Fine-Grained Multitask Allocation for Participatory Sensing With a Shared Budget

Jiangtao Wang, Yasha Wang, Daqing Zhang, Leye Wang, Haoyi Xiong, Abdelsalam Helal, *Fellow, IEEE*, Yuanduo He, and Feng Wang

***Abstract*—For participatory sensing, task allocation is a crucial research problem that embodies a tradeoff between sensing quality and cost. An organizer usually publishes and manages multiple tasks utilizing one shared budget. Allocating multiple tasks to participants, with the objective of maximizing the overall data quality under the shared budget constraint, is an emerging and important research problem. We propose a fine-grained multitask allocation framework (MTPS), which assigns a subset of tasks to each participant in each cycle. Specifically, considering the user burden of switching among varying sensing tasks, MTPS operates on an attention-compensated incentive model where, in addition to the incentive paid for each specific sensing task, an extra compensation is paid to each participant if s/he is assigned with more than one task type. Additionally, based on the prediction of the participants' mobility pattern, MTPS adopts an iterative greedy process to achieve a near-optimal allocation solution. Extensive evaluation based on real-world mobility data shows that our approach outperforms the baseline methods, and theoretical analysis proves that it has a good approximation bound.**

***Index Terms*—Fine-grained, multitask allocation, participatory sensing (PS).**

I. Introduction

CROWDSOURCING, a term coined in [1], is defined by as the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people. With an increasing demand for real-time environmental information in smart cities (e.g., air quality, noise level, and traffic condition) [2] and the proliferation of mobile devices, a special case of crowdsourcing, called the *participatory sensing* (PS) [3] (or mobile crowd-sensing [4]) has become an effective way to sense and collect information at a low deployment cost. In PS systems, *organizers* publish sensing tasks onto a mediation platform, while *participants* with mobile devices sense and report the information of their surroundings. Data quality and cost are two opposing factors for PS. *Managing the tradeoff between data quality and cost through task allocation is a fundamental and crucial research problem* in PS.

Several research studies were proposed recently which aimed at either maximizing the data quality with budget constraints (see [5]–[7]), or minimizing the sensing cost while guaranteeing a minimum level of data quality (see [8]–[10]). Most of them assume that each task runs independently, so that they focused on the optimal allocation of a single task without considering the co-existence of multiple concurrent tasks.

With the increasing demand for PS, it is a common phenomenon that an organizer publishes and manages multiple concurrent PS tasks with one shared budget [11]. For example, instead of building a specialized system for each single PS task, the city government may choose to handle multiple tasks, such as air quality monitoring, noise level reporting, and traffic condition detection, in a multitask PS system with a shared pool of participants and a total shared budget. *How to allocate tasks to participants, with the objective of maximizing the overall data quality (ODQ) of multiple tasks under the shared budget constraint*, becomes an important and emerging research problem.

A straightforward solution to tackle the above problem is to decompose the multitask allocation into a set of single-task allocation problems. First, the total shared budget is divided into several parts based on single-task importance. Then, by directly adopting the existing techniques (see [5]–[7]), the allocation of each single task is optimized under its own budget constraint independently. However, such solution cannot guarantee that the ODQ is optimal [11], because the predefined budget for each task may be inadequate or in excess to achieve a minimum level of data quality requirement.

Instead of adopting the above straightforward solution, recent studies [11], [12] proposed a novel strategy. Specifically, they attempt to effectively select a subset of participants from a huge user pool. Each selected participant is given an equal and fixed reward, and they have to complete all sensing tasks as long as their mobile devices have the required sensors, while the unselected participants are excluded from the entire sensing process. However, such approach has a disadvantage under an increasing number of tasks, in which *the selected participants may undertake too many tasks*, which could lead to excessive energy consumption, costly bandwidth, attention overdraw, among other concerns [11]. Moreover, the number of assigned tasks for each participant may differ based

Manuscript received June 7, 2016; revised August 22, 2016; accepted September 6, 2016. Date of publication September 9, 2016; date of current version January 10, 2017. *(Corresponding author: Yasha Wang.)*

J. Wang, Y. Wang, D. Zhang, Y. He, and F. Wang are with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wangyasha@pku.edu.cn; daqing.zhang@telecom-sudparis.eu).

L. Wang is with Telecom SudParis, 91011 Évry, France.

H. Xiong is with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA.

A. Helal is with the Computer Science Department, University of Florida, Gainesville, FL 32611 USA.

Digital Object Identifier 10.1109/JIOT.2016.2608141

on the diverse sensing capabilities of their mobile devices, thus, giving an equal reward for each selected participant is unfair.

Therefore, in this paper, *instead of requiring selected participants to complete all sensing tasks and excluding unselected ones, we consider a more fine-grained multitask allocation approach for PS*. We include all candidate participants in the process of PS by *assigning a subset of tasks to each participant in each cycle, with the objective of maximizing the ODQ of multiple PS tasks under a total shared budget constraint*. Specifically, our approach is based on the following assumptions.

1) *Incentive Cost:* In this paper, incentives are given to the participants based on two principles as follows.
   a) *"More Pay for More Work" Principle:* The more samples a participant reports, the more rewards s/he will get.
   b) *Task Switching Compensation Principle:* When switching to a new task, the participants have to make extra efforts in reading the task description, observing the phenomenon, collecting and sharing data, and so on.

   Thus, it is reasonable to give extra reward to compensate such extra effort. Based on the above two principles, we adopt a fine-grained attention-compensated incentive model, which consists of the following two components.
   a) *Task-Specific Base Incentive:* A task-specific amount of incentive is given to the participants for each sample s/he contributes. Because completing different tasks requires different levels of human attention and device-oriented resources, it varies from one task to another.
   b) *First-Time Participation Bonus Incentive:* If a participant is assigned with a task that has never been completed by him/her before, s/he will get an extra bonus, which is proportional to the task-specific base incentive. The detail of the incentive model is described in Section III.
2) *Overall Data Quality:* The ODQ of multiple PS tasks is defined as the weighted sum of the data quality of each single task. The weights characterize the importance of each task, which, we assume, are set by the organizer. For each single task, the data quality is calculated as the average quality in each temporal-spatial cell.[1] In this paper, similar to [13], we use the number of samples to characterize the achieved data quality in the temporal-spatial cell. The details of data quality metric definition are described in Section III.

To further demonstrate the research issues and challenges, a motivating case is illustrated as follows. The city government of Beijing has developed a platform to provide real-time and location-dependent information about urban environment based on the PS paradigm. There are three concurrent tasks (see Table I) running on the platform published by the government with a total budget constraint of 1000 RMB. All tasks share the same subareas and cycles, and each task has its task-specific data quality requirement. To more simply describe the problem, we focus on a specific downtown area during a specific time period (e.g., one week). This downtown area is divided into 20 virtual subareas, and the region covered by a cell tower is regarded as one subarea. The entire duration is divided into several equal-length cycles (e.g., one hour per cycle). There are 500 mobile users registered as candidate participants, and their historical records of connection to the cell towers are utilized by the platform only for the purpose of task allocation after anonymity process. An appropriate subset of tasks in each cycle is assigned to each participant with the total budget constraint, and is predownloaded into each participant's smartphone. To reduce the disturbance, the preassigned tasks will only be pushed to the participants online when they connect to cell towers (placing a call or sending a text message). In this manner, the participants would notice the message of new tasks after finishing the call/text sending. The system adopts the aforementioned attention-compensated incentive model (see Table I).

TABLE I
THREE TASKS ON THE PLATFORM

		Incentives	
Task	**Participant Involvement**	**base**	**bonus**
Noise Level Report	Select noise level (quiet/medium/noisy)	0.1	0.1
Traffic Detection	Take pictures	0.2	0.2
Air Quality Monitoring	Take phone out and collect	0.25	0.25

From the above use case, the platform needs to first combine the participants, sensing cycles and tasks into participant-cycle-task triples, where each triple refers to the task allocation of a specific sensing task to the specific participant in a specific sensing cycle. Then the platform needs to select a set of triples from all combined ones to maximize the ODQ under the total budget constraint, with the attention-compensated incentive model in mind. To solve this research problem there are at least the following challenges that must be overcome.

A. Optimizing Multitask Allocation While Lowering the Computation Complexity

First, as the task assignment is determined offline, we need to develop an algorithm to predict the participant's connection to the cell tower and located subarea in each cycle based on historical data from the telecom operator. Second, even if we can predict the mobility and connections accurately, it is still very difficult to perform multitask allocation because of the larger search space compared to the state-of-the-art studies. Take the abovementioned use case as an example: with 500 candidate participants, ten tasks, and ten cycles, there are 50 000 task-participant-cycle triples in total for our fine-grained multitask allocation problem, so that at most

[1]Generally speaking, for PS task, the organizer usually specifies the target sensing area consisting of a set of subareas, and sensing duration divided into equal-length sensing cycles. Thus the PS task consists of a set of temporal-spatial cells.

2^{50000} task assignment plans need to be evaluated. In contrast, [11] and [12] only select a subset of participants, so that the size of its search space is 2^{500}, which is much smaller than ours. Thus, it is impossible to find the optimal one through a brute-force approach. Instead, a mechanism to obtain a near-optimal multitask assignment while lowering the computation complexity is required.

B. Estimating the Incentive Cost of Each Task-Participant-Cycle Triple During the Search for Near-Optimal Solution

No matter what approach is adopted, during the search process, the utility of each task-participant-cycle triple needs to be evaluated, to which the incentive of each triple is relevant. However, by adopting the attention-compensated incentive model, the incentive cost of triple (t, p, c) depends on whether the participant p has completed the task t before cycle c. As we cannot foresee this in the search process, it is difficult to compute the incentive cost of each triple. Therefore, we need to design a task allocation process that can iteratively approximate the incentive of each triple.

With the abovementioned research objective and challenges, the main contributions of this paper are as follows.

1) We study a fine-grained multitask allocation problem for PS. Instead of requiring selected participants to complete all sensing tasks and excluding unselected ones, all candidate participants are involved in the allocation process. The objective is to assign a subset of tasks to each participant per cycle, with sensing quality optimization goal and a total shared budget constraint.
2) We propose a novel framework, named MTPS, which adopts a two-stage offline multitask allocation process. MTPS first predicts participant's mobility and connection to the cell tower in each cycle based on historical data from the telecom operator. It then adopts an iterative greedy process to optimize the allocation. Theoretical analysis shows that the algorithm can achieve near-optimality with low computational complexity.
3) We evaluate our approach extensively with mobility traces from the real world. We verify that the proposed approach outperforms baseline methods achieving higher ODQ under various settings.

The rest of this paper is organized as follows. Section II reviews the related research activities. Section III establishes the system models of our framework. Section IV describes the proposed two-step framework for multitask allocation. Section V extensively evaluates the performance of the proposed strategy by real-trace-driven simulations, and finally, Section VI concludes this paper with discussions of future work directions.

II. Related Work

A. Single-Task Allocation Optimization

A group of research studies aimed at maximizing the data quality of PS with certain constraints. Reddy *et al.* [5], [14] first studied the research challenge of participant recruitment in PS, and they proposed a coverage-based recruitment strategy to select a predefined number of participants so as to maximize the spatial coverage. Singla and Krause [6] proposed a novel adaptive participant selection mechanism for maximizing spatial coverage under total incentive constraint in community sensing with respect to privacy. Cardone *et al.* [7] developed a mobile crowdsensing platform, where a simple participant selection mechanism was proposed to maximize the spatial coverage of crowdsensing with a predefined number of participants. Xiong *et al.* [13] considers a new version of task-assignment problem, where the optimization goal is to maximize the number of calls (sensing data) for certain location sets under an overall budget constraint.

Another group of research studies, on the other hand, aimed at minimizing the cost while ensuring a certain degree of the data quality. Zhang *et al.* [2] studied offline participant selection in piggyback mobile crowdsensing for probabilistic coverage. Their goal was to select minimum number of participants to guarantee that the selected participants made enough number of calls at a certain percentage of the target locations over a long fixed sensing period. Similarly, Xiong *et al.* [8], [15] investigated how to assure the full coverage over several cell towers with the minimum number of users. Karaliopoulos *et al.* [9] studied the user recruitment for mobile crowdsensing over opportunistic networks, whose goal was to minimize the total cost even as ensuring the full coverage of point-of-interest. Hachem *et al.* [10] proposed a participant selection framework for PS, which reduced the number of selected participants by predicting mobile users' future locations in the next time slot (or sensing cycle) based on their current location and recent trajectory. Philipp *et al.* [16] and Wang *et al.* [17] leveraged the spatial or temporal correlations to reduce the number of participants required in each sensing cycle, while meeting certain coverage quality constraints. Recently, Song *et al.* [18] proposed a QoI-aware and energy-efficient participant selection strategy, which aimed at balancing the quality of Information and energy cost.

The research objective of above studies is to optimize the allocation of a single PS task, which have not addressed the competition among multiple concurrent tasks on the shared resource, i.e., the total budget.

B. Multitask Allocation Optimization

Although early research focused on the allocation of a single PS task, recent studies explore the multitask allocation problem.

Yang *et al.* [19] proposed an approach to select a subset of participants to minimize total latency of sensing tasks while ensuring the quality of data. In [11] and [12], the goal was to select a subset of participants to best provide the QoI satisfaction metrics for all tasks under the total budget constraints. It first assumes that each participant is paid a fixed amount of incentives, and then the selected participants need to join in all sensing cycles and complete all tasks. Different from [11] and [12], the approach proposed by [20] and [21] selected a subset of participants in each cycle to perform all tasks. In [20], the goal was to minimize total cost even as ensuring different levels of coverage of multiple tasks, whereas [21] aimed at minimizing the overall sensing cost of mobile devices with heterogeneous sensing capabilities even as completing all location-specific tasks. References [22] and [23] study the problem of multitask participant selection under two situations: 1) participant selection based on participants' intentional movement for time-sensitive tasks and 2) unintentional movement for delay-tolerant tasks.

Whether the selection was for the whole cycles or each cycle, the objective of the above study was formulated in such a manner as to effectively select an appropriate subset of participants from a huge user pool to perform all sensing tasks, while unselected ones do nothing. In contrast, the goal of this paper was to assign a subset of tasks to each participant in each cycle, which is more fine-grained.

III. System Model

In this section, we first define the incentive model and data quality metrics, and then clarify the mathematical problem formulation.

A. Incentive Model

As the extra effort in completing new tasks is considered, we adopt a novel incentive model consisting of the following two components.

1) *Task-Specific Base Incentive:* For each collected sample of task i, a task-specific amount of incentive, denoted as α_i, is paid to the participant who contributes it.
2) *First-Time Participation Bonus Incentive:* If a participant reports one sample for a task i that has never been completed by him/her before, he/she will get an extra incentive denoted as β_i, which is proportional to the task-specific base incentive, that is, $\beta_i = \tau * \alpha_i$, where τ is a positive coefficient defined by the platform or organizer.

Thus the incentive cost of a task-participant-cycle triple (t, p, c) is defined as follows:

$$\text{cost}(t,p,c) = \begin{cases} \alpha_t * K, & \text{if } p \text{ has completed } t \text{ previously} \\ \alpha_t * K + \tau * \alpha_t = (\tau + K)\alpha_t, & \text{else} \end{cases} \quad (1)$$

where t, p, and c denotes task, a participant, and a cycle, respectively, and $K(K \neq 0)$ is the number of samples p reports for task t in cycle c. For example, for the participant shown in Fig. 1, we set $\tau = 1$, and the base incentive of task A, B, and C is 0.1, 0.2, and 0.25 RMB, respectively. Besides, we assume that the participant connects to the cell tower at least one time when moving into a new subarea (the grid in Fig. 1), and completes each assigned task. Then, with the mobility traces and task assignment in Fig. 1, this participant can earn $(0.2 \times 2 + 0.2) + (0.25 \times 2 + 0.25) = 1.35$ RMB in cycle 1, $(0.1 \times 2 + 0.1) + 0.2 \times 2 = 0.7$ RMB in cycle 2, and $(0.1 + 0.2 + 0.25) \times 2 = 1.1$ RMB in cycle 3.

B. Data Quality Metric

1) Assumptions: For each PS task, the organizer needs to specify the target sensing area, which often consists of a set

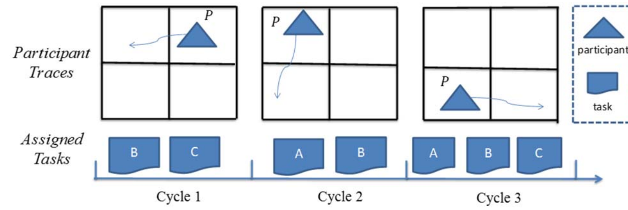


Fig. 1. Task allocation and incentive cost: an example.

of subareas. The organizer also needs to specify the sensing durations (e.g., five days), which is usually divided into equal-length cycles (e.g., each cycle lasts an hour). We call a specific subarea in a certain cycle as a *spatial-temporal cell*.

Assumption 1: The data quality of the PS task is associated with the number of samples in each spatial-temporal cell, but will saturate when the number of samples reaches a certain threshold. As humans are involved in the sensing loop, the data quality of a single sample may be inaccurate; thus for each spatial-temporal cell, a number of samples may be needed. First, because of human subjectivity, different participants may contribute different data for the same phenomenon. For example, a task requires participants in the supermarkets to manually select the crowd level (very crowded/crowded/middle/less crowded) and report to the server. In the same place and at the same time, participants may select different levels as it is subjective. Second, participants may deliberately contribute inaccurate data for their own benefits; for example, a leasing agent may intentionally contribute fabricated low noise readings to promote the properties in a particular suburb. Therefore, to ensure data quality, we need to collect more than one sample for each spatial-temporal cell. However, if we increase the number of samples above a certain degree, the quality of the data might not increase anymore. In fact, it is difficult to determine this degree, so that according to [13], we assume that it is identical among all spatial-temporal cells of a specific task, which is set by the organizer.

Assumption 2 (Different Types of Tasks Have Different Requirements in Terms of the Amount of Samples): For example, to represent the status of a subarea, the amount of samples the noise level report task needs is larger than that of the air quality monitoring task. This is because the noise level may be quite diverse even between nearby locations, whereas the distribution of air quality is much more uniform.

2) Quality Metric Definition: Based on the above-mentioned two assumptions, we define the data quality of multiple PS tasks as follows.

We first set an upper threshold $\text{Max}T_x$ for a specific task x, based on which we define the following concepts.

Definition 1 (Effective Amount of Samples): The effective number of samples for a specific task x on a subarea i and in a cycle j by the set of task-participant-cycle triples V is defined as follows:

$$E^x_{i,j}(V) = \min\left\{S^x_{i,j}(V), \text{Max}T_x\right\} \quad (2)$$

where $S^x_{i,j}(V)$ is the actual number of samples for a specific task x on a subarea i and in a cycle j by the set of task-participant-cycle triples V.

Definition 2 (Data Quality Index Per Cell): Based on Definition 1, the data quality of a specific task x on a spatial-temporal cell (subarea i and cycle j) by the set of task-participant-cycle triples V is defined as data quality index per cell (Cell_DQI) as follows:

$$\text{Cell\_DQI}^x_{i,j}(V) = 1 - \frac{\text{Max}T_x - E^x_{i,j}(V)}{\text{Max}T_x} = \frac{E^x_{i,j}(V)}{\text{Max}T_x}. \quad (3)$$

Based on Definitions 1 and 2, this paper proposes a novel metric to evaluate the data quality for multiple concurrent PS tasks, named data quality index (DQI), which is described as follows.

Definition 3 (Data Quality Index): Data quality of a specific task x by the set of task-participant-cycle triples V is defined as the average of the data quality per spatial-temporal cell. The mathematical formula of DQI of task x can be defined as follows:

$$\text{DQI}_x(V) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\text{Cell\_DQI}^x_{i,j}(V)}{M*N} \quad (4)$$

where M and N are the number of subareas and cycles, respectively. In this manner, the achieved data quality of a task x ranges from 0 to 1, where 0 indicates that requirements for task x are not satisfied at all, whereas 1 means that requirements for task x are fully satisfied.

C. Problem Formulation and Transformation

The goal of this paper is to select an appropriate set of task-participant-cycle triples, to maximize the DQI of each task under the total budget constraint, which is formulated as follows.

There are H participants on the platform, which are denoted by the set $P = \{p_1, p_2, \ldots, p_H\}$. There are a total of X PS tasks denoted by the set $T = \{t_1, t_2, \ldots, t_X\}$. Besides, the entire sensing duration is divided into Y equal-length cycles denoted as $C = \{c_1, c_2, \ldots, c_Y\}$. We denote all task-participant-cycle triples as a full set $V = T \times P \times C$, where$(t_j, p_i, c_k) \in V$. Thus, given a full set V and total budget B, the multitask assignment problem can be formulated as selecting a subset of V denoted as $V^*(V^* \subseteq V)$ with the following optimization goal and constraint:

$$\begin{aligned} \max \quad & F(V^*) = \left[\text{DQI}_1(V^*), \text{DQI}_2(V^*), \ldots, \text{DQI}_X(V^*)\right]^T \\ \text{s.t.} \quad & \sum_{(t_j,p_i,c_k)\in V^*} \text{cost}(t_j, p_i, c_k) \le B \end{aligned}$$

where $F(V^*)$ is a vector of objective functions, and each element is the DQI metric of a certain task collected by V^*.

Apparently, it is a multiobjective optimization (MOO) problem, whose optimal solution may not exist. Then, Pareto optimality can be used to describe solutions for MOO problems. A solution is Pareto optimal if it is not possible to move from that solution and improve at least one objective function, without detriment to any other objective function. A simple but efficient problem transformation for MOO problems is the weighted sum method [24]. By using it, we set weights w_i for each task and maximize the following composite objective function:

$$\text{ODQ}(V) = \sum_{i=1}^{X}\text{DQI}_i(V) * w_i \quad (5)$$

where w_i is the weight characterizing the importance of each task,[2] weight sum is kept as 1, and X is the total number of tasks. The objective function (5) is called the ODQ in this paper.

If all weights are positive, as assumed in [24], then maximizing the single-objective function (5) provides a Pareto optimal solution for the original MOO. Hence, the MOO problem can be converted to a single-objective optimization problem to select a set $V^*(V^* \subseteq V)$, denoted as

$$\begin{aligned} V^* = \arg\max_{v} \quad & \sum_{i=1}^{X}\text{DQI}_i(V) * w_i \\ \text{s.t.} \quad & \sum_{(t_j,p_i,c_k)\in V^*} \text{cost}(t_j, p_i, c_k) \le B. \end{aligned} \quad (6)$$

IV. MTPS Framework

A. Overview

Our framework, named MTPS, follows a centralized approach, where a central server collects and stores the volunteering mobile users' historical call traces in the target area, and the server selects task-participant-cycle triples. The assignments are predownloaded before PS task execution. After the task is executed, the assigned tasks are pushed to the participants when he/she connects to the cellular tower (calling/text sending). To solve the above multitask allocation problem, we propose a two-phase solution to determine the task assignment offline, which is shown in Fig. 2 and works as follows.

In phase 1, MTPS predicts each user's mobility and connection to the cellular tower, using the historical data collected from the telecom operator. Specifically, this step estimates the probability $\text{Pro}_{p,i,j}$ of the participant p connecting at least once to each cell tower s_i in cycle c_j.

In phase 2, MTPS takes an iterative greedy process to optimize the allocation. In each of the iterations, it incrementally selects task-participants based on the utility function, which is calculated by the prediction results in phase-1 and the estimated ODQ. The utility function is based on the *ODQ improvement per incentive cost*. To estimate the cost of each task-participant-cycle according to our proposed incentive model, we use different cost functions in each of the iterations. In the first iteration, we use a cost function considering only base incentive. Subsequent iterations use different cost functions to reselect a new set of task-participant-cycle triples. Phase 2 stops after the ith iteration, when the estimated ODQ of the combination set obtained is not optimal than that of the $(i-1)$th iteration.

[2] In realistic scenarios, the weights are relevant to many factors. This paper does not focus on how to determine the weights. Instead, we assume that it has already been predefined by the platform.

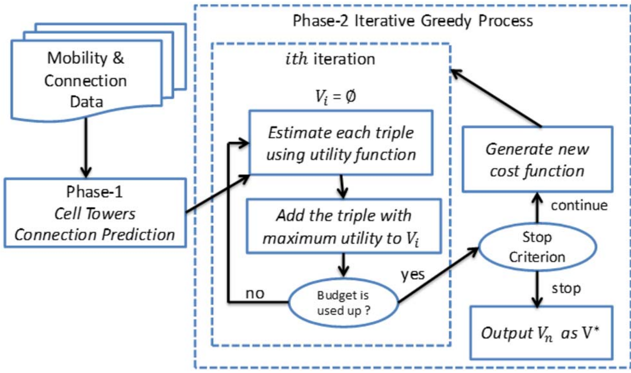


Fig. 2. MTPS: an overview.

B. Design Details

1) Cell Towers Connection Prediction: In this step, we predict the number of samples during cycle $c_j$at subarea s_i by calculating the probability of each participant connecting to the tower at least one time in each sensing cycle.

We first map each participant's historical call traces onto N sensing cycles. Then we count the average number of connection by each participant $p \in P$ at each cell tower s_i in each cycle c_j, which is denoted as $\lambda_{p,i,j}$. For example, to estimate $\lambda_{p,i,j}$ for sensing cycle c_j from 08:00 to 09:00 of a specific day, we count the average number of connections by p at s_i during the same period in the historical records.

Assuming that the connection sequence follows an inhomogeneous Poisson process [25], the probability of a participant p to connect n times to cell tower s_i in sensing cycle c_j can be modeled as

$$\text{Pro}_{i,j}(p, n) = \lambda^n_{p,i,j} * e^{\lambda_{p,i,j}} \big/ n!. \tag{7}$$

Therefore, we can estimate the probability of participant p connecting at least once during cycle c_j at s_i as follows:

$$\text{Pro}_{i,j}(p) = \sum_{n=1}^{\infty} \text{Pro}_{i,j}(p, n) = 1 - e^{-\lambda_{p,i,j}}. \tag{8}$$

Thus we predict that the probability of participant p providing one sample for each assigned task during cycle c_j at s_i as

$$\alpha_{i,j} = 1 - e^{-\lambda_{p,i,j}}. \tag{9}$$

2) Greedy Search for Each Iteration: The inner loop of phase 2 (see Fig. 2) shows the process of ith iteration ($i = 1, 2, \ldots$). Given the full task-participant-cycle triples set V, total budget B, and a utility function, the algorithm selects a set of task-participant-cycle triples incrementally as follows.

1) The algorithm first initializes the temporal task-participant-cycle triples set $V_i = \emptyset$, and selects the single task-participant-cycle triple $(t_j, p_i, c_k) \in V$ with the maximum utility and adds it into the solution, that is, $V_i = V_i \cup \{(t_j, p_i, c_k)\}$.
2) The algorithm then selects one unselected task-participant-cycle triple $(t_{j'}, p_{i'}, c_{k'}) \in V \backslash V_i$ having the maximal utility when combining with V_i using Utility$_i$ and adds it into the solution, that is, $V_i = V_i \cup \{(t_{j'}, p_{i'}, c_{k'})\}$.
3) The algorithm calculates the remaining budget $B_{\text{remain}}(V_i)$. Then the algorithm continues selecting another unselected combination until the remaining budget is insufficient to select one more combination.

At the end of the ith iteration, the algorithm obtains a set of task-participant-cycle triples (i.e., V_i) with the given budget B.

3) Utility Function Calculation: As this paper adopts a fine-grained attention-compensated incentive model, different task-participant-cycle triples have different corresponding cost. Therefore, we should define the utility increase of a certain task-participant-cycle triple as the ODQ improvement per incentive cost. Specifically, during the ith iteration ($i = 1, 2, 3, \ldots$), given the set of incrementally selected task-participant-cycle triples V_i, the utility for adding a triple $(t_{j'}, p_{i'}, c_{k'}) \in V \backslash V_i$ is calculated as follows:

$$\begin{aligned}\text{Utility}_i\big((t_{j'}, p_{i'}, c_{k'})\big|V_i\big) \\ = \frac{\Big\{\text{DQI}_{t_{j'}}\big(\big\{(t_{j'}, p_{i'}, c_{k'})\big\} \bigcup V_i\big) - \text{DQI}_{t_{j'}}(V_i)\Big\} * w_{t_{j'}}}{\text{cost}\big(t_{j'}, p_{i'}, c_{k'}\big)}\end{aligned} \tag{10}$$

where $\text{cost}(t_{j'}, p_{i'}, c_{k'})$ is the submodular incentive cost of the triple $(t_{j'}, p_{i'}, c_{k'})$.

By adopting the proposed incentive model, the incentive cost of a certain task-participant-cycle triple depends on whether the participant has completed the task in previous cycles. As we cannot foresee this in the searching process, it is difficult to compute the incentive cost of each combination. Thus, we must use different cost functions in each of the iterations.

1) *Cost Function Calculation for the First Iteration:* In the first iteration, we do not consider the bonus, so that the cost function of $(t_{j'}, p_{i'}, c_{k'})$ is calculated as $\text{cost}_i(t_{j'}, p_{i'}, c_{k'}) = \alpha_{t_{j'}} * K$, where $\alpha_{t_{j'}}$ is the base incentive of task $t_{j'}$.
2) *Cost Function Calculation for the ith Iteration ($i \geq 2$):* During the ith iteration ($i = 2, 3, \ldots$), we consider the bonus incentive when calculating the cost function. Specifically, we evaluate the cost of a triple [i.e., $\text{cost}_i(t_{j'}, p_{i'}, c_{k'})$] in the ith iteration based on the obtained triple set in the $(i-1)$th iteration using the following rules.
   a) If the triple $(t_{j'}, p_{i'}, c_{k'})$ has been selected in the $(i-1)$th iteration, then its estimated incentive cost is calculated by the decrease of estimated total cost in the $(i-1)$th iteration with the deletion of triple $(t_{j'}, p_{i'}, c_{k'})$.
   b) Else, its incentive cost is estimated through two steps on the total estimated incentive cost in the $(i-1)$th iteration: first, add the cost of triple $(t_{j'}, p_{i'}, c_{k'})$ and then minus the cost of other triples.

The mathematical formula is defined as follows:

$$\mathrm{cost}\left(t_j', p_i', c_k'\right) = \begin{cases} \mathrm{TC}(V_{i-1}) - \mathrm{TC}\left(V_{i-1} \backslash \left\{\left(t_{j'}, p_{i'}, c_{k'}\right)\right\}\right) \\ \qquad \text{if } \left(t_{j'}, p_{i'}, c_{k'}\right) \in V_{i-1} \\ \mathrm{TC}(V_{i-1}) + \mathrm{cost}\left(t_{j'}, p_{i'}, c_{k'}\right) \\ - \sum_{e \in V_{i-1}} \{\mathrm{TC}(V_{i-1}) - \mathrm{TC}(V_{i-1} \backslash \{e\})\}, \text{ else} \end{cases} \tag{11}$$

where the cost function $\mathrm{TC}(V)$ is the estimated total cost of a given task-participant-cycle triples V.

4) Overall Data Quality Estimation: As Fig. 2 illustrates, the stopping criterion is based on the comparison of the estimated ODQ between the current iteration and the previous one. Thus we need to estimate the ODQ for a given output task-participant-cycle triples set at the end of each iteration.

Given a set of task-participant-cycle triples set V', the ODQ it achieves is estimated as follows:

$$\mathrm{ODQ}\left(V'\right) = \sum_{x=1}^{G} \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \min\left\{S_{i,j}^{x}\left(V'\right), \mathrm{Max}T_x\right\}}{M * N * \mathrm{Max}T_x} * w_x \tag{12}$$

where $S_{i,j}^{x}(V)$ and $\mathrm{Max}T_x$ are defined in Section III, and G, M, and N are the total number of tasks, subareas, and cycles, respectively. Based on the mobility prediction result in phase-1, $S_{i,j}^{x}(V')$ is estimated as

$$S_{i,j}^{x}\left(V'\right) = \sum_{(x, p_{i'}, j) \in V'} \left(1 - e^{-\lambda_{p_{i'}, i, j}}\right). \tag{13}$$

5) Algorithm Analysis: We analyze the above iterative greedy approach in terms of the optimization goal and computation complexity, respectively. As approached in Propositions 1 and 2, both the objective function $\mathrm{ODQ}(V)$ and total cost function $\mathrm{TC}(V)$ are submodular functions over set V. Therefore, according to the theory of submodular function maximization under the submodular knapsack constraint [26], our approach can achieve the $(1 - e^{-1})$ near-optimality bound. For example, if the brute-force enumeration can obtain an ODQ of 0.9 with a certain total budget, then our approach can achieve an ODQ of $(1 - e^{-1}) * 0.9 \approx 0.57$ with the same budget.

Proposition 1: The overall data quality $\mathrm{ODQ}(V)$ is a submodular function.

Proof: According to the diminishing return property of submodular functions [26], $\mathrm{Cell\_DQI}_{i,j}^{x}(V)$ is a submodular over the set V, Therefore, based on the linear property[3] of submodular function, the $\mathrm{ODQ}(V)$, as the linear sum of $\mathrm{Cell\_DQI}_{i,j}^{x}(V)$, is a submodular function. ■

Proposition 2: The total cost function $\mathrm{TC}(V)$ is a submodular function.

Proof: Denote T as the set of all tasks, $t \in T$, P is the set of all participants, $p \in P$, and C is the set of all cycles, $c \in C$. Denote $V = T \times P \times C$, $A \subseteq B \subseteq V$, and $(t, p, c) \in V \setminus B$. Denote the task-specific base incentive of task t as α_t, so that based on the definition in Section III, the first-time participation bonus incentive of task t is $\tau * \alpha_t$. If the task t in combination (t, p, c) is new to all the elements of B, then $\mathrm{TC}(A \cup \{(t, p, c)\}) - \mathrm{TC}(A) = (1 + \tau) * \alpha_t$, and $\mathrm{TC}(B \cup \{(t, p, c)\}) - \mathrm{TC}(B) = (1 + \tau) * \alpha_t$. Otherwise, $\mathrm{TC}(A \cup \{(t, p, c)\}) - \mathrm{TC}(A) = \alpha_t$ or $(1 + \tau) * \alpha_t$, and $\mathrm{TC}(B \cup \{(t, p, c)\}) - \mathrm{TC}(B) = \alpha_t$. Thus, in all cases, it holds that $\mathrm{TC}(A \cup \{(t, p, c)\}) - \mathrm{TC}(A) \geq \mathrm{TC}(B \cup \{(t, p, c)\}) - \mathrm{TC}(B)$. According to the definition in [26], *the total cost function* $\mathrm{TC}(V)$ *is a submodular function*. ■

The described iterative greedy process consists of two nested loops. In our experimental study, which was also demonstrated in [26], the outer loop commonly runs 5–8 iterations and runs two iterations in the best case. The inner loop runs $H * X * Y$ iterations in the worst case (the budget is extremely adequate, so that all task-participant-cycle triples are selected), where H, X, and Y are the number of participants, tasks, and cycles, respectively.

[3]The linear sum of multiple submodular functions is still a submodular function.

V. Evaluation

In this section, we report the evaluation results using large-scale real-world cell tower connection traces to verify the effectiveness of our approach. We first introduce baseline methods for evaluation. Then we present the datasets and the basic experiment settings. Finally, the detailed evaluation results with respect to the baseline methods are presented and compared.

A. Baselines

We provide the following three baseline task allocation methods for comparative studies.

1) Random Allocation: This method randomly selects task-participant-cycle triples one by one, until the total budget is used. We repeat the random allocation process 20 times, and select the best achieved ODQ as the final result.

2) MaxODQ: This method adopts a greedy algorithm without considering the cost. Thus, it incrementally selects task-participant-cycle triples based on the ODQ *improvement*, until the total budget is used up.

3) MaxODQ/Cost: This method adopts similar greedy process as MaxODQ, but with the consideration of the cost when calculating the utility of a task-participant-cycle triple. It incrementally selects triples based on the ODQ *improvement per incentive cost*. However, different from our approach, this method does not execute iteratively to approximate the incentive cost of each triple.

B. Datasets and Experiment Setups

The dataset used for evaluation is the D4D dataset [27], which contains 50 000 users' phone records of connections to the cell towers (calling or sensing text messages) from Cote d'Ivoire, and each record includes user id, connection time, and cell tower. All these users are reselected randomly every two weeks with anonymized user ids and totally 10 two-week periods of records are stored in the dataset. In each two-week period, our experiment uses the records in the first week for task allocation, and we tested the overall system utility of selected task assignment instance set using records in the

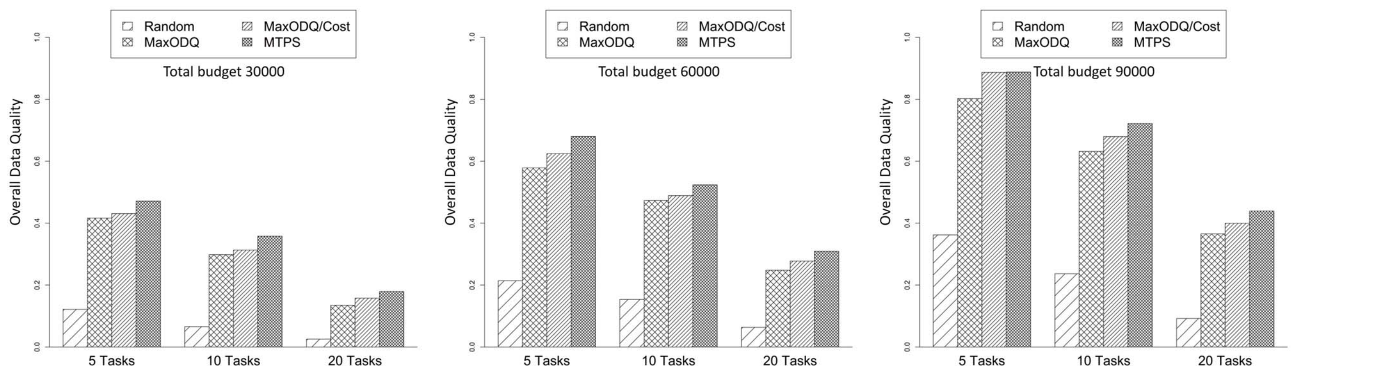


Fig. 3. ODQ comparison.

second week. Specifically, we extracted records of the downtown area (20 cell towers with 1500 mobile users). We further assume that each task executes for five days from Monday to Friday in a week, runs ten cycles every working day from 8:00 to 18:00, with each cycle lasting 1 h. Thus the total period consists of 50 cycles. After the datasets are extracted, there are several parameters that need to be set. First, we evaluate our approach and baseline methods under the different total incentive budget and number of tasks.

1) We set the total budget as 30 000, 60 000, and 90 000 RMB, respectively, and we randomly generated the task-specific base incentive of each task from 1 to 3 RMB. The task switching bonus incentive is set to be equal to the task-specific base.
2) We create 5, 10, and 20 virtual PS tasks with randomly generated weights from 0 to 1 and threshold (i.e., MaxT_x defined in Section III) from integer 1 to 5.

We repeated the experiment 20 times for each setting of total budget and number of tasks, obtained the selected triples on the training data (the first week), and calculated the ODQ achieved on the testing data (the second week).

C. Overall Data Quality Comparison

Fig. 3 shows the ODQ of different methods in various settings. Because of space limitation, we show only the average evaluation results of the targeted region. From the comparisons shown in Fig. 3, we can observe the following.

1) In all settings and datasets, our approach outperforms the baseline methods. Specifically, our approach gained on average 323% higher overall utility than random allocation, 18% higher than MaxODQ, and 9% higher than MaxODQ/Cost.
2) The advantage of our approach is related to the total budget and the number of tasks, especially compared to the random allocation. With the increase of the number of tasks and decrease of total budget, the advantage of our approach compared to the random allocation becomes more obvious. This is reasonable because the relationship between the number of tasks and the total budget has a significant impact on the importance of selecting task-participant-cycle triples. Let us take an extreme case as the example: if the budget is adequate enough to select all combinations, then all the methods can achieve the optimal solution (i.e., selecting all triples).

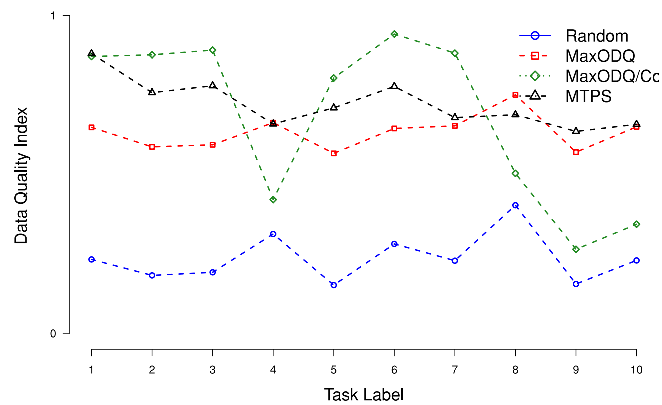


Fig. 4. Data quality of each single task.

D. Data Quality of Each Task

In this section, we evaluate the data quality of each single task (i.e., DQI) for our approach and baseline methods. Due to space limitation, we only show the results in Fig. 4, where there are ten equally important tasks and the total shared budget is set as 90 000 RMB.[4]

From the comparison shown in Fig. 4, we can observe the following.

1) If the tasks are randomly allocated, the achieved DQI of each single task is significantly lower than MTPS. Thus, the ODQ is obviously lower than MTPS.
2) For most of the tasks, the achieved DQI of MaxODQ is lower than MTPS, thus, the ODQ shows the same trend.
3) For the MaxODQ/Cost method, the variance of each task's achieved DQI is bigger than MTPS.

Certain tasks achieve a DQI that is near 0.95, while some others only achieve a DQI around 0.25. The ODQ obtained by MTPS is, however, larger than MaxODQ/Cost.

E. Number of Assigned Tasks Per Participant

Instead of selecting a subset of participants and allocating all tasks to them, we study a more "fine-grained" multiallocation by assigning a subset of tasks to each participant per cycle, which is an important feature of MTPS. However, in extreme cases, if MTPS assigns all tasks to the same set of participants

[4]The results under other settings indicate similar pattern.

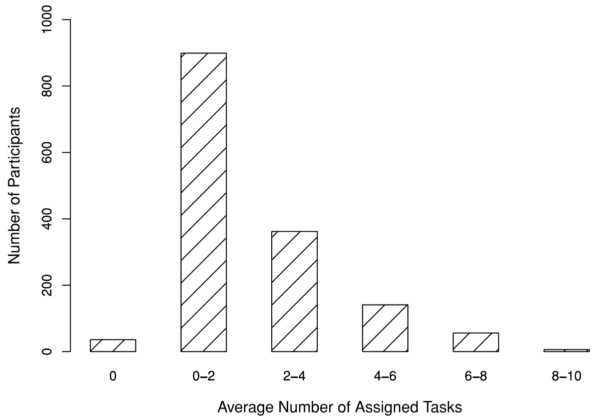


Fig. 5. Distribution of the number of assigned tasks.

every cycle, then MTPS will degenerate to the coarse-grained participant selection. Thus, we calculate the average number of assigned tasks per cycle for each participant. Fig. 5 shows the result under the setting that there are ten tasks and total budget is 90 000 RMB.[5] From the figure, we can see that MTPS will not assign all tasks to the same set of participants. On the contrary, most of the participants are involved and assigned with different average number of takes per cycle.

F. Different Density of Participants

In this section, we evaluate the impact of participants' density on the performance of different approaches when fixing the budget as 90 000 and the number of tasks as 10. We vary the total number of participants from 500, 1000, 1500 to 2000. Fig. 6 shows the ODQ under different number of participants. We can see the following.

1) The performance of all approaches increases as the number of participants gets bigger, because more assignment options are provided under the same budget constraint.
2) MTPS still outperforms other baseline methods when varying the density of participants.

G. Computation Time

In this section, we evaluate the computation time of our approach and baseline methods, and show how fast each method could accomplish the multitask allocation procedure. We carried out experiments using a laptop with an Intel Core i7-4710HQ Quad-Core CPU and 16 GB memory. Different algorithms were implemented with the Java SE platform on a Java HotSpotTM 64-Bit Server.

Table II presents the average time consumed using the above dataset with the setting of 20 tasks and 90 000 RMB. From Table II, we can see that although our approach took longer time than other methods, we believe it is worthy, because the improved performance is for the overall utility of multiple tasks rather than a single task, so that the performance improvement should be considered in the multitask platform context with significant number of tasks. As the task allocation process is completed offline and the sequential algorithm was run on a laptop, shorter computation time can be easily achieved by running on a more powerful computer or using parallel algorithms.

[5]The evaluations under other settings show the similar result.

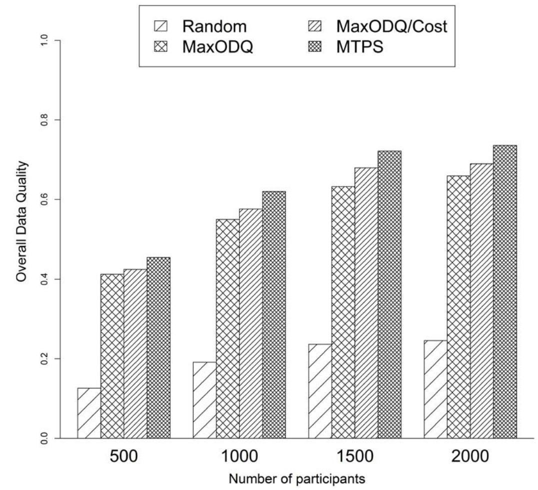


Fig. 6. Performance comparison under different participants' density.

TABLE II
COMPUTATION TIME EVALUATION RESULT IN MINUTES

Our Approach	Random Allocation	MaxODQ	MaxODQ/Cost
14.1	0.6	3.7	4.9

VI. CONCLUSION

This paper proposes a fine-grained multitask allocation framework for PS under a total budget constraint. It operates on an attention-compensated incentive model, which takes the extra compensation for completing new tasks into account. It first predicts each user's mobility and connection to the cell towers using the historical data collected from the telecom operator. It then takes an iterative greedy process to select a set of task-participant-cycle triples. Extensive evaluation based on real-world mobility traces has demonstrated that our approach outperforms the baseline methods.

In future research, we intend to improve MTPS in the following aspects.

1) *More Complex Constraints and Models:* We plan to improve upon this research by considering more complex constraints, which may exist in some applications. For example, because completing PS tasks usually requires explicit human involvement, the number of assigned tasks for each participant in a cycle has limits. Moreover, a participant may not be able to complete a certain task as his/her mobile device may not be equipped with a corresponding sensor. Another area of improvement is developing more sophisticated attention-compensated incentive models. We proposed and used a simple model that gives additional incentives for new tasks, but we realize that even with all non-new tasks, juggling more tasks concurrently require more attention than the sum of attentions required by individual tasks.

2) *Privacy Preserving Mechanisms:* To obtain the cell tower connection prediction models, our approach currently collects and analyzes the raw historical data of mobile users. This, however, leads to privacy concerns. One way to reduce the privacy concerns is to provide predictive models only rather than raw data. Or the client software running on the user's device can collect raw data, but only upload predictive models for task allocation.
3) *Heterogeneity of Multiple Tasks:* This paper assumes that all sensing tasks share the same target sensing region and cycles. However, in real-world application scenarios, different tasks may be required to executed at different locations or with different length of cycles. Thus, we will consider the heterogeneity of multiple tasks in both the spatial and temporal dimensions in future work.

REFERENCES

[1] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, 2006.

[2] D. Zhang, H. Xiong, L. Wang, and G. Chen, "Crowdrecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Seattle, WA, USA, 2014, pp. 703–714.

[3] J. A. Burke *et al.*, "Participatory sensing," in *Proc. Center Embedded Netw. Sens.*, Boulder, CO, USA, 2006, pp. 117–134.

[4] B. Guo *et al.*, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surveys*, vol. 48, no. 1, p. 7, 2015.

[5] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Pervasive Computing*. Heidelberg, Germany: Springer, 2010, pp. 138–155.

[6] A. Singla and A. Krause, "Incentives for privacy tradeoff in community sensing," in *Proc. 1st AAAI Conf. Human Comput. Crowdsourcing*, Palm Springs, CA, USA, 2013, pp. 1–12.

[7] G. Cardone *et al.*, "Fostering participAction in smart cities: A geo-social crowdsensing platform," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 112–119, Jun. 2013.

[8] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, "EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 39, 2015.

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

[10] S. Hachem, A. Pathak, and V. Issarny, "Probabilistic registration for large-scale mobile participatory sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, San Diego, CA, USA, 2013, pp. 132–140.

[11] B. Zhang, Z. Song, C. H. Liu, J. Ma, and W. Wang, "An event-driven QoI-aware participatory sensing framework with energy and budget constraints," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 42, 2015.

[12] 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.

[13] H. Xiong, D. Zhang, G. Chen, L. Wang, and V. Gauthier, "Crowdtasker: Maximizing coverage quality in piggyback crowdsensing under budget constraint," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, St. Louis, MO, USA, 2015, pp. 55–62.

[14] S. Reddy *et al.*, "Using context annotated mobility profiles to recruit data collectors in participatory sensing," in *Location and Context Awareness*. Heidelberg, Germany: Springer, 2009, pp. 52–69.

[15] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi, "EMC^3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint," *IEEE Trans. Mobile Comput.*, vol. 14, no. 7, pp. 1355–1368, Jul. 2015.

[16] D. Philipp, J. Stachowiak, P. Alt, F. Dürr, and K. Rothermel, "DrOPS: Model-driven optimization for public sensing systems," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, San Diego, CA, USA, 2013, pp. 185–192.

[17] L. Wang *et al.*, "CCS-TA: Quality-guaranteed online task allocation in compressive crowdsensing," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Osaka, Japan, 2015, pp. 683–694.

[18] Z. Song *et al.*, "QoI-aware energy-efficient participant selection," in *Proc. 11th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, Singapore, 2014, pp. 248–256.

[19] F. Yang *et al.*, "Heterogeneous task allocation in participatory sensing," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, 2015, pp. 1–6.

[20] H. Li, T. Li, and Y. Wang, "Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks," in *Proc. IEEE 12th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Dallas, TX, USA, 2015, pp. 136–144.

[21] X. Lu, D. Li, B. Xu, W. Chen, and Z. Ding, "Minimum cost collaborative sensing network with mobile phones," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, 2013, pp. 1816–1820.

[22] Y. Liu *et al.*, "TaskMe: Multi-task allocation in mobile crowd sensing," in *Proc. 2016 ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp 2016)*, Heidelberg, Germany, Sep. 12–16, 2016.

[23] B. Guo, Y. Liu, W. Wu, Z. Yu, and Q. Han, "Activecrowd: A framework for optimized multitask allocation in mobile crowdsensing systems," *IEEE Trans. Human–Mach. Syst.*, to be published.

[24] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: New insights," *Struct. Multidiscipl. Optim.*, vol. 41, no. 6, pp. 853–862, 2010.

[25] J. Weinberg, L. D. Brown, and J. R. Stroud, "Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data," *J. Amer. Stat. Assoc.*, vol. 102, no. 480, pp. 1185–1198, 2007.

[26] R. K. Iyer and J. A. Bilmes, "Submodular optimization with submodular cover and submodular knapsack constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2436–2444.

[27] V. D. Blondel *et al.* (2012). Data for development: The d4d challenge on mobile phone data. *Comput. Sci.*

Jiangtao Wang received the Ph.D. degree from Peking University, Beijing, China, in 2015.

He is currently a Post-Doctoral Researcher with the Institute of Software, School of Electronics Engineering and Computer Science, Peking University. His current research interests include ubiquitous computing, urban data analytics, and software engineering.

Yasha Wang received the Ph.D. degree from Northeastern University, Shenyang, China, in 2003.

He is a Professor and the Associate Director of the National Research and Engineering Center of Software Engineering, Peking University, Beijing, China. He has authored or co-authored over 50 papers in prestigious conferences and journals, such as ICWS, UbiComp, and ICSP. He has accomplished several key national projects on software engineering and smart cities, as a Technical Leader and a Manager. Cooperating with major smart-city solution providing companies, his research has been adopted in over 20 cities in China. His current research interests include urban data analytics, ubiquitous computing, software reuse, and online software development environment.

Daqing Zhang received the Ph.D. degree from the University of Rome "La Sapienza," Rome, Italy, in 1996.

He is a Professor with Peking University, Beijing, China, and Telecom SudParis, Évry, France. His current research interests include context-aware computing, urban computing, and mobile computing.

Prof. Zhang served as the General or Program Chair for over ten international conferences. He is an Associate Editor for *ACM Transactions on Intelligent Systems and Technology*, the IEEE TRANSACTIONS ON BIG DATA, and others.

Leye Wang received the B.Sc. and M.Sc. degrees in computer science from Peking University, Beijing, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Institut Mines-Télécom/Telecom SudParis, Évry, France, and Université Pierre et Marie Curie, Paris, France, in 2016.

His current research interests include mobile crowdsensing, social networks, and intelligent transportation systems.

Haoyi Xiong received the Ph.D. (highest Hons.) degree in computer science from Telecom SudParis, Évry, France, and the University of Paris VI, Paris, France, in 2015.

He is currently an Assistant Professor with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO, USA (formerly the University of Missouri at Rolla). From 2015 to 2016, he was a Post-Doctoral Research Associate with the Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA, USA. His current research interests include ubiquitous computing, cyber-human systems, and large-scale optimization and decision making.

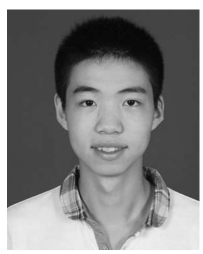
Dr. Xiong was a recipient of the Best Paper Award of IEEE UIC 2012 and the Outstanding Ph.D. Thesis Runner-Up Award of CNRS SAMOVAR 2015.

Abdelsalam (Sumi) Helal (F'15) received the Ph.D. degree in computer sciences from Purdue University, West Lafayette, IN, USA.

He is currently a Professor with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA, where he is also the Director of the Mobile and Pervasive Computing Laboratory. His current research interests include pervasive and mobile computing, Internet of Things, smart spaces, smart city concepts, smart health and technologies for successful aging, and persuasive computing.

Yuanduo He received the bachelor's degree from Peking University, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree at the Institute of Software, School of Electronics Engineering and Computer Science.

His current research interests include ubiquitous computing, mobile computing, and data mining.



Feng Wang is currently pursuing the undergraduation degree at the School of Electronic Engineering and Computer Science, Peking University, Beijing, China.

His current research interest includes mobile crowd sensing.