenge to the maximal profit for the CMCS platform, since it is hard to predict the reconstruction quality due to the dynamic features and mobility of participants. In response to such challenges, we propose two profit-driven online participant selection mechanisms for the given task model and participant model. In ProSC, the sub-profit in each slot is maximized during the sensing period of a task, by combing a statistical-based quality prediction method and a repetitive cross-validation algorithm. In ProSC+, we jointly optimize the number of required participants and their spatial distribution to further improve the converging property. Finally, we conduct comprehensive evaluations, the results indicate the effectiveness and efficiency of our mechanisms.**

I. Introduction

Mobile Crowdsensing (MCS) has attracted substantial attentions recently, due to the rapid development of smartphones and the embedded sensors [1]. There exist various MCS applications, such as city monitoring [2], smart transportation [3], emergency alarming [4], etc. In this paper, we focus on the environmental monitoring application. A typical MCS framework consists of two parts, i.e., the platform in the cloud and participants with smart devices. The platform is responsible for task publishing, data collection, information processing, and service providing. Participants can be either data providers or service consumers.

A crucial problem of MCS is to maximize the platform's profit for a given sensing task, i.e., the charge of the sensing task minus the payments to participants that execute the task. The charge is determined by the data quality of the sensing results, while the payments are related to selected participants. In this paper, we utilize the spatial-temporal coverage as an essential metric for measuring the data quality, which is popularly adopted in recent literature [5] [6] [7]. For example, given an environmental monitoring task, the platform always wants to obtain accurate and comprehensive sensing data across the entire target area. If the sensing data of a target area is low rank, we call it compressive [8]. The data reconstruction is an effective strategy to improve the data quality with a limited amount of sensing data in compressive MCS (CMCS). If we divide the target area into multiple cells, the data reconstruction process means to deduce those unsensed cells from the collected data of sensed cells. Since we take the coverage as a metric to measure the data quality, the reconstruction quality indicates the spatial-temporal coverage after the data reconstruction process. Given a data reconstruction method, the selection of sensed cells determines the reconstruction quality as well as the total payments to participants. Therefore, a proper participant selection mechanism is essential and should be carefully designed to maximize the profit of the platform. Although the participant selection problem have been discussed richly in traditional MCS system [9] [10] [11], the related research in compressive MCS is not enough.

978-1-5386-2542-2/18/$31.00 ©2018 IEEE

It is of great challenge to design the profit-driven participant selection mechanism in CMCS, since it is hard to predict the data quality due to the dynamic features and mobility of participants. 1) The dynamic features of participants indicate that the platform should determine whether or not to select a participant in a real time. The platform should predict the profit brought by the arrival participants, and select those who can bring the maximum profit. In the traditional MCS system, the profit prediction is based on the coverage increase, which can be calculated intuitively. However, in the CMCS system, the profit prediction is influenced by both the coverage increase and the reconstruction results. 2) The mobility, which is an important feature of participants in CMCS, brings uncertain spatial-temporal distribution of collected data. Planned mobile trajectories for participants would make it possible to sense more representative spatial and temporal data, and to achieve high reconstruction quality with less cost. Nevertheless, the planned trajectories are restricted within mobility constraints. The relative distributions produced by multiple restricted participant trajectories are variable in different slots, which makes the reconstruction quality prediction difficult.

In this paper, we aim to exploit dynamic and mobile participants to execute sensing tasks and maximize the profit of the platform. This is characterized as the POPS problem. In response to this essential problem, we predict the reconstruction quality based on an exponential-based method and propose to maximize the sub-profit in each slot to approximately approach the maximum profit of all slots for a given task. The linear programming method and repetitive cross-validation algorithm are used in ProSC to alternately optimize the essential parameter

λ and participant quantity. To further improve the converging property of our mechanism, we then develop an entropy-based algorithm to jointly optimize the number of employed participants and the participant distribution in each slot. Specifically, the proposed mechanisms in this paper are orthogonal with all existing data reconstruction methods [8] [12]. The combination of our mechanisms and proper reconstruction methods can further improve the reconstruction quality and maximize the platform's profit. The main contributions of the paper can be summarized as follows:

- We formulate the online participant selection problem under given task model and participant model, aiming to maximize the profit of the CMCS platform.
- We predict the reconstruction quality in compressive environmental monitoring applications with an exponential-based method.
- To tackle the proposed POPS problem, we first develop an intrinsic mechanism (ProSC) to maximize the sub-profit in each slot under random participant selection. Then we propose a distribution-aware mechanism (ProSC+) to further improve the converging property.
- We conduct extensive evaluations with a real dataset and three reconstruction methods. The evaluation results indicate the effectiveness and efficiency of our mechanisms.

II. Problem Formulation

A. The model of sensing tasks

Consider that the CMCS platform publishes a sensing task (such as the temperature monitoring), including the sensing area and the task duration. We divide the sensing area into N cells with the same size and split the task duration into M slots with the same length. For any cell, the sensed data is assumed unchanged during a slot but may vary across different slots. To reduce the payments, the platform would not employ sufficient participants to sense all cells in every slot. That is, just a part of appropriate cells are selected to be sensed by participants in each slot. Each participant is required to upload the sensed data of a cell one time during one slot.

Definition 1: (data matrix) Let $x_{ij}^{(t)}$ denote the measured result (such as temperature) of a cell at i row and j column of the sensing area in t^{th} slot, where $1{\le}i{\le}n$, $1{\le}j{\le}m$, and $1{\le}t{\le}M$. It is common to take a snapshot and stack the columns in slot t to form a column vector. Those column vectors of all slots are compiled into the columns of a larger matrix $X_{N\times M}$, where $N{=}n{\times}m$. Note that, we use x (lowercase) to indicate the spatial data matrix in each slot, and X (capital) to indicate the spatial-temporal data matrix in this paper.

Definition 2: (sensing matrix) The sensing matrix C records the collected sensing results. According to the defined data matrix X and the participant selection matrix S (which will be introduced in Section II-B; $S_{ij}{=}1$ if the cell i is sensed by participants in slot j, else $S_{ij}{=}0$), the sensing matrix can be desined as $C{=}X\circ S$, where $\circ$ denotes the element-wise product of two matrices.

Definition 3: (reconstruction quality) Suppose the reconstructed matrix can be denoted as $\widehat{X}$, it is deduced from the sensing matrix C via the data reconstruction. The reconstruction quality Q of $\widehat{X}$ is defined as the correct coverage ratio.

$$Q=\frac{\sum\limits_{i,j}\left|\bigcup\left\{\hat{X}_{ij}\,\middle|\,\frac{(\hat{X}_{ij}-X_{ij})}{X_{ij}}\le\chi\right\}\right|}{N\times M} \tag{1}$$

Definition 4: (profit) Suppose the sensing cost of different participants to execute the sensing task in a cell of a slot is the same, the payment is determined by the cost, and is denoted as c. If a cell is correct, the gain of it is defined as g, else the gain is 0. Based on the given parameters, the profit of the CMCS platform can be calculated as:

$$P{=}g{\times}(N{\times}M){\times}Q{-}c{\times}\sum_{i,j}S_{ij}$$

Note that, the data matrix X is the ground truth of the target area, which can not be got in the real-world. More details about data reconstruction can be seen in [7] [8]. Hence, it is challenging to predict the reconstruction quality or profit.

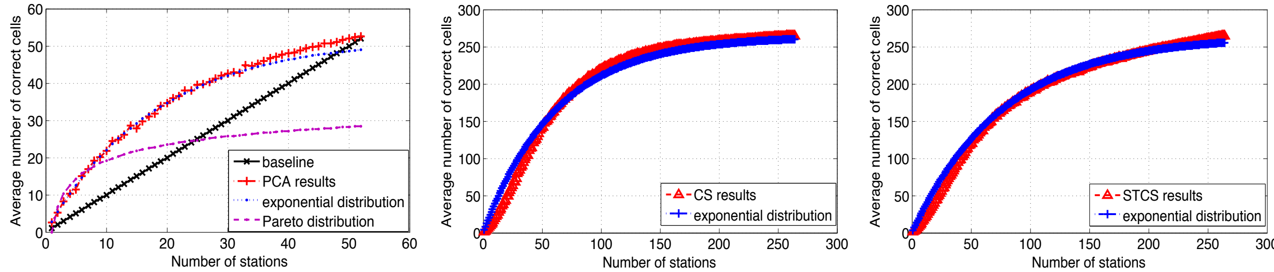
B. The model of mobile participants

Definition 5: (selection matrix) Suppose there are total U participants who are selected to execute the sensing task. Let $S^{(u)}$ denote the selection matrix of participant u ($1{\le}u{\le}U$), $S_{ij}^{(u)}{=}1$ if the cell i is sensed by participant u at j^{th} slot. Thus, we have selection matrix $S{=}\sum_u S^{(u)}$. We set $S_{ij}{=}1$ if $S_{ij}{\ge}1$. It means repeated sensing of one cell in the same slot by multiple participants is regarded as a waste.

The participants need to report their basic information to the platform as they appear. For any participant u, the start time, the start location, the destination, and the deadline are denoted as $t_s^{(u)}$, $l_s^{(u)}$, $l_d^{(u)}$, and $t_d^{(u)}$, respectively. $1{\le}t_s^{(u)}{\le}M$, $1{\le}t_d^{(u)}{\le}M$. Once a participant registers for the task, we assume that they will execute the sensing task until reaching their destinations. The CMCS platform makes trajectory plans for selected participants to make sure that each of them can reach the destination before the deadline. Hence, we have $S_{l_s t_s}^{(u)}{=}1$ and $S_{l_d t}^{(u)}{=}1$, where $1{\le}t{\le}t_d^{(u)}$. However, the participants come to the platform dynamically, the total number and distribution of participants are unknown at first. It is challenging to determine which participant to select aiming to maximize the profit.

C. The POPS problem formulation

In this paper, our objective is to maximize the profit of the platform for a sensing task through data reconstruction, with the above models of tasks and participants. To achieve the objective, we propose the Profit-driven Online Participant Selection (POPS) problem in CMCS: Given the task area N and duration M, the gain g of a correct cell, the payment c of sensing a cell, the quality requirement α, the budget B of the platform, and participant mobility information $t_s^{(u)}$, $l_s^{(u)}$, $l_d^{(u)}$, $t_d^{(u)}$, we determine the selection matrix S to maximize



(a) PCA (b) CS (c) STCS

Fig. 1: Statistic results of the reconstruction quality in the temperature dataset

the profit of the CMCS platform. Hence, our POPS problem can be formulated as:

$$\begin{aligned}
&\max \ \mathrm{Q} - \tfrac{\mathrm{c}}{\mathrm{M\times N\times g}} \times \sum_{\mathrm{i,j}} \mathrm{S_{ij}} \\
&s.t. \\
&Q \geq \alpha; \\
&c \times \sum_{i,j} S_{ij} \leq B; \\
&\forall u, S_{l_s t_s}^{(u)}{=}1; \\
&\forall u, S_{l_d t}^{(u)}{=}1, 1{\leq}t{\leq}t_d^{(u)};
\end{aligned} \tag{2}$$

where $1{\leq}i{\leq}N, 1{\leq}j{\leq}M, 1{\leq}u{\leq}U$. The first two constraints indicate the task requirement and budget limitation, respectively. The reconstruction quality must beyond the threshold and the total payments should not exceed the budget. The last two constraints indicates the mobility constraints of participants. Each participant should reach the destination before the deadline under the planned trajectory.

III. Reconstruction Quality Prediction

Recently, researchers have proposed many reconstruction methods for various applications. In this paper, we consider three representative methods, including PCA (Principal Component Analysis), CS (Compressive Sensing), and STCS (Spatial and Temporal Compressive Sensing). PCA is a popular reconstruction method for most applications. CS and STCS have been used in the field of environment monitoring [7] [8].

We first implement the PCA reconstruction method on the real-life temperature dataset [13] in one slot and compute the reconstruction quality under the different numbers of sensing cells. For each setting of the number of sensing cells, we select such number of cells randomly and conduct 100 rounds of experiments and calculate the reconstruction quality on average. To estimate the relationship between the reconstruction quality and the number of sensing cells, we evaluate two representative cumulative distribution functions: the exponential distribution and the Pareto distribution. We adopt the maximum likelihood estimation (MLE) method to estimate the parameters of the two distributions and perform the Akaike test to decide the best one. The evaluation results are illustrated in Fig. 1(a). We can see that the relationship between the reconstruction quality and the number of sensing cells closely resembles the exponential distribution.

We further evaluate the reconstruction quality by the CS and STCS reconstruction methods in all slots. Fig. 1 plots the reconstruction results, under reconstruction methods and the number of sensing cells. We find that all of such kinds of evaluation results are all consistent with the exponential distributions. The root cause is that the number of deduced cells increases at first with more sensed cells. However, the number overlapped deduced cells increases, too, as the number of sensed cells large enough. That makes the reconstruction quality initially has a high rate of increase, but as the sensed cells more and more, the increase rate taper off, until all the cells are correct. Hence, it is reasonable to infer that the relationship between the reconstruction quality and the number of sensing cells follows the exponential distributions with different parameter settings. Accordingly, we can utilize the exponential distribution to estimate the reconstruction quality, when the number of sensing cells is given.

The cumulative distribution function of the exponential distribution can be expressed as follows:

$$F(y;\lambda) = \begin{cases} 1 - e^{-\lambda y}, & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{3}$$

According to our observations, the reconstruction quality can be estimated by the following equation:

$$Q' = max\{\frac{y}{N \times M}, F(y;\lambda)\} \tag{4}$$

where y is the number of sensing cells, $y{=}\sum_{i,j} S_{ij}$. In this paper, we can infer the true reconstruction quality Q by assuming that $Q{\approx}Q'$.

To solve the POPS problem, we need to derive the ideal value of the parameter λ for X, then we can predict the reconstruction quality based on selected participants. However, the value of λ for the x (the spatial data matrix in a slot) and X (the spatial-temporal data matrix across all the slots) are different even using the same reconstruction method. The reason is that the data reconstruction is based on the matrix structure. Consider that those mobile participants come to the platform dynamically. It is infeasible to predict the sensing matrix C due to the lack of prior knowledge; hence, we can not get the exact value of λ for the X directly. Fortunately, we find that the λ for x in different slots ($x^{(1)}$, $x^{(2)}$, $x^{(3)}$, ...) is almost the same, we denote it as λ_c. For this reason, we propose to maximize the sub-profit in each slot and utilize the sub-profit sum across all slots to approximate the maximum profit of the CMCS platform. Thus, according to the exponential-based quality prediction measurement and the gradual approximation method, the objective function in each slot is $1-e^{-\lambda_c(\sum\limits_{i,j} s_{ij})}-\frac{c}{n\times m\times g}\sum\limits_{i,j} s_{ij}$, the quality threshold constraint and the budget constraint can be combined as $-\frac{\ln(1-\alpha/M)}{\lambda_c}\leq\sum\limits_{i,j} s_{ij}\leq\frac{B}{M\times c}$.

IV. Participant Selection mechanisms

A. ProSC: an intrinsic mechanism

The objective function of the POPS problem in each slot experiences two variables, λ_c and $\sum_{i,j} s_{ij}$. The variable $\sum_{i,j} s_{ij}$ includes the participant quantity and distribution. In ProSC, we suppose that the platform would guide each selected participant to move along the shortest route from the start location to the destination. Based on this assumption, the distribution of selected participants can be known beforehand. Therefore, only the λ_c and participant quantity y ($y=\sum_{i,j} s_{ij}$) should be determined. The main idea of our mechanism is that, we iteratively adjust the values of λ_c and y, until the profit in each slot converges to the maximum value.

There are three steps in each slot. In the first step, we optimize y under a given value of λ_c. If λ_c is determined, the y is easy to get since the objective function in each slot is convex, and a traditional linear programming algorithm can be used. Note that, the initial value of λ_c in the first slot is determined randomly. In the second step, we would select a set of mobile participants to carry on the sensing task. Some participants may leave the target sensing area after certain slots, while new participants may enter the sensing area at any time. If there are y' participants at the $(t-1)^{th}$ slot, $y-y'$ new participants would be selected in the t^{th} slot. In ProSC, we select the $y-y'$ new participants randomly, the y' existing participants would move along their shortest routes and continue to execute the sensing task in the next slot.

In the third step, we optimize the value of λ_c through repetitive cross-validation based on collected data in each slot. Cross-validation is a common method to estimate parameters [7]. We take an example to illustrate the main idea. Consider that we have recruited k participants to sense k cells in a slot and the sensing data have been collected. Thus, we select $k-1$ cells from the k sensing cells, execute the data reconstruction and deduce the data of the left one cell. The ground truth of k sensing cells are all known, we can compare the reconstruction results of the k cells with the ground truth of them, and get the error rate θ of the k sensed cells. As discussed in [7], the observed error rate satisfy the normal distribution around the actual error rate, $\theta_i \sim N(\vartheta,\sigma^2)$, where ϑ denotes the ground truth of the actual error rate. Therefore, we repeat the above cross-validation process p rounds, and use the average value of observed error rates to estimate the actual error rate, that is $\vartheta\approx\bar{\theta}$, and $\bar{\theta}=(\sum\limits_{i=1}^{k\times p}\theta_i)/(k\times p)$. Thus, the value of the parameter λ_c can be estimated based on the error rate, $\lambda_c=\ln(1-\vartheta)/(-k)$. The value of λ_c is used in the next slot to optimize the number of required participants.

Repeat the above steps, the parameter λ_c and participant quantity y will be updated in each slot, while the profit will converge to the maximum value in a few slots. The selection matrix S compiled by $S^{(u)}$ of all slots would be an approximate solution to the POPS problem. However, the participant distribution has not been considered. In ProSC, the routes of some participants may be similar with each other or even overlapped in some slots. This phenomenon will incur lower reconstruction quality and higher error rate under the same number of participants. The uneven error rate in different slots may lead to the un-converging of our mechanism.

B. ProSC+: a distribution-aware mechanism

To tackle the un-converging problem in the ProSC, we optimize the participant distributions in ProSC+. Intuitively, it would be better if those selected participants disperse in the target sensing area, so as to sense more different cells. For this reason, those similar routes should be avoided in each slot. Ideally, the most representative cells should be sensed. Wang et al. [7] propose to select one more representative sensing cell in each slot, until the data quality is satisfied. This method is not suitable for mobile and dynamic participants. On the other hand, uniform coverage is discussed currently and the entropy was proposed as a metric to illustrate the dispersion degree of participant distribution [6] [14].

To maintain a stable dispersed distribution in all slots, we utilize the 2D entropy to guide the participant distribution in ProSC+. The 2D entropy has been widely used in the field of image segmentation [15], since it could illustrates the information confusion and spatial correlation of pixels in an image. The more confusion the image, the higher the entropy value. Similarly, we can formalize the sensing results in a slot as an image, and the number of participants in each cell acts as the gray value of a pixel. The more decentralized the participants, the higher the entropy value. The 2D entropy (E) can be calculated as:

$$E=-\sum_a\sum_b\frac{f(a,b)}{n\times m}\times\log_2(\frac{f(a,b)}{n\times m}) \tag{5}$$

where a denotes the number of participants in a cell, b denotes the average number of participants in the neighborhoods of a cell, and $f(a,b)$ is the repetition number of 2-tuples (a,b).

We divide the participants in a slot into two kinds, existed participants and new participants. For the existed participants, the trajectory planning is restricted by the mobility constraints. 1) the movement distance between two slots should not exceed the mobile capability of each participant. 2) each participant must reach the destination before the deadline under allocated sensing cells. Suppose there are y' existed participants, and the speed of each participant can be calculated through historical data. To select a proper set of sensing cells in each slot, we first assign a selection pool containing potential cells for each

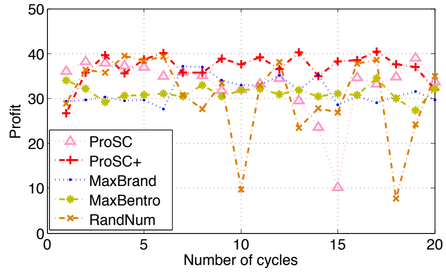


Fig. 2: Profit compare.

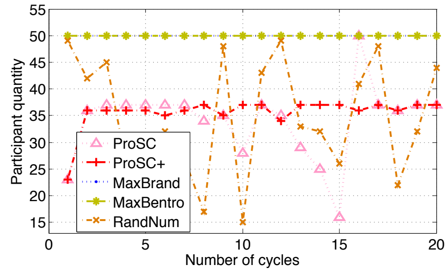


Fig. 3: Participant quantity compare.

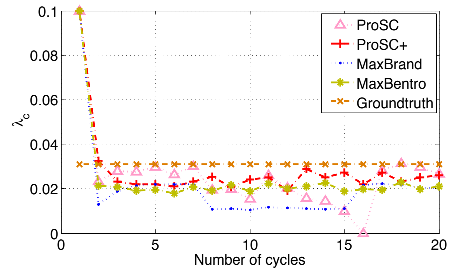


Fig. 4: The λ_c prediction.

participant u ($1 \leq u \leq y'$), then choose a optimal cell as the sensing cell.

$$pool_t^{(u)} = \bigcap_l \begin{aligned} &(\bigcup_l (\bar{v}^{(u)} \geq d(l' \,\Big|\, S^{(u)}_{l'(t-1)}=1 \,, l \,\Big|\, S^{(u)}_{lt}=1) \,, \\ &\bigcup_l (\bar{v}^{(u)} \times (t_d^{(u)} - t) \geq d(l \,\Big|\, S^{(u)}_{lt}=1 \,, l_d^{(u)}))) \end{aligned} \tag{6}$$

where $\bar{v}^{(u)}$ denotes the average speed of participant u. The slot length is formulated as 1. $d(i,j)$ denotes the Manhattan distance between location i and j. For the $y-y'$ new participants, we should select proper set from the start locations of all candidates. It means all the start locations of candidates, whose start time is t, can be regarded as the selection pool of new participants.

$$pool_t^{+} = \bigcup_{w(w \neq u)} l_s^{(w)} \,\Big|\, t_s^{(w)} = t \tag{7}$$

We combine the selection process of the two kinds of participants, and formulate the selection problem in each slot as: pick y' proper cells from the selection pool of each existed participant and $y-y'$ cells from the start locations of candidate participants, aiming to maximize the E of all the y cells. This selection problem is a knapsack problem, and the greedy algorithm is an effective resolution to this kind of problem. The value of E, however, is a global parameter; hence, a progressive plain greedy algorithm cannot be directly used. To tackle the selection problem, we adopted an iteration based greedy algorithm. We first select y cells randomly and compute E, then repeat the selection process for I_e rounds, where I_e is the threshold. At last, the participant distribution with the largest E within the I_e rounds is selected as an approximately optimal solution. The calculation of the participant quantity y and the λ_c in ProSC+ is the same with those in ProSC.

V. Performance Evaluation

A. Experiment setup

Dataset: We evaluate our mechanisms based on a real-life temperature dataset. The temperature dataset was delivered by the SensorScope project [13]. The project deploys 88 sensors across the EPFL campus to collect environmental parameters from 2006.11.01 to 2007.05.09. The sampling period of each sensor is 30 seconds. The target sensing area is about $500m \times 300m$. We divide the sensing area as 10×10 cells, and find that 52 cells are covered by sensors. If a cell is covered by multiple sensors, the average value of those sensors acts as the sensing result in that cell.

To enable our experiments, we first reconstruct the whole data matrix based on the real dataset, and use the synthetic data as the ground truth to carry on the following test. The reason is that the number of cells covered by stations is limited in the real dataset. We can not simulate the participant trajectories with these limited and separated sensing cells. Note that, the reconstruction method to synthesize data and experiments are different in our paper. The synthesizing process may lead to a higher value of the parameter λ_c in our experiments than in a real situation, yet the theories discussed in this paper is not impacted. Since the dataset do not contain the participants information, we randomly generate the mobility settings for each participant, including the start time, the start location, the end location and the deadline.

Baselines: To compare with our mechanisms, we use the following three baselines.

- **MaxBrand**: All budget is used to recruit participants. The new participants in each slot are selected randomly. The trajectories of selected participants are the shortest routes.
- **MaxBentro**: All budget is used to recruit participants. When new participants are needed, the ones with maximum entropy value are selected. The trajectories of selected participants are guided through the platform to cells with a highest entropy value.
- **RandNum**: The number of participants in each slot is determined randomly, the participant trajectories are not considered in this baseline.

B. The profit and participant quantity compare

Fig. 2 illustrates the profit results of our algorithms and baselines. The maximum achievable profit computed offline is around 40. We can find that the profits of all five mechanisms vibrate in different slots. The reason is that the sensing data in different slots is variable. The profit of ProSC around the 15^{th} slot performs large offset, because the participant trajectories are similar with each other in those slots. The profit of ProSC+ is more stable than ProSC, and it is higher than other mechanisms in most slots. Since the number of participants in the RandNum mechanism is selected randomly within the budget, the peak value of profit achieved by RandNum in different slots can be seen as the possible maximum profit.

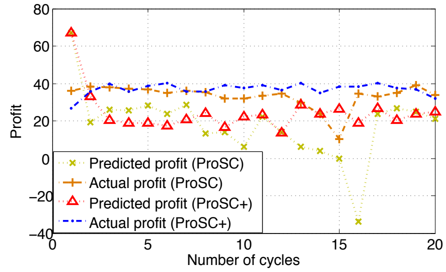


Fig. 5: Predicted and actual profit.

The profit of ProSC+ after the second slot is always near the maximum profit, which proves the effectiveness of our mechanism.

Fig. 3 describes participant quantity compare of different mechanisms in each slot. The optimal ground truth of participant quantity under our settings is 36. As all the budget is used to recruit participants in MaxBrand and MaxBentro, the participant quantity of these two mechanisms are all equal to the budget. The results of ProSC+ and ProSC are similar in most slots, except that the latter one vacillates hardly in some slots. We can see that, the participant quantity of ProSC+ is stable around the ground truth 36 after the second slot. It means that the convergence rate of ProSC+ is within two slots, which prove the effectiveness of our mechanism.

C. The prediction of λ_c and profit

Fig. 4 shows the predicted value of the parameter λ_c in each slot. Note that, the ground truth of the λ_c in each slot under our settings is 0.031. We can see that the predicted λ_c in ProSC and MaxBrand perform great fluctuation through different slots, due to the lackness of the distribution controlling process. The predicted λ_c in ProSC+ and MaxBentro is more stable, and the former one is always closer to the ground truth than the latter one. Especially, the prediction error of ProSC+ is always less than 0.01 after the second slot. According to the quantity prediction method introduced in Section IV, the quantity prediction error is within ± 1 under the illustrated λ_c prediction error. Therefore, the prediction method introduced in our paper is proven effective to determine the participant quantity in each slot.

Fig. 5 represents the compare results of the predicted profit and actual profit. We can find that both the predicted profit and actual profit of ProSC fluctuate obviously in different slots, while those of ProSC+ are more stable. However, the predicted profit of both ProSC and ProSC+ are always lower than their actual profits. The reason is that the parameter λ_c is underestimated (Fig. 4), as we predict λ_c through repetitive cross-validation based on the limited sensing data. To predict the profit more accurately, a better method is needed to predict the parameter λ_c.

VI. Conclusions

In this paper, we focus on maximizing the profit of the CMCS platform, which employs dynamic and mobile participants to sense a part of sensing cells and deduce the results of all unsensed cells via the data reconstruction. We characterize such an optimization problem, denoted as POPS, with the quality requirement of the sensing task and the mobility constraints of participants. We then propose two methods to solve the POPS problem. We conduct evaluations on a real-life dataset and the results demonstrate the effectiveness and efficiency of our mechanisms.

Acknowledgment

The work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61422214, 61772544, 61379144 and 61402513.

References

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile Crowdsensing : Current State and Future Challenges," *IEEE Communications Magazine*, no. November, pp. 32–39, 2011.

[2] J. Cherian, J. Luo, H. Guo, S.-S. Ho, and R. Wisbrun, "Parkgauge: Gauging the occupancy of parking garages with crowdsensed parking characteristics," *Proceedings of the 17th IEEE International Conference onMobile Data Management (MDM)*, vol. 1, pp. 92–101, 2016.

[3] S. Morishita, S. Maenaka, D. Nagata, M. Tamai, K. Yasumoto, T. Fukukura, and K. Sato, "SakuraSensor: quasi-realtime cherry-lined roads detection through participatory video sensing by cars," *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*, pp. 695–705, 2015.

[4] T. Ludwig, C. Reuter, and V. Pipek, "What you see is what i need: Mobile reporting practices in emergencies," *Proceedings of the 13th European Conference on Computer Supported Cooperative Work (ECSCW)*, pp. 181–206, 2013.

[5] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," *Proceedings of the 34th IEEE International Conference on Computer Communications (INFOCOM)*, pp. 2542–2550, 2015.

[6] S. Ji, Y. Zheng, and T. Li, "Urban sensing based on human mobility," *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*, pp. 1040–1051, 2016.

[7] L. Wang, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, and Y. Wang, "CCS-TA: Quality-Guaranteed Online Task Allocation in Compressive Crowdsensing," *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pp. 683–694, 2015.

[8] L. Kong, M. Xia, X.-Y. Liu, G. Chen, Y. Gu, M.-Y. Wu, and X. Liu, "Data Loss and Reconstruction in Wireless Sensor Networks," *Proceedings of the IEEE 32th International Conference on Computer Communications (INFOCOM)*, pp. 1–1, 2013.

[9] H. Li, T. Li, F. Li, W. Wang, and Y. Wang, "Enhancing participant selection through caching in mobile crowd sensing," *Proceedings of the IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)*, pp. 1–10, 2016.

[10] S. Yang, F. Wu, S. Tang, T. Luo, X. Gao, L. Kong, and G. Chen, "Selecting most informative contributors with unknown costs for budgeted crowdsensing," *Quality of Service (IWQoS), 2016 IEEE/ACM 24th International Symposium on*, pp. 1–6, 2016.

[11] X. Yin, Y. Chen, and B. Li, "Task assignment with guaranteed quality for crowdsourcing platforms," *Quality of Service (IWQoS), 2017 IEEE/ACM 25th International Symposium on*, pp. 1–10, 2017.

[12] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-Temporal Compressive Sensing and Internet Traffic Matrices," *Network*, vol. 20, no. 3, pp. 267–278, 2009.

[13] F. Ingelrest, G. Barrenetxea, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, "Sensorscope: Application-specific sensor network for environmental monitoring," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 2, p. 17, 2010.

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

[15] J. Fan, R. Wang, L. Zhang, D. Xing, and F. Gan, "Image sequence segmentation based on 2d temporal entropic thresholding," *Pattern Recognition Letters*, vol. 17, no. 10, pp. 1101–1107, 1996.