Optimal policy for composite sensing with crowdsourcing

Bei Zhao, Siwen Zheng and Jianhui Zhang

Abstract

The mobile crowdsourcing technology has been widely researched and applied with the wide popularity of smartphones in recent years. In the applications, the smartphone and its user act as a whole, which called as the composite node in this article. Since smartphone is usually under the operation of its user, the user's participation cannot be excluded out the applications. But there are a few works noticed that humans and their smartphones depend on each other. In this article, we first present the relation between the smartphone and its user as the conditional decision and sensing. Under this relation, the composite node performs the sensing decision of the smartphone which based on its user's decision. Then, this article studies the performance of the composite sensing process under the scenario which composes of an application server, some objects, and users. In the progress of the composite sensing, users report their sensing results to the server. Then, the server returns rewards to some users to maximize the overall reward. Under this scenario, this article maps the composite sensing process as the partially observable Markov decision process, and designs a composite sensing solution for the process to maximize the overall reward. The solution includes optimal and myopic policies. Besides, we provide necessary theoretical analysis, which ensures the optimality of the optimal algorithm. In the end, we conduct some experiments to evaluate the performance of our two policies in terms of the average quality, the sensing ratio, the success report ratio, and the approximate ratio. In addition, the delay and the progress proportion of optimal policy are analyzed. In all, the experiments show that both policies we provide are obviously superior to the random policy.

Keywords

Conditional sensing, crowdsourcing, partially observable Markov decision process, smartphone

Introduction

With the proliferation of personal smart devices, such as smartphone, human is able to capture information/event from the physical world with smartphones more easily than before.[1–4] Embedded with a rich set of sensors, the current smartphone can support increasing applications across a wide variety of domains, such as crowdsensing,[1,5–7] environmental monitoring,[8] and social networks.[9] These applications can be classified into two major classes: participatory sensing (user is directly involved) and the opportunistic sensing (user is not involved).[5,10,11] In the participatory sensing, user can act as the preliminary sensor and decision-maker before his or her smartphone implements a certain sensing task. For example, users make decisions whether to take part in an application, and then operate his or her smartphone to implement the application.[12–14]

College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China

Corresponding author:
Jianhui Zhang, College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China.
Email: jhzhang@ieee.org

Most of the previous works on crowdsensing take the smartphone into consideration, only a small part works suggest that crowdsensing should also include user as the sensor instead of just sensor carrier and operator.[15–17] For example, Wang et al.[16] took human as sensor and studied their behavior's affecting the sensing data quality. But there are a few articles noticed that humans and their smartphones depend on each other. There are two questions should be focused on the relationship with humans and their smartphones. The first is how to describe the relation between the smartphone and its user during smartphone sensing. The second is how two improve the performance of the smartphone sensing by exploiting the relation. As we all know, human has more powerful ability of recognition than the smart device and plays a key role before the process of smartphone sensing. In this article, we propose a framework to clarify the relation, and then study the performance improvement of the crowdsensing under a scenario, where users are willing to have good experience to take part in the crowdsensing. Since smartphones are under control of its user, its sensing decision is made after its user's willingness. We design the framework as *conditional sensing* as shown in Figure 1, where each user takes the action "sleeping" if he or she is not willing to taking part in the smartphone sensing. The scenario studied in this article represents a class of common applications in the participatory sensing, where some users are asked to implement a certain task, such as to detect the interesting object/event around them. We further investigate the case where the users have limited cost to implement the task, and hope a certain success implementation probability, denoted by ζ.

Summary of key contributions

The key contributions of this article are listed as follows:

1. This article studies the relationship between human and smartphone during the smartphone sensing, and proposes the framework: composite sensing.
2. We study the scenario of the object detection, and formulate the composite sensing problem, that is, how to improve the user experience under the framework of composite sensing as the partially observable Markov decision process (POMDP). We also design a new scheme, called composite sensing policy, to solve the composite sensing problem and get the maximal overall sensing quality.
3. We provide the theoretical and experimental analysis for the composite sensing policy. The
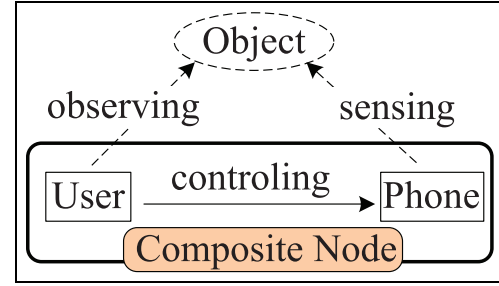

Figure 1. Compound node.

theoretical optimization of the policy is guaranteed while the experimental results certificate the performance of the optimal and myopic policies we proposed.

Road map

This article is organized as follows. The related works are reviewed in section "Related work." Section "Preliminaries" presents the composite sensing and system models. We formulate the composite sensing problem and map it as the POMDP in section "Composite sensing problem." The composite sensing policy for the problem is designed and the theoretical performance of the policy is presented in section "Composite sensing policy." The performance of our solution is also evaluated by the experiment in section "Experiment results." The work of the whole article is summarized in section "Conclusion."

Related work

Today's smartphone is embedded in a number of specialized sensors, including camera, global positioning system (GPS), digital compass, and so on. It can sense the environmental information, and share the information with the friend of the smartphone holder or report to a certain server.[13] It has become not only the core communication device in people's daily life but also a smart sensing device for environmental monitoring, smart transportation systems, social networks, and so on.[10] Its applications are thus widely exploited and are extended to many more areas than before. According to the awareness and involvement of the user in the architecture as sensing device custodians, the smartphone applications can be classified into two major classes: participatory sensing (user is directly involved) and the opportunistic sensing (user is not involved).[10] The participatory sensing includes both the smartphone and its holder into the significant decision stages in the sensing application. One type of relation between the

smartphone and its holder is the composite sensing proposed in this article.

Participatory sensing

A wide range of environmental information, such as road traffic, can be sensed and disseminated by ordinary citizens with smartphones. It brings a new way for the development of many application areas, such as environmental monitoring and social networks. The interesting examples include road traffic monitoring,[18] SmartPhoto,[17] and Ear-phone.[14] Rana et al.[14] designed an *end-to-end* participatory urban noise mapping system called Ear-phone. The key idea of Ear-phone is to crowdsource the collection of urban noise to people, who carry smartphones equipped with sensors and location-providing GPS receivers. In the end-to-end system, the urban noise is sent to a central server. A noise map is reconstructed and then is provided to the end user. In VTrack, some participatory drivers with smartphone send its location estimated by WI-FI or GPS to a central server in real time, and the server provides the real-time routes with the minimal travel time to users.[19] Mohan et al.[18] have presented TrafficSense to monitor road and traffic conditions in a setting where there are much more complex varied road conditions (e.g. potholed roads), chaotic traffic (e.g. a lot of braking and honking), and a heterogeneous mix of vehicles (two wheelers, three wheelers, cars, buses, etc.). Wang et al.[17] proposed a framework, called SmartPhoto, to quantify the quality (utility) of crowdsourced photos based on the accessible geographical and geometrical information (called metadata), including the smartphone orientation, position, and all related parameters of the built-in camera. The sensed photos are sent to a server by the participators and different rewards are feedback to them because the smartphone orientation and position cause the different sensing qualities. There are increasingly new applications appearing, such as CrowdAtlas, for generating a high quality map by crowdsourcing.[20] For more details on smartphone sensing, we refer interested readers to the survey articles.[2,10] From the observation from the related works on smartphone applications, we can find the following features: (1) sensing result report: many smartphone applications require the participators to report their sensed information to central servers; and (2) human acts sensor: in the smartphone applications with the participatory sensing, human is a key part of the systems in these applications, and makes key stages of the decision to sense the environmental information. Not all users are willing to be participators and not all of their sensing results have equal value because the smartphone types and sensing conditions may be different.[13,16]

Human as sensor

Human's decision is the necessary part of the smartphone applications with the participatory sensing, and has great affection on the sensing result. For example, SmartPhoto needs humans to observe the Event of Interesting (EoI) and then take pictures.[17] Most of the current smartphone sensing applications are based on voluntary participation.[13,21] In these applications,[13] humans estimate the incentive reward at first, and then operate their smartphone to participate if satisfied or they observe the EoI at first, and then decide to collect and report the information about the EoI if it is observed and satisfies requirement.[14,18] Zhao et al. have showed that mobile crowdsourced sensing (MCS) is a new paradigm that takes advantage of pervasive smartphones to efficiently collect data, enabling numerous novel applications. They proposed incentive mechanisms which are necessary to attract more user participation to achieve good service quality for an MCS application.[21] ND Lane et al. have surveyed some existing mobile phone sensing algorithms, applications, and systems. They also discussed the emerging sensing paradigms, and formulated an architectural framework for discussing a number of the open issues and challenges emerging in the new area of mobile phone sensing research.[2] The smartphones' decisions base on their users' observation and decision. It is an underlying phenomenon in the applications of smartphone sensing. Wang et al.[16] used humans as sensors, and studied their decisions affecting the sensing data quality. Although human makes a key decision in the smartphone applications with participatory sensing, most of the previous works make simply an assumption on human's decision or ignore the humans' decision. Furthermore, the participator's decision and its relationship with its smartphone are fairly considered and researched.

Preliminaries

Object, observing, and sensing model

This article concerns a set V of composite nodes to sense a set of m objects. The object in this article can be a target, such as the famous building,[17] and the EoI, such as the cellular or the Wi-Fi signal.[22] As shown in Figure 2(a), each object is assumed to have an orientation, and K aspects. Let the parameter θ, $\theta \in \{1, \ldots, K\}$, denotes the aspect that facing one node. For example, $\theta = 2$ means that the second aspect of the object o_j faces the node. When the node takes the action to sense one θ of the object's aspects, the action results in a certain sensing quality $q(\theta)$, $0 \leqslant q(\theta) \leqslant 1$. In this article, the sensing quality is defined as the function of the aspect as given by the following equation

$$q(\theta) = \frac{1 - \theta}{K}, \quad \theta = 1, \ldots, K \qquad (1)$$
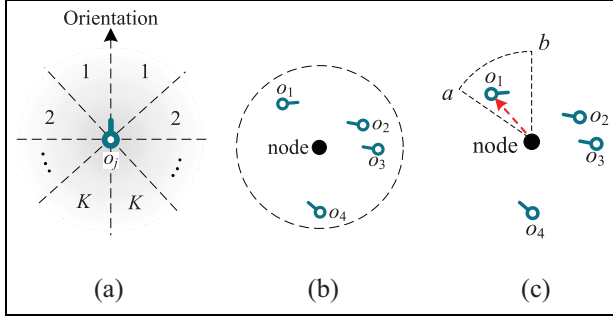
Figure 2. System model: (a) object model, (b) observing model, and (c) sensing model.

Figure 3. Each node has four directions to choose: (a) o_1, (b) o_2, (c) o_3, and (d) o_4.

Each user's observing range is modeled as a disk as shown in Figure 2(b) and the smartphone's sensing range is modeled as a fan-shaped sensing area in Figure 2(c). They have the same radius since the user would not notice the object out of the observing range. The smartphone can fix a direction to sense one of the objects in its sensing range as shown in Figure 2(c). Let the object ID denotes the direction that the node chooses. The example in Figure 3 shows that the node has the directions as many as the number of the objects.

Conditional sensing

In the crowdsourcing applications with the participatory sensing, the smartphone must be under the control of its user. Each user acts the preliminary sensor, and implements the composite operation with his or her smartphone as a whole. We call such a whole as a *composite node* (*node* in brief) as shown in Figure 1. In each node, the user can make observing decision $\alpha \in \{0(sleeping), 1(observing)\}$ to observe the state of the objects in the composite node's sensing range, and then the smartphone can make sensing decision $\beta = \{0(non - sensing), 1(sensing)\}$. The node implements the composite sensing: *conditional decision-making*. The sensing decision is based on the observing decision as shown in Figure 4. By the observing decisions $\alpha = 0$, the node sleeps. Otherwise, the user observes the objects' states, and obtains the observation outcome $\Theta_{j,k}(\tau)$: $\theta_j(\tau) = k$, where τ is the time slot in the period T. Given the observation outcome $\Theta(\tau)$, the smartphone makes the sensing decision. If the sensing decision is $\beta_{j,k}(\tau) = 1$, the smartphone chooses the direction o_j object to sense. Otherwise, the node turns to sleep. The observing and sensing decisions compose the decision space A, that is, $A = \{\alpha, \beta\}$.

In this following context, we present the composite sensing from the view of an arbitrary node. The objects refer to these in the sensing range of the node.
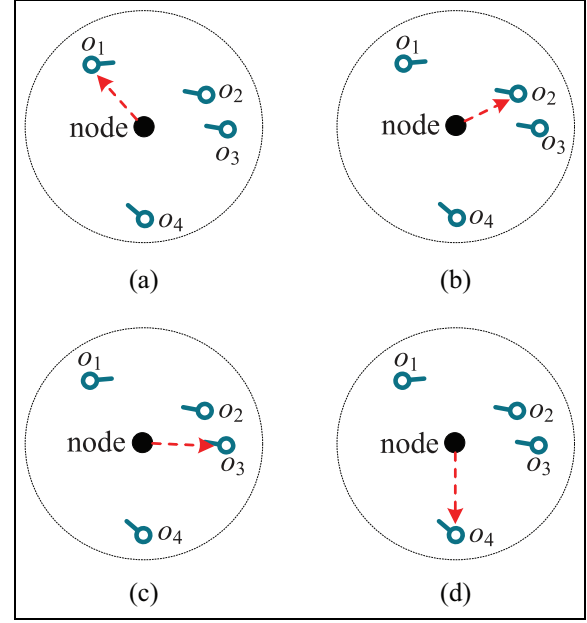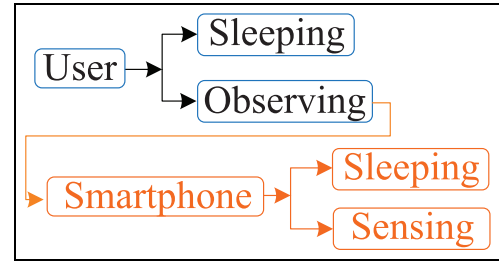

Figure 4. Composite detection.

System model

This article studies the scenario where the nodes and objects are static and uniformly randomly deployed in the interested area. With an additional server, these nodes and objects compose the *composite sensing system*. In each time slot, each object o_j is in either of two states: disappear and appear. The object state is clarified by the following two concepts: object state and system state.

Definition 1

Object state. The object state indicates the appearance of an object o_j in each time slot τ, and is denoted by $z_j(\tau)$, where $z_j(\tau) \in \{0(disappear), 1(appear)\}$.

The design of the optimal observing and sensing decision uses the definition of the object state. When an object is in the state: disappear, that is, $z_j(\tau) = 0$, it cannot be observed by any node. When the object is in the state: appear, that is, $z_j(\tau) = 1$, it can be observed

Table 1. Symbol and meaning.

Symbol	Description	Symbol	Description
v	Compound node	ψ	Probability of reward
o	Object	θ	Object's aspect
m	Number of objects	s	System state
B	Belief vector	μ	Belief state
Π	State space	z	Object state
τ	Time slot	W	New belief vector
$\flat$	Probability	P	Big probability
α	User decision	β	Smartphone decision
A	α set	Ψ	Vector of ψ
F	Value function	c	Composite decision
Q	Q function	u	Observing, report result
g	Node group	Φ	Object state vector
E	Expectation	η	Node sensing direction
K	Number of aspects	γ	Success report probability
q	Sensing quality	ζ	Threshold for γ
r	Reward	ω	Element of the vector
T	Period	φ	Element of Φ

and one of its K aspects faces one node. Assume that each object has the equal transition probability among the disappear state and the K aspects, that is, $p(\theta'|\theta) = p(\theta|\theta')$, and $p(z = 0|\theta) = p(\theta|z = 0)$, and its state transition is independent of other objects. Suppose that there are m objects around the node. The definition of the system state is given as below.
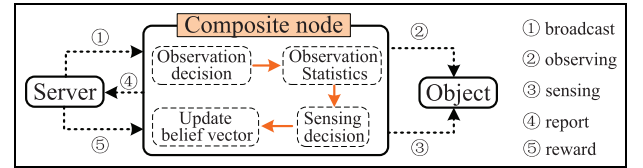
Definition 2

System state. The system state is the collection of the states of the m objects, and is denoted by $s(\tau)$, where $s(\tau) = \{z_j(\tau), j = 1, \ldots, m\}$.

Given a sequence of time slots $\tau \in T$, this article assumes that the system states $s(\tau)$ form a Markov chain with the state space $\Pi = \{0, 1\}^m$. To achieve reward, each node observes and senses the objects around it, and then reports the sensing results to the server. Let $\gamma_j^i(\theta = k)$ denotes the report of the node v_i for the object o_j when o_j's kth aspect faces the node v_i. The sensing quality of the report $\gamma_j^i(\theta = k)$ is thus $q_j^i(\theta = k)$. If the report is accepted by the server, it returns the acknowledgment of the node with a certain reward. In this article, the server adopts the non-separable sensing quality rule in equation (2) as the rule to choose the reporting from the nodes. By the function, the server accepts the maximal sensing quality for the same object among the nodes' reporting for the same object

$$q_j = \max_{v_i} q_j^i(\theta = k) \qquad (2)$$

where q_j^i is the sensing quality reported by the node v_i for the object o_j, and there may be more than one node sensing the same object o_j simultaneously. By the sensing quality rule in equation (2), the report $\gamma_j^i(\theta_j = k)$

Figure 5. The composite sensing system.

can be successful if any other report $\gamma_j^{i'}(\theta_j = k')$ for the same object o_j has no aspect with higher quality, that is, $k' \leqslant k$. Let $\gamma_j(\theta_j = k) \in \{0, 1\}$ denotes the *reported state* of the object o_j, which means that there is no report with the aspect higher than k if $\gamma_j(\theta = k) = 1$. Otherwise, $\gamma_j(\theta = k) = 0$. Most of symbols and their meaning are summarized in Table 1.

Composite sensing problem

This section presents the composite sensing process with the goal to maximize the overall sensing quality, and then maps it as a POMDP.

Compound sensing system

The structure of the composite sensing system, illustrated in Figure 5, implements the crowdsourcing task, which is implemented including four parts: *task broadcast, composite sensing process, report*, and *reward*.

Task broadcast. The application server broadcasts some advertisements to the users and to attract them to

participate in the task: to sense the objects in their sensing ranges. After the node accepts the task, it implements the composite sensing process to maximize the reward returned from the server.

Composite sensing process. Each node implements the composite sensing process, which is composed of conditional decisions made in a series of time slots. In each time slot, the observing decision $\alpha(\tau)$ is first made according to the historical observation and decisions, stored in the historical information vector $H(\tau)$. Based on its outcome, the sensing decision $\beta(\tau)$ is then made.

   Observing decision. At the beginning of each time slot τ, the node makes the observing decision. If the observing decision is made to be sleeping, the smartphone has to choose the *sleeping* sensing decision either in this slot. Otherwise, the user chooses one direction, that is, one object o_j, to observe. If the object's state is appearance, that is, $z_j(\tau) = 1$, the node can observe its orientation as shown in Figure 2(b). After the observation, the node obtains the observation outcome: the object state $z_j(\tau)$ and its orientation $\theta_j(\tau)$. Given the system state $s(\tau) = s$ and the observing decision $\alpha = 1$, the conditional PMF (probability mass function) of observation outcome, $\theta_j(\tau) = k$, for the object o_j is given by

$$
\begin{aligned}
p_o(k|s) &\stackrel{\Delta}{=} p\{\theta_j(\tau) = k|s(\tau) = s\} \\
&= \begin{cases} p(\theta_j(\tau) = k|z_j = 1), & \text{if } z_j = 1, \text{then } k > 0 \\ 0, & \text{if } z_j = 0, \text{then } k = 0 \end{cases}
\end{aligned} \quad (3)
$$

where $k = 0$ indicates that no aspect cannot be observed when the object o_j disappears, and $p(\theta_j(\tau) = k|z_j = 1)$ is the conditional probability that the observing outcome is $\theta_j(\tau) = k$ when the object state is $z_j = 1$.

   Sensing decision. After the observing decision for the object o_j, the node makes the sensing decision $\beta_j(\tau)$ in the slot. If the observing decision $\alpha(\tau) = 0$, the sensing decision must be sleeping, that is, $\beta_j(\tau) = 0$. Otherwise, the smartphone makes the sensing decision according to the observation outcome: $\theta_j(\tau) = k$. The node makes the sensing decision to sense the object o_j, that is, $\beta_j(\tau) = 1$, with the following probability

$$
p_s(\beta_j(\tau) = 1) = \begin{cases} p(\beta_j(\tau) = 1|\theta_j(\tau) = k), & \text{if } k > 0 \\ 0, & \text{if } k = 0 \end{cases} \quad (4)
$$

where the conditional probability $p(\beta_j(\tau) = 1|\theta_j(\tau) = k) \in [0, 1]$.

Report. After the sensing decision is made to achieve the sensing result $q(\theta_j(\tau))$, the result is reported to the server. The server chooses the report with maximal sensing quality for the same object by the rule given in equation (2), and the server feedbacks the reward to the reporting node. In this case, the node's report is called a successful report. Denoted the successful report for the object o_j by $\psi_j(k)$ when the observation outcome is $\theta_j(\tau) = k$ and $k > 0$. The node with the successful report can thus obtain some reward from the server, and counts its successful report probability, denoted by $p_r(\gamma_j(k))$. Recall that the successful report can be obtained only after the observing decision, sensing decision, and report are taken. So the successful report probability $p_r(\gamma_j(k))$ can be formulated as the following equation

$$
\begin{aligned}
\psi_j(k) &= p(\gamma_j(k) = 1|\beta_j(\tau) = 1) \\
&\quad p(\theta_j(\tau) = k|z_j = 1)p(z_j = 1|s(z_j = 1|s(\tau) = s) = s) \\
&= p_s(\beta_j(\tau) = 1)p_o(k|s)p(z_j = 1|s(\tau) = s)
\end{aligned} \quad (5)
$$

where the last equality is obtained by equations (3) and (4)

$$
p(\gamma_j(k) = 0|\beta_j(\tau) = 1) = 1 - p_r(\psi_j(k)) \quad (6)
$$

Reward. The reward, denoted by $r(\tau)$, for the successful report is defined to be a monotonically increasing function with the aspect. This article uses the sensing quality as the reward, which means that the successful report with higher sensing quality obtains higher reward according to equation (1). Recall the definition of the composite sensing process in section "Conditional sensing," the reward can be obtained only after the observing decision $\alpha(\tau) = 1$ and the sensing decision $\beta_j(\tau) = 1$. Then, the immediate reward $r(\tau)$ in slot τ can be given by

$$
r(\tau) = \alpha(\tau)\beta_j(\tau)q(\theta = k), \quad k = 1, \ldots, K \quad (7)
$$

   Notice that the node chooses only one object to sense each time if its sensing decision $\beta > 0$. It is willing to choose the object that can result in the sensing quality and probability of the successful reporting as high as possible. The objects have their own states: appear or disappear, which compose of the state space Π. They switch between the states from one time slot τ to next time slot $\tau + 1$ with some probabilities p_Π.

Convert to POMDP

The composite sensing process can be mapped as the POMDP. In the process, the node observes only a part of the objects around it, and the report result cannot be

directly known after it reports the sensing result to the server. The system states thus cannot be fully observable. In the following, this article formulates the composite sensing process as the POMDP by a tuple $\langle \Pi, \Theta, A, P, q \rangle$:

- Π is a set of objects' states in the node's sensing range.
- Θ is a finite set of sensing and report results, that is, $\theta, \gamma \in \Theta$.
- A is the decision space, that is, $A = \{\alpha, \beta\}$, $\forall \alpha \in \{0, 1\}, \beta = \{0, 1, \ldots, K\}$.
- P is a set of the system state transition probabilities: $P = \{p(s'|s)\}, \forall s, s' \in \Pi$.
- $q(\theta)$: $A \times \Pi \to (0, 1]$ is the sensing quality function.

Belief vector. In the composite sensing process, the node makes the decision according to the historical information $H(\tau)$ at the beginning of each time slot. The historical information vector $H(\tau)$ is updated in each time slot τ. As time goes on, the size of $H(\tau)$ grows quite big. Smallwood et al.[23] showed that the conditional probability, denoted by $B(\tau)$, of the system states of the objects around the node based on its decision and observation history $H(\tau)$ can be a sufficient statistic of these objects' historical states. $B(\tau)$ is named as the *belief vector* of the node for the states of the objects around it at the end of each time slot $\tau - 1$, and is defined as $B(\tau) \stackrel{\Delta}{=} [\mu_s(\tau)]_{s \in \Pi}$. Each element $\mu_s \in B(\tau)$, called belief state, is the conditional probability (given the observing and sensing history) that the objects' state is s at the beginning of slot $\tau + 1$ prior to the state transition. $B(\tau)$ can be updated based on $B(\tau - 1)$ and the decisions and report results in the slot τ. We introduce an *updating function* $\mathcal{T}$ to implement the updating of the belief vector, that is, $B(\tau) = \mathcal{T}(B(\tau - 1)|\Theta(\tau), A(\tau))$.

This article adopts a reward-based updating function $\mathcal{T}$: $B(\tau + 1) = \mathcal{T}(B(\tau), \Theta(\tau), A(\tau), \Psi(\tau))$. Based on the Bayes' rule, the update of $B(\tau + 1)$ is calculated in two cases. When the observing decision makes the node to sleep, that is, $\alpha = 0$, the belief vector is updated based solely on the underlying Markovian model of the object state, that is, $B(\tau + 1) = \mathcal{T}(B(\tau)|\alpha = 0)$. The belief element is updated by the following equation

$$\mu_s(\tau + 1) = \sum_{s' \in \Pi} \mu_{s'}(\tau) p(s|s') \qquad (8)$$

When the user takes the observing decision $\alpha(\tau) = 1$, it can observe the system state $s(\tau) = z(\tau)$ with the probability as equation (3). The information state can be updated by the Bayes' rule:[24] when the node is in the state s' at slot τ, the belief state is the probability that the state is in the state s at slot $\tau + 1$

$$\mu_s(\tau + 1) = \frac{\sum\limits_{s' \in \Pi} \mu_{s'}(\tau) p(s|s') p_o(k|s)}{P(\gamma|s, (\alpha, \beta))} \qquad (9)$$

where the denominator is a normalizing constant and is given by the sum of the numerator overall values of $s \in \Pi$ as the following equation

$$\mu_s(\tau + 1) = \frac{\sum\limits_{s' \in \Pi} \mu_{s'}(\tau) p(s|s') p_s(k|s)}{\sum\limits_{s' \in \Pi} \mu_{s''}(\tau) p_s(k|s'')} \qquad (10)$$

where $p_s(k|s)$ is given according to equation (4).

Objective. The composite sensing policy is a sequence of decision couples: $\langle \alpha(\tau), \beta(\tau) \rangle$, $\tau \in T$. The optimal policy, denoted by $\langle \alpha(\tau), \beta(\tau) \rangle^*$, $\tau \in T$, is to maximize the expected overall sensing quality in T under the constraint of the successful reporting probability threshold ζ. It is equivalent to finding the optimal policy for the finite constrained POMDP. Recalling the immediate reward given in equation (7), the goal of the optimal policy is given by

$$q(T) = \max E\left[\sum_{\tau \in T} r(\tau) \Big| B(0)\right] \qquad (11)$$

$$\text{s.t.} \quad p_r(\gamma_j(k)) \geqslant \zeta \quad \forall v_j \in V, k = 1, \ldots, K \qquad (12)$$

where $B(0)$ is the initial belief vector for the object states, and ζ is the threshold for the success report probability.

Composite sensing policy

Some previous works, such as the one-pass algorithm,[23] can carry out the sequence of the optimal decision. The computation complexity required to obtain the optimal decision increases exponentially with the size of the state space, and can be very high for the general POMDP.[25] One of the alternative methods for addressing this problem is to design the myopic policy.[25] Myopic policy focuses on the immediate reward and ignores the impact of current policy on future rewards. Generally, the myopic policy is suboptimal. In this section, we explore some specific properties of the composite sensing system: monotonicity and the independence between the action and object states. With these properties, the computation for the optimal policy given in this section can be simplified.

Value function

The key step of making the composite decision is to measure how good the previous decision is. Value function can express the objective in equation (11) explicitly

as functions of the belief vector B and the observing and sensing decision $\langle\alpha,\beta\rangle$. Let $F(B(\tau),A)$ denotes the *value function*, which is the maximum expected total reward that can be accumulated starting from τ given the belief state $B(\tau)$. To make the decision $\langle\alpha,\beta\rangle$ in each time slot τ can accumulate the reward started from τ with two parts: the immediate reward given in equation (7) and the maximum expected future reward $F(B(\tau+1),A)$. Considering all possible system states $s\in\Pi$ and the successful report probability in equation (5), and then maximize over all possible decisions in A, we can arrive the value function in the following equation

$$F_T(B) = \max_{\langle\alpha,\beta\rangle\in A}\sum_{s\in\Pi}\mu_s(\tau)r(B,\langle\alpha,\beta\rangle)$$

$$F_\tau^*(B) = \max_{\langle\alpha,\beta\rangle\in A}\sum_{s\in\Pi}\mu_s(\tau)\sum_{\gamma_j(k)=0}^{1}p_r(\gamma_j(k))$$

$$\left[\gamma_j(k)r(B,\langle\alpha,\beta\rangle) + F_{\tau+1}^*(\tau(B,\langle\alpha,\beta\rangle))\right], \quad d\forall\tau\in T \quad (13)$$

where the first term in the right of the equation denotes the expected immediate reward $r(B,A)$, and the future reward $F_\tau(B(\tau+1))$ can be calculated by the future belief vector $B(\tau+1)$ with the Bayes' rule.[26,27] The immediate reward $r(B,A)$ is achieved in current time slot by taking the sensing action, and is given as $r(B,A)=\psi(\gamma|s)q_j$.

Optimal composite sensing policy

This section analyzes the properties of the composite sensing process, which includes: (1) monotonicity of value function and (2) monotonicity of success report probability. With these properties, we can obtain an explicit optimal design for the composite sensing process and a deterministic optimal sensing policy in Lemma 2, and observing policy in Lemma 3.

Lemma 1

Monotonicity of success report probability. Given the sensing decision $\beta=1$, the success report probability $p_r(\gamma(k))$ increases with the observing outcome $\theta(\tau)=k$, that is, $p_r(\psi_j(k'))\geqslant p_r(\psi_j(k))$ for $k'\geqslant k$.

The proof of Lemma 1 is referred to Appendix 1.

Theorem 1

Monotonicity of value function. The value function $F(B,\theta)$ is monotonically increasing with the aspect θ, that is, $F(B,\theta')\geqslant F(B,\theta)$ for $\theta'\geqslant\theta$. The proof of Theorem 1 is referred to Appendix 1.

Recall that the object of the composite sensing process is to maximize the overall reward under the constraint of the successful sensing probability as given in

equation (11). If there is no constraint, the node would always make the composite sensing to wake up in each time slot so as to maximize the overall outcome. With the constraint given in equation (12), the composite sensing must be decided carefully. Since the successful report possibility increases monotonically with the aspect θ as claimed in Lemma 1, there must be an aspect, denoted by $\theta(\tau)=\bar{k}$, such that the following condition is satisfied given the observing outcome $\theta(\tau)>0$

$$\exists\theta=\bar{k}:p(\psi_j(\bar{k}))\geqslant\zeta \quad \text{and} \quad p(\psi_j(\bar{k}-1))<\zeta \quad (14)$$

According to equation (6), the successful sensing probability is affected by both the observing and sensing decisions. By Lemma 1, the sensing decision $\beta=1$ with higher the observing outcome $\theta(\tau)=k$ can result in higher success report probability $p(\psi_j(k'))$. According to Theorem 1, the value function monotonically increases with the success report probability $p(\psi_j(k'))$. Therefore, we can make a threshold-structured optimal sensing decision, which is given by the below lemma.

Lemma 2

Optimal sensing decision. Given the observing outcome $\theta(\tau)=k$, the optimal sensing decision β is given as follows

$$\beta(\tau) = \begin{cases} 1, & \text{if } k\geqslant\bar{k} \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

where the threshold aspect $\bar{k}$ is defined in equation (14).

The next is to design the optimal observing decision, which chooses the best object to observe in each time slot since there are m objects. It is easy to find that there is definitively no chance to obtain the reward if the object is in the state of disappearance, that is, $z=0$. Lemma 2 shows the optimal sensing decision, that is, the sensing decision must be taken only if the observing outcome is $\theta(k)=1$, $k\geqslant\bar{k}$ in order to satisfy the constraint in equation (12). For the constraint composite sensing process, the observing decision has to choose the object, whose state is $z=1$ and the aspect $\theta(k)=1$, $k\geqslant\bar{k}$. The threshold of the aspect $\theta(\bar{k})=1$ divides the object states into two groups denoted by $\tilde{z}=1$ and $\tilde{z}=0$. In the first group $\tilde{z}=0$, the object states includes $z=0$ or $z=1$ and the aspect $\theta(k)=1$, $k<\bar{k}$. In the second group $\tilde{z}=1$, the object states includes $z=1$ and the aspect $\theta(k)=1$, $k\geqslant\bar{k}$. For each object o_j, we also define two transition probabilities: σ_j and ϕ_j, between the two *group states*, as follows

$$\sigma_j(\tau) \triangleq p(\tilde{z}(\tau)=1|\tilde{z}(\tau-1)=1)$$
$$\phi_j(\tau) \triangleq p(\tilde{z}(\tau)=1|\tilde{z}(\tau-1)=0) \quad (16)$$

The above two probabilities can be calculated and updated from the transition probability of the system states given in equation (9) or (10)

$$\sigma_j(\tau) = \sum_{k=\bar{k}}^{K} \sum_{k'=\bar{k}}^{K} p(z = \theta(k), s(\tau+1)|z = \theta(k'), s(\tau))$$

$$\phi_j(\tau) = \sum_{k=\bar{k}}^{K} \sum_{k'=0}^{\bar{k}-1} p(z = \theta(k), s(\tau+1)|z = \theta(k'), s(\tau))$$

$$(17)$$

Because one object's states are independent of others', the probability that the object o_j is in the group state $\widetilde{z}(\tau+1) = 1$ can be updated according to the observing outcome in previous time slot by the following equation

$$p(\widetilde{z}(\tau+1) = 1)$$

$$= \begin{cases} \sigma_j, & \text{if } \alpha_j = 1; k \geq \bar{k} \\ \phi_j, & \text{if } \alpha_j = 1; k < \bar{k} \\ \sigma_j(\tau) + (\phi_j - \sigma_j)p(\widetilde{z}(\tau) = 1) & \text{if } \alpha_j = 0 \end{cases}$$

$$(18)$$

We have the following lemma to determine the optimal observing decision.

Lemma 3

Optimal observing decision. Suppose that there are m objects. Given the observing outcome in the previous slot $\tau - 1$, the optimal observing decision is to observe the object with the $[\min\{\sigma, \phi\}, \max\{\sigma, \phi\}]$. The optimal observing decision in time slot τ is to choose the object o_j to observe, where $o_j = \arg_{o_j} \max \psi_j(k), \quad \forall k = 1, \ldots, K$.

Proof. According to the definition of the group state $\widetilde{z}(\tau+1)$, the observing decision lets the node active when the object state is $z(\tau) = 1$ and $\theta(\tau) = k, \quad k \geq \bar{k}$. Thus, the constraint is satisfied by the observing decision. Thus, the object state, which results in the maximal value function, must be contained in the group state.

Next, we prove by induction that the value of the observing decision given in Lemma 3 is maximized. According to the system model in section "System model," the object states have the equal transition probability among its states. The transition probability does not change with time. When the observing decision makes the node to sleep, that is, $\alpha(\tau) = 0$, there is no chance to outcome any observing result by equation (3), that is, $p_o(k|s) = 0$. Thus, $p = 0$ given in equation (6). So $F(p, s) = 0) = 0$ in this case. When the observing decision makes the node to observe the object, that is, $\alpha(\tau) = 1$, the belief vector can be updated by equation (9). Thus, we have $B(\tau+1) = \mathcal{T}(B(\tau)|\alpha = 1)$. Since the observing decision in Lemma 3, the probability $p(z_j = 1|s(\tau) = s)$ in equation (6) for the object state in the group state $\widetilde{z} = 1$ can be maximized. With $F(p, s) = 1$. For each object, the transition probabilities between any two states are equal, that is, $p(z_i|z_j) = p(z_j|z_i)$. Therefore, the observing decision given in Lemma 3 is optimal.

Optimality of myopic policy

A myopic policy does not consider the impact of the current action on the future or long-term reward, and focuses solely on maximizing the expected immediate reward. It is usually suboptimal for the general POMDP. The myopic policy need not estimate the future reward so that the computation complexity can be reduced. In this article, the myopic policy only cares the impact on the next time slot so we modify the value function as the following equation

$$F_\tau^*(B) = \max_{\langle \alpha, \beta \rangle \in A} \sum_{s \in \Pi} \mu_s(\tau) \sum_{\gamma_j(k)=0}^{1} p_r\big(\gamma_j(k)\big)\gamma_j(k)r(B, \langle \alpha, \beta \rangle)$$

$$(19)$$

The description of the myopic policy is quite similar to the optimal one except that equation (13) in step 5 of Algorithm 1 is replaced by equation (19).

Algorithm I. Optimal policy.

Input: Initial belief vector $B(0)$ and ζ.
Output: Overall quality $q(T)$.
1: List all possible information states $(B, p_r(\gamma_j(k)))$, $v_j \in V, k = 1, \cdots, K$, that each node may go through. Let $\mathcal{B}$ include all such states such that the constraint in inequality (12).
2: Let $= 0$ for all states $(B, p_r(\gamma_j(k)))$ with $p_r(\gamma_j(k)) > \zeta$, $v_j \in V, k = 1, \cdots, K$;
3: **while** $\tau < = T$ **do**
4:   **if** $\mathcal{B}$ is nonempty **then**
5:     Compute the value function for the state $(B, p_r(\gamma_j(k))) \in \mathcal{B}$ with equations (9) and (13);
6:     Get the maximal quality of all object and remove its state from set $\mathcal{B}$;
7:   **end if**
8:   $\tau = \tau + 1$
9: **end while**

Experiment results

In this section, we conduct numerical and simulation to verify the performance of our optimal and myopic policy by comparing it with a randomized algorithm, which is just to select some objects in each round randomly. We numerically analyze the impact of various parameters such as the average quality, sensing ratio,
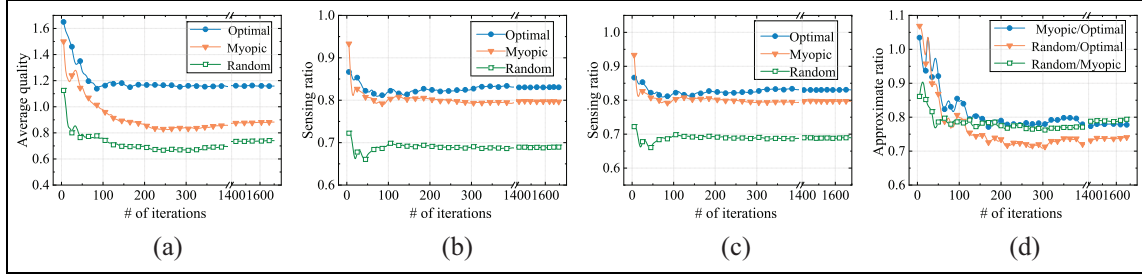
Figure 6. Convergence of the optimal, myopic, and random policies: (a) average quality, (b) sensing ratio, (c) success report ratio, and (d) approximate ratio.

success report ratio, and the algorithm approximate ratio under proposed algorithms in terms of the number of iterations and different thresholds. Besides, we give the progress proportion and delay analysis of the optimal policy.

Evaluation setup

To better validate the performance of our proposed algorithms, we build a test bed and conduct field experiments. Our evaluation field is divided into three disks according to composite node v_1, v_2, v_3 with its observing range. Seven objects are uniformly and randomly deployed in the field. The possible states of the seven objects in each time slot are: appear or disappear. The state in different time slots has no effect on each other. If an object appears, the orientation is also randomly distributed, and the orientation in different time slots is also independent of each other. In the following Figures 6 and 7, we consider the average quality, sensing ratio, success report ratio, and the algorithm approximate ratio as metrics for evaluation under various parameters: the number of iterations and thresholds.

Performance comparison

Average quality. Figure 6(a) shows the average quality obtained by the optimal, myopic, and randomized policies, respectively, under the different number of iterations and fixed threshold value $\zeta = 0.1$. After almost 200 iterations, the optimal policy gets a stable average quality, about 1.19. Besides this, the average quality achieved by the myopic policy is about 0.88 after nearly 500 iterations. In contrast, the average quality of the random policy is about 0.73 after 500 iterations, which is much lower than other policies as shown in Figure 6(a).

As shown in Figure 7(a), we evaluate the average qualities obtained by the optimal and myopic policies compared with the random policy when we set various thresholds and keep the fixed 1500 iteration times. With

the threshold increasing, the optimal strategy always maintains a good expectation value of the average quality about 1.2. In contrast, myopic and random policies show insufficient performance. When the threshold is between $[0, 0.3]$, the myopic policy gets the average quality about 0.92 and the random policy gets it about 0.79. When the threshold is greater than 0.3, their average qualities drop badly.

Sensing ratio. As mentioned in equation (12), when the success report probability is less than the threshold ζ, we will not take the sensing action in the optimal and myopic policies. Figure 6(b) counts the ratio of the number of sensing actions to the number of observing actions with the number of iterations increasing from 0 to 1700. It reflects the sensing probability obtained by the optimal, myopic, and randomized policies after observing objects. Again, we set the threshold with the fixed value $\zeta = 0.1$. After almost 300 iterations, the optimal policy gets a stable sensing ratio, about 84%. Besides this, the sensing ratio obtained with the myopic policy is about 80% after nearly $20k$ iterations. In contrast, the sensing ratio of the random policy is about 68% after 300 iterations, which is much lower than other policies as shown in Figure 6(a).

As shown in Figure 7(b), we evaluate the sensing ratio obtained by the optimal and myopic policies compared to the random policy when we set various threshold values and keep the fixed number of iterations, that is, 1500. The optimal policy always maintains a good sensing ratio with about 80%. In contrast, the myopic policy and random policy show insufficient performance. When the threshold is between $[0, 0.4]$, the sensing ratio gets by the myopic policy is about 65% and the sensing ratio gets by the random policy is about 54%. When the threshold is greater than 0.4, their sensing ratio performance drops badly.

Success report ratio. As mentioned in equation (2), the server only accepts the maximal sensing quality for a same object among the nodes' reporting for it.
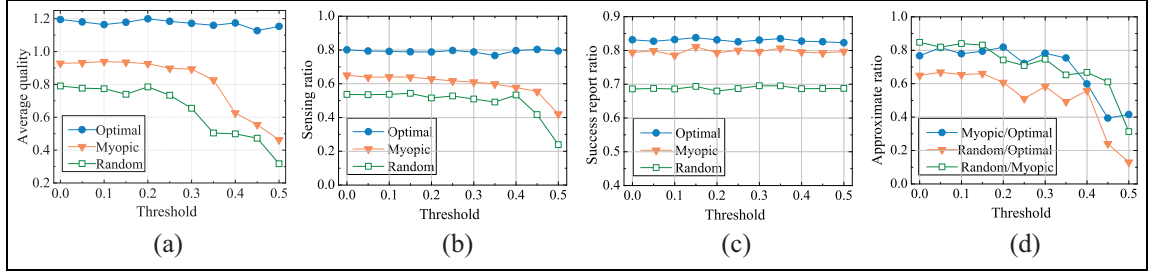
Figure 7. Performance under different thresholds with fixed 1500 iterations: (a) average quality, (b) sensing ratio, (c) success report ratio, and (d) approximate ratio.

Therefore, the success report ratio is also one of the criteria to evaluate how good a strategy is. Figure 6(c) counts the ratio of the number of success reports to that of observing actions with the number of iterations increasing from 0 to 1700. We set the threshold value to be 0.1. It reflects the success report probability obtained by the optimal, myopic, and randomized policies after observing objects. After almost 300 iterations, the optimal policy gets a stable success report ratio about 82%. Besides this, the success report ratio by the myopic policy is about 79% after nearly 400 iterations. In contrast, the success report ratio of the random policy is about 67% after 300 iterations, which is obviously lower than other policies as shown in Figure 6(c).

As shown in Figure 7(c), we evaluate the success report ratio obtained by the optimal and myopic policies compared to the random policy when we set various threshold values and keep the fixed 1500 iterations. The optimal strategy always maintains a good success report ratio about 84% and the myopic policy shows insufficient performance with 79% success report ratio. However, the success report ratio of the random policy is only about 68%.

Approximate ratio. The approximation ratio can measure the performance difference of our policies. It reflects the performance of the optimal, myopic, and randomized policies clearly. Again, we set the threshold value of ζ to be 0.1. In Figure 6(d), the blue curve shows the approximate ratio between the myopic and optimal policies with the number of iterations increasing from 0 to 1700. It is obvious that the performance of the optimal and myopic policies goes stable after nearly 200 iterations. The approximation ratio of the myopic and optimal policies is about 78% finally. The orange curve shows the approximation ratio between the random and optimal policies with the number of iterations increasing from 0 to 1700. It is obvious that the performance of the random and optimal policies gets stable after nearly 200 iterations. The approximation ratios of the myopic and optimal policies are about 73%. The green curve shows the approximation ratio between the

random and myopic policies with the number of iterations increasing from 0 to 1700. After nearly 150 iterations, the performance of the random and myopic policies gets stable. The final approximation ratio of the myopic and optimal policies is about 77%.

As shown in Figure 7(d), we evaluate the approximation ratio among the optimal, myopic, and randomized policies when we set various thresholds and keep the fixed number of iterations, that is, 1500. In Figure 7(d), the blue curve shows the approximate ratio between the myopic and optimal policies with the threshold varies between $[0, 0.5]$. The approximation ratios of the myopic and optimal policies are about 78% and relatively stable in the interval $[0, 0.35]$. When the threshold is greater than 0.35, the approximation ratio suddenly drops to around 40%. The orange curve shows the approximation ratio between the random and optimal policies with the threshold varies between $[0, 0.5]$. The approximation ratio of the random and optimal policies is about 60% and relatively stable in the interval $[0, 0.4]$. When the threshold is greater than 0.4, the performance suddenly drops to around 20%. The green curve shows the approximation ratio between the random and myopic policies with the threshold varies between $[0, 0.5]$. The approximation ratios of the random and myopic policies are about 70% and relatively stable in the interval $[0, 0.45]$. When the threshold is greater than 0.45, the performance suddenly drops to around 30%.

Delay and progress proportion. To review the complex perceptual system in Figure 5, the server goes through five steps from the start of broadcasting to feedback rewards to the object. In this experiment, we use delay to represent the time from the beginning of the broadcast to the end of the feedback. As shown in Figure 8, we observe that the delay of the optimal policy increases significantly as the number of objects increases. In addition, after several hundred iterations, the delay of the optimal policy is basically stable. In this experiment, it is assumed that we need the optimal strategy to complete the calculation of 1500 iterations.
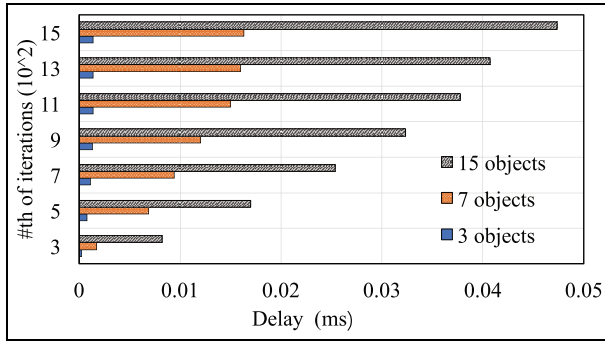
Figure 8. Delay of the optimal policy with 1500 iterations.

Figure 9. Progress proportion of the optimal policy.

The progress proportion represents the percentage of the number of completed iterations to the total 1500 iterations under a particular timestamp. As shown in Figure 9, we observe that with the increase in the number of objects, the time to complete the fixed 1500 iterations of the optimal policy is significantly extended.

The main trends in the results are summarized as follows:

- The average quality, sensing ratio, success report ratio, and other indicators obtained by the optimal policy and myopic policy tend to be stable.
- Compared with the myopic policy, some indicators of the optimal policy reached the stability earlier.
- The effect of threshold setting on myopic policy and random policy is much greater than that of optimal policy.
- With the increase in objects' number in the experimental scene, the delay increases significantly and the progress proportion slows down significantly.

Conclusion

This article observed the phenomenon of composite sensing with user as sensor in crowdsourcing. The phenomenon usually happens and has not been well studied. We thus proposed the framework: composite sensing, and then map it as a POMDP problem. The composite sensing policy is proposed and analyzed theoretically and experimentally. The theoretical optimization of the policy is guaranteed. In this article, we discuss the case where the smartphone can choose one direction to sense in each time slot. We take another case as a future work, where the smartphone may choose one or more directions to sense in each time slot. Compared with traditional methods, the use of this method in large-scale environmental data has yet to be verified and optimized.
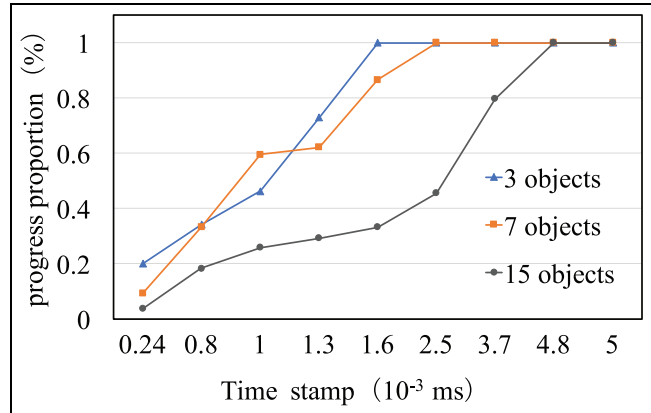
ORCID iD

Siwen Zheng (iD) https://orcid.org/0000-0001-9224-2119

References

1. Chatzimilioudis G, Konstantinidis A, Laoudias C, et al. Crowdsourcing with smartphones. *IEEE Internet Comput* 2012; 16(5): 36–44.
2. Lane ND, Miluzzo E, Lu H, et al. A survey of mobile phone sensing. *IEEE Commun Mag* 2010; 48(9): 140–150.
3. Ganti RK, Ye F and Lei H. Mobile crowdsensing: current state and future challenges. *IEEE Commun Mag* 2011; 49(11): 32–39.
4. Zhang J, Lu P, Li Z, et al. Distributed trip selection game for public bike system with crowdsourcing. In: *Proceedings of the IEEE conference on computer communications (INFOCOM)*, Honolulu, HI, 16–19 April 2018, pp.2717–2725. New York: IEEE.
5. Zhang H, Zhang Z, Chen X, et al. Energy efficient joint source and channel sensing in cognitive radio sensor networks. In: *Proceedings of the 2011 IEEE international conference on communications (ICC)*, Kyoto, Japan, 5–9 June 2011, pp.1–6. New York: IEEE.

6. Li Z, Zhang J, Gan J, et al. Large-scale trip planning for bike-sharing systems. *Pervas Mob Comput* 2019; 54: 16–28.

7. Wu H, Zhang Z and Jiao C. Learn to sense: a meta-learning based distributed sensing algorithm in wireless sensor networks. In: *Proceedings of the 2018 10th international conference on wireless communications and signal processing (WCSP)*, Hangzhou, China, 18–20 October 2018, pp.1–6. New York: IEEE.

8. Bhatta A and Mishra AK. GSM based hand-held comm-sense-sensor for environment monitoring. In: *Proceedings of the 2016 11th international conference on industrial and information systems (ICIIS)*, Roorkee, India, 3–4 December 2016. New York: IEEE.

9. Yang Z, He S, Chen J, et al. Localize online social network user via social sensing. In: *Proceedings of the 2nd international workshop on social sensing*, Pittsburgh, PA, 18–21 April 2017. New York: ACM.

10. Khan WZ, Xiang Y, Aalsalem M, et al. Mobile phone sensing systems: a survey. *IEEE Commun Surv Tut* 2013; 15(1): 402–427.

11. Wang J, Zhang Z and Hanzo L. Joint active user detection and channel estimation in massive access systems exploiting Reed–Muller sequences. *IEEE J Select Top Signal Process* 2019; 13: 739–752.

12. Becken S, Stantic B, Chen J, et al. Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. *J Environ Manage* 2017; 203(1): 87–97.

13. Yang D, Xue G, Fang X, et al. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In: *Proceedings of the 18th annual international conference on mobile computing and networking (Mobi-Com 2012)*, Istanbul, 22–26 August 2012, pp.173–184. New York: ACM.

14. Rana RK, Chou CT, Kanhere SS, et al. Ear-phone: an end-to-end participatory urban noise mapping system. In: *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*, Stockholm, 12–16 April 2010, pp.105–116. New York: ACM.

15. Srivastava M, Abdelzaher T and Szymanski B. Human-centric sensing. *Philos Trans Roy Soc A Math Phys Eng Sci* 2012; 370(1958): 176–197.

16. Wang D, Amin MT, Li S, et al. Using humans as sensors: an estimation-theoretic perspective. In: *Proceedings of the 13th international symposium on information processing in sensor networks*, Berlin, 15–17 April 2014, pp.35–46. New York: IEEE.

17. Wang Y, Hu W, Wu Y, et al. SmartPhoto: a resource-aware crowdsourcing approach for image sensing with smartphones. In: *Proceedings of the 15th ACM international symposium on mobile ad hoc networking and computing*, Philadelphia, PA, 11–14 August 2014, pp.113–122. New York: ACM.

18. Mohan P, Padmanabhan VN and Ramjee R. Traffic-Sense: rich monitoring of road and traffic conditions using mobile smartphones. In: *SenSys'08*, Raleigh, North Carolina, USA, 5–7 November 2008.

19. Thiagarajan A, Ravindranath L, LaCurts K, et al. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In: *Proceedings of the 7th ACM conference on embedded networked sensor systems*, Berkeley, CA, 4–6 November 2009, pp.85–98. New York: ACM.

20. Wang Y, Liu X, Wei H, et al. CrowdAtlas: self-updating maps for cloud and personal use. In: *Proceeding of the 11th annual international conference on mobile systems, applications, and services*, Taipei, Taiwan, 25–28 June 2013. New York: ACM.

21. Zhao D, Li XY and Ma H. How to crowdsource tasks truthfully without sacrificing utility: online incentive mechanisms with budget constraint. In: *Proceedings of the IEEE conference on computer communications (INFO-COM)*, Toronto, ON, Canada, 27 April–2 May 2014. New York: IEEE.

22. Sensorly, http://www.sensorly.com (accessed November 2014).

23. Smallwood RD and Sondik EJ. The optimal control of partially observable Markov processes over a finite horizon. *Oper Res* 1973; 21(5): 1071–1088.

24. Ross S, Pineau J, Chaib-draa B, et al. A Bayesian approach for learning and planning in partially observable Markov decision processes. *J Mach Learn Res* 2011; 12: 1729–1770.

25. Shani G, Pineau J and Kaplow R. A survey of point-based POMDP solvers. *Auton Agent Multi Agent Syst* 2013; 27(1): 1–51.

26. Maritz JS and Lwin T. *Empirical Bayes methods*. Abingdon: Routledge, 2018.

27. Butler T, Jakeman J and Wildey T. Combining push-forward measures and Bayes' rule to construct consistent solutions to stochastic inverse problems. *SIAM J Sci Comput* 2018; 40(2): A984–A1011.

Appendix 1

Proof for Lemma 1

Lemma 4

Given the sensing decision $\beta = 1$, the success report probability $p_r(\gamma(k))$ increases with the aspect $\theta = k$, that is, $p(\gamma_j(k') = 1|\beta_j(\tau) = 1) \geqslant p(\gamma_j(k) = 1|\beta_j(\tau) = 1)$ for $k' \geqslant k$.

Proof

Recall that the nodes and objects are uniformly and randomly deployed in the interested area in section "Preliminaries." Suppose there are n nodes, which can sense the same object o_j. The o_j's direction is random and fixed in an arbitrary time slot. In each aspect, there are n/K nodes on average. There may be more than one node in each aspect. The nodes in the same aspect of the object o_j have the same sensing result and same report result. Denote the node in the kth aspect by v_k, $k = 1, \ldots, K$, and the reporting probability of the node v_k by p_k. The node in the lower aspect of the object can sense the object with higher quality according to the object model in section "Object, observing,

and sensing model." When the node v_k can obtain the successful report, the server does not accept the reports of the other nodes $v_i, i < k$. So the probability that v_k's report is successful is $p(\gamma^j|k) = \prod_{i=1}^{k-1}(1 - p_i)p_k$. It is easy to notice that $p(\gamma^j|k)$ is non-increasing with k since $0 \leqslant p_i \leqslant 1$. Notice that the lemma also holds when there is less than one node on average in each aspect. According to equation (5), this lemma is correct.

For the states s_k and $s_{k'}$ with $k \geqslant k'$, there is at least one object under the state s_k. Its aspect facing the node is lower than the lowest aspect of the objects under the state $s_{k'}$ according to Definition 2. By the composite sensing policy, the node senses only one object each time. It refers to sense the object with the lowest aspect in s_k rather than that under the state s_k, and then have higher successful reporting according to the above proof.

Lemma 5

Monotonicity of success report probability. Given the sensing decision $\beta = 1$, the success report probability $p_r(\gamma(k))$ increases with the aspect $\theta = k$, that is, $p_r(\psi_j(k')) \geqslant p_r(\psi_j(k))$ for $k' \geqslant k$.

Proof

Recall the assumption in section "System model." When the object o_j appears, that is, $z_j = 1$, any aspect of the object may face the node with the same probability, that is, $p(\theta|\theta') = p(\theta'|\theta)$. For two aspects $\theta_j(\tau) = k'$ and $\theta_j(\tau) = k$, we thus have

$$p(\theta_j(\tau) = k'|z_j = 1)p(z_j = 1|s(\tau) = s)$$
$$= p(\theta_j(\tau) = k|z_j = 1)p(z_j = 1|s(\tau)) \quad (20)$$

According to the definition of the successful report probability in equation (6) and Lemma 1, we can prove the lemma.

Proof for Theorem 1

The value function in equation (13) contains two parts: the expected immediate reward and future reward. The two parts have the property of monotonicity with the aspect. We thus can prove the Theorem 1 by the following lemmas.

Lemma 6

Monotonicity of immediate reward. Given the sensing decision $\beta = 1$, the expected immediate reward $r(B, A)$ monotonically increases with the aspects θ.

Proof. In the value function in equation (13), the expected immediate reward is given the following equation

$$E[r(\tau)|B(\tau), A] = \sum_{s \in \Pi} \mu_s(\tau) \sum_{\gamma_{j,k}(\tau) = 0}^{1} \psi(\gamma_{j,k}(\tau))r(B, \langle \alpha, \beta \rangle) \quad (21)$$

Consider two composite decisions: $A = \langle \alpha, \beta \rangle$ and $A' = \langle \alpha', \beta' \rangle$ such that the observing outcomes under the two composite decisions are $\theta(k)$ and $\theta(k)$, $k \geqslant k'$, respectively. Notice that the reward is zero when $\gamma_{j,k}(\tau) = 0$. Comparing the expected immediate reward resulting from the two decisions, we have

$$E[r(\tau)|B(\tau), A] - E(r(\tau)|B(\tau), A')$$
$$= \sum_{s \in \Pi} \mu_s(\tau)\psi(\gamma_{j,k}(\tau) = 1)r(B, A)$$
$$\quad - \sum_{s \in \Pi} \mu_s(\tau)\psi(\gamma_{j,k'}(\tau) = 1)r(B, A')$$
$$= \sum_{s \in \Pi} \mu_s(\tau)\{\psi(\gamma_{j,k}(\tau) = 1)r(B, A) - \psi(\gamma_{j,k'}(\tau) = 1)$$
$$\quad r(B, A')\} \geqslant 0$$

$$\quad (22)$$

This finishes the proof.

Recall that the server chooses the report with the maximal sensing outcome by sensing quality rule in section "Related work." The fail report cannot obtain any immediate reward, that is, the accumulated value $F(B(\tau))$ cannot increase if the report is not successful in τ. We thus have the following claim.

Claim 1. Suppose that the accumulated value up to time slot $\tau + 1$ is $F(B(\tau + 1))$, and the accumulated value up to τ is same to that in $\tau + 1$ with probability of $p(\gamma_j(k) = 0|\beta_j(\tau) = 1)$ if the report $\gamma(k)$ is implemented, where $p(\gamma_j(k) = 0|\beta_j(\tau) = 1)$ is given in equation (6).

Theorem 2

Monotonicity of value function. The value function $F(B, \theta)$ is monotonically increasing with the success report probability p_r, that is, $F(B(\tau), p_{r'}) \geqslant F(B(\tau), p_r)$ for $p'_r \geqslant p_r$.

Proof. We prove the theorem by mathematical induction. From the value function given in equation (13), we can induce that the theorem is correct by Lemma 6. Suppose that the theorem is correct for every slot $\tau > \tau_0$, that is, for two success report probabilities $p_r \geqslant p'_r$, the following inequality is satisfied

$$F(B(\tau_0), p_r) \geqslant F(B(\tau_0), p'_r) \quad (23)$$

Since the failed report cannot obtain any reward according to Claim 1, equation (23) also indicates

$$F(B(\tau_0), 1 - p_r) \geqslant F(B(\tau_0), 1 - p'_r) \quad (24)$$

Comparing two value functions with the two success report probabilities p_r and p'_r, we have

$$F(B(\tau_0), p_r) - F(B(\tau_0), p'_r)$$

$$= \sum_{s \in \Pi} \mu_s(\tau) \{ p_r r(B(\tau_0), p_r) - p'_r r(B(\tau_0), p_{r'})$$

$$+ \sum_{\gamma_j(k) = 0}^{1} [p_r F_{\tau+1}(\mathcal{T}(B, p_r)) - p_{r'} F_{\tau+1}(\mathcal{T}(B, p'_r))] \}$$

$$\geqslant \sum_{s \in \Pi} \mu_s(\tau) \sum_{\gamma_j(k) = 0}^{1} [p_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p_r))$$

$$- p'_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p'_r))]$$

$$= \sum_{s \in \Pi} \mu_s(\tau) [p_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p_r))$$

$$+ (1 - p_r) F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), 1 - p_r))$$

$$- p'_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p'_r))$$

$$- (1 - p'_r) F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), 1 - p'_r))]$$

$$\geqslant \sum_{s \in \Pi} \mu_s(\tau) [p'_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p'_r)) \qquad (25)$$

$$+ (1 - p'_r) F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), 1 - p'_r))$$

$$- p'_r F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), p'_r))$$

$$- (1 - p'_r) F_{\tau+1}(\mathcal{T}(B(\tau_0 + 1), 1 - p'_r))] \geqslant 0$$

The last second inequality is because $p_r \geqslant p'_r$. The last inequality is because of equations (23) and (24). This finishes the proof.