Quantifying Worker Reliability for Crowdsensing Applications: Robust Feedback Rating and Convergence

Hong Xie, *Member, IEEE* and John C. S. Lui, *Fellow, IEEE*

Abstract—Worker reliability estimation is a fundamental problem in crowdsensing applications. This paper proposes a robust feedback rating approach to estimate worker reliability explicitly. In this approach, the requester provides a feedback rating to reflect the quality of the sensor data submitted by each worker. The aggregation of each worker's historical feedback ratings serves as a reliability estimate. The challenges are: *(1) Feedback ratings are subjected to noise; (2) Workers' cognitive bias in task selection leads to sensor data quality variations.* We develop a mathematical model to quantify rating noise from requesters and the degree of cognitive bias of workers in task selection. We derive sufficient conditions, under which the aggregate rating is asymptotically accurate in estimating worker reliability, via stochastic approximation techniques. These conditions identify a class of asymptotically accurate rating aggregation rules for crowdsensing applications. We further derive the minimum number of ratings needed to guarantee a given reliability estimation accuracy, via martingale theory. Via extensive experiments: (1) We reveal fundamental understandings on how various factors such as rating noise influence the minimum number of ratings needed to achieve certain accuracy; (2) We show that our feedback rating approach improves air quality index estimation accuracy by as high as 50 percent over the a typical baseline algorithm.

Index Terms—Crowdsensing, worker reliability, feedback rating, stochastic approximation, martingale

1 Introduction

Crowdsensing applications [1] are becoming increasingly popular due to the prevalence of mobile devices such as smartphones, which are equipped with sensors such as GPS, compass, etc. Crowdsensing is a paradigm that outsources the task of sensor data collection to a crowd of mobile device users, and it is much cheaper than hiring data sensing professionals. Crowdsensing has been applied to a variety of applications such as air quality monitoring, health care, smart transportation, etc. In general, a crowdsensing platform is composed of requesters, workers (i.e., mobile device users) and data sensing tasks. Requesters post data sensing tasks on the crowdsensing platform, workers use their mobile devices to collect sensor data accordingly and transmit the data back to the requester when it is done. Finally, the requester aggregates the data to produce an estimate on the *true information*, e.g., the ground truth air quality.

However, the data collected by workers is subjected to noise [2]. This noise is caused by a variety of sources such as the innate quality of the sensor in a mobile device, the skill or expertise of a worker, the amount of efforts that a worker exerts and the difficulty of a sensing task. As a consequence, the data collected by different workers can vary greatly or even be conflicting. In order to improve the accuracy of estimating the *true information*, it is important to collect sensor data from reliable workers or adjust the estimate toward reliable workers [3], [4]. However, worker reliability is *unknown* to the requester.

This paper proposes a new alternative: *utilizing a feedback rating approach to estimate workers' reliability.* In this approach, the crowdsensing platform operator (i.e., the company who provides the crowdsensing platform to requesters and workers) deploys a feedback rating system and a rating aggregation rule. After a sensing task is completed, the requester provides a feedback rating to each worker reflecting the quality of her submitted data. A higher feedback rating implies that a worker's data is of higher quality. Finally, a rating aggregation rule summarizes each worker's feedback ratings to produce an indicator on her reliability, which is public to all requesters and the worker herself.

One essential component of a feedback rating system is the sensor data quality assessment, which assists the requester to assign a feedback rating. There are three typical approaches to assess sensor data quality. The first approach directly applies some previous sensor data quality assessment algorithms such as [5] and [6]. In particular, Yang *et al.* [5] proposed a clustering based approach to assess sensor data quality. Peng *et al.* [6] designed an expectation-maximization algorithm to assess sensor data quality. The second approach requires requesters to use their own knowledge or expertise to evaluate the sensor data quality. In fact, this

- *Hong Xie is with the Chongqing Key Laboratory of Software Theory and Technology, Chongqing University, Shapingba, Chongqing 400044, China. E-mail: xiehong2018cqu.edu.cn.*
- *John C. S. Lui is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong. E-mail: cslui@cse.cuhk.edu.hk.*

Manuscript received 22 Aug. 2020; revised 22 Mar. 2021; accepted 3 Apr. 2021. Date of publication 12 Apr. 2021; date of current version 5 Dec. 2022. (Corresponding author: Hong Xie.)
Recommended for acceptance by J. Tang.
Digital Object Identifier no. 10.1109/TMC.2021.3072477

method is deployed in real-world crowdsourcing systems like Upwork[1] and Freelancer.[2] The third approach estimates the true information by some previous algorithms such as [2], [3], [5], and then assess sensor data quality via calculating the variation or distance between the sensor data and the *estimate* of the true information. We use the following simplified example to illustrate this data quality assessment method.

Example 1. Assume the air quality index (AQI) is measured in the range $[0, 100]$. Suppose three workers w_1, w_2 and w_3 report the AQI of 10, 60 and 80 respectively, and the requester estimates the ground truth AQI via the simple average, i.e., $(10 + 60 + 80)/3 = 50$. We are aware of sophisticated ground truth estimation method [2], [3], [5], but here we use the simple average for simplicity of presentation. Consider a numerical feedback rating metric in the range [0,1]. One simple rating protocol is: the worker w_1 receives a feedback rating $1 - |50 - 10|/100 = 0.6$, w_2 receives $1 - |50 - 60|/100 = 0.9$, and w_3 receives $1 - |50 - 80|/100 = 0.7$. Suppose the crowdsensing platform operator uses the "average scoring rule" to estimate worker reliability. E.g., if w_1, w_2 and w_3 have only one feedback rating, their reliability scores are 0.6, 0.9 and 0.7 respectively. Suppose w_1, w_2 and w_3 participate another sensing task and earn feedback ratings of $0.15, 0.85$ and 0.75 respectively, their reliability scores become $(0.6 + 0.15)/2 = 0.375$, $(0.9 + 0.85)/2 = 0.875$ and $(0.7 + 0.75)/2 = 0.725$.

One advantage of this feedback rating approach is that each worker is tagged with an explicit reliability score (please refer to Section 5 for more discussions on the advantage of the feedback rating approach). This reliability score can be used to improve sensor data aggregation [2], task matching or recommendation [3], or incentive mechanism design [7], etc.

The challenge is *how to "accurately" estimate worker reliability via feedback ratings.* For the crowdsensing platform operator, it is non-trivial to select an appropriate rating aggregation rule because some rating aggregation rules may not be accurate or robust in estimating worker reliability. What makes it challenging is that feedback ratings are subjected to noise. All of these three aforementioned sensor data quality assessment methods can cause feedback rating noise. The reason is that the accuracy of the sensor data quality assessment algorithms [5], [6] is not 100 percent, requesters are subjected to human bias or errors in evaluating sensor data quality [8], and the accuracy of the true information estimation methods [2], [3], [5] is not 100 percent. The following simplified example illustrates rating noise.

Example 2 (Rating noise). Suppose the ground truth AQI in Example 1 is 15. Under this ground truth AQI, feedback ratings should be: w_1 receives $1 - |15 - 10|/100 = 0.95$, w_2 receives $1 - |15 - 60|/100 = 0.55$, and w_3 receives $1 - |15 - 80|/100 = 0.35$. Comparing with the feedback rating in Example 1, the feedback rating noise or error for w_1 is $0.95 - 0.6 = 0.35$, for w_2 is $0.55 - 0.9 = -0.35$, and for w_3 is $0.35 - 0.7 = -0.35$.

1. https://www.upwork.com/
2. https://www.freelancer.com/

Note that the rating noise may also lead to cognitive bias in task selection [9], e.g., the worker selects a task which may not be appropriate at her reliability level. As a consequence, it results in higher variations in sensor data quality as illustrated in the following example.

Example 3 (Cognitive bias). Suppose workers' skill level is measured in a range $[0, 5]$, and the ground truth skill level of w_2 is 3. Due to noise or bias in feedback ratings, she receives a sequence of feedback ratings as: $0.9, 0.85, 0.98, 0.9, 0.9$. In this situation, she may over estimate her own skill level to be of 4. This over assessment may lead her to select more challenging tasks (i.e., fits for workers having skill level of 4 as above). As a consequence, she submits a low quality sensor data because her ground truth skill level is only of 3.

Hence, rating noise and cognitive bias in task selection exists in a feedback rating system. This paper aims to explore a robust rating aggregation rule to handle rating noise and cognitive bias in task selection. We do not restrict to any instance of feedback rating assigning methods (or sensor data quality estimation methods), but instead we develop a model to abstract these methods via rating noise and cognitive bias in task selection and study how to design robust rating aggregation rules. We aim to answer: *(1) How to develop a mathematical model to characterize the rating noise and cognitive bias in task selection? (2) Under what conditions, an aggregate rating can accurately reflect the worker reliability? (3) What's the speed of revealing the worker reliability?* Our contributions are:

- We develop a mathematical model to capture key factors that affect the accuracy of feedback ratings in reflecting the sensor data quality. Our model quantifies the *degree of rating noise*and *cognitive bias*.
- We derive sufficient conditions, under which the aggregate rating is asymptotically accurate in estimating worker reliability, via the stochastic approximation theory. These conditions enable us to identify a class of *asymptotically accurate rating aggregation rules*for crowdsensing applications. We further derive the *minimum number of ratings* needed to guarantee a given reliability estimation accuracy, via the martingale theory.
- We conduct extensive experiments and reveal that the averaging scoring rule is *sub-optimal* in terms of estimation accuracy, and a *recency aware rating aggregation rule* can significantly improve the accuracy under the same number of ratings. To achieve the same estimation accuracy, the minimum number of ratings needed *decreases slightly* as the degree of cognitive bias increases, while it *increases significantly* as the degree of rating bias increases. Our feedback rating approach improves AQI estimation accuracy by as high as 50 percent over the URP algorithm [3], especially when the sensor data is noisy.

The reminder of this paper organizes as follows. Section 2 presents the crowdsensing system model, rating behavior model, task selection behavior model and rating aggregation rule model. Section 3 presents the theoretical analysis on the accuracy of worker reliability estimation using feedback ratings. Section 4 presents the experimental results. Section 5 discusses the related work and Section 6 concludes.

2 System Model

We start with a baseline model of estimating worker reliability. Then, we formulate models to characterize the cognitive bias of task selection and the rating noise from requesters.

2.1 The Baseline Model

Worker Reliability. Consider a crowdsensing platform where requesters post data sensing tasks and workers are allowed to select any posted task to solve. Without loss of generality, we focus on one worker denoted by w in our presentation. We denote the skill level of worker w by

$$s \in \mathcal{S} \triangleq [0, s_{\max}], \tag{1}$$

where $s_{\max} \in \mathbb{R}_+$ denotes the maximum possible skill level. The skill level s captures the innate quality of the sensor in a mobile device, the expertise of a worker, ..., etc. A larger value of s implies that a worker is more skilled. We denote the reliability of worker w by

$$\theta \in \mathcal{M} \triangleq [0, M], \tag{2}$$

where $M \in \mathbb{R}_+$ denotes the maximum possible reliability level. The reliability θ reflects the average quality of the sensor data submitted by worker w. The rationality of this modeling of average quality is to capture uncertainty in sensor data quality submitted by a worker. Due to human factors such as consistency of workers in solving tasks [10], [11] and the data collection environment such as weather, a worker may submit sensor data with different qualities even when she repeats the same task twice. For example, $\theta = \mathbb{E}[M(1-X)]$, where X is a random variable denoting the normalized error of the data submitted by worker w. A larger value of θ implies a more reliable worker. We formally model workers' reliability θ on a task with certain skill requirement as

$$\theta \triangleq f(s, v), \tag{3}$$

where $v \in \mathcal{S}$ denotes overall skill requirement of the data sensing tasks that the worker w selects to solve, and $f : \mathcal{S} \times \mathcal{S} \to \mathcal{M}$ denotes the reliability function. Also, v can be interpreted as the average skill requirement of tasks selected by the worker w, who knows the real skill value s of herself. The rationality behind this model of average skill requirement is to capture that workers may select different tasks which may require different skill levels. A larger value of v implies that on average the worker w tends to select tasks with higher skill requirement. The function $f(s, v)$ models the average reliability θ of the worker w. It is increasing in s and decreasing in v, capturing that on average a worker's reliability increases (or the quality of submitted data increases) if she is more skilled, while her reliability decreases (or the quality of submitted data decreases) if she selects data sensing tasks with higher skill requirement (i.e., a higher chance that a selected task is not appropriate for her skill level). Workers has financial incentives to select tasks with high skill requirement [7], [12], [13]. Note that the function f is unknown to the worker and the crowdsensing platform operator. Workers need to learn from their crowdsensing experience to select appropriate tasks. Unless stated explicitly, we assume that all the functions in this paper are continuously differentiable. One possible example of f can be expressed as

$$f(s, v) = M \times \frac{s - v + s_{\max}}{2 s_{\max}}, \tag{4}$$

which is a linear reliability function. One can check that this linear form of $f(s, v)$ is well-defined, i.e., $f(s, v) \in \mathcal{M}$.

One alternative to model worker reliability is via a stair function $F(s, V) = M \mathbb{1}_{\{s \geq V\}}$, where V denotes the skill requirement of a task selected by worker w. Here, V is a random variable capturing the uncertainty in the skill requirement. The average reliability of worker w can be derived as $\mathbb{E}[F(s, V)] = M\mathbb{P}[s \geq V]$. Suppose V follows a normal distribution with mean v and variance 1. Then, the average worker reliability $\mathbb{E}[F(s, V)]$ is a function of the mean v of this normal distribution. Furthermore, the average worker reliability $\mathbb{E}[F(s, V)]$ is decreasing in v. Namely, the average worker reliability $\mathbb{E}[F(s, V)]$ is a special case of Equation (3). This example shows that Equation (3) is reasonable and flexible.

Feedback Ratings and Reliability Estimation. Note that the worker reliability θ is unknown to requesters and the crowdsensing platform operator. To reveal a worker's reliability, a feedback rating system is deployed by the crowdsensing platform operator. After a sensing task is completed, the requester provides a feedback rating to each worker reflecting the quality of her sensor data. The feedback rating is recorded in the profile of a worker. Without loss of generality, we consider a cardinal rating metric $\mathcal{M}$. A higher rating means that the sensor data is of higher quality. Let $R_i \in \mathcal{M}$ denote the ith rating of the worker w. For example, $R_i = M(1 - X_i)$, where $X_i \in [0, 1]$ is a random variable denoting an estimate of the normalized error of the data submitted by worker w. The crowdsensing platform operator deploys a weighted average scoring rule to estimate the worker w's reliability from past ratings, i.e.,

$$\Theta_i \triangleq \frac{\sum_{j=1}^{i} \alpha_j R_j}{\sum_{j=1}^{i} \alpha_j}, \ \forall i \in \mathbb{N}_+, \tag{5}$$

where Θ_i denotes an estimate of θ, and $\alpha_j \in \mathbb{R}_+$ denotes the weight of the jth rating. The aggregate rating Θ_i is public to all workers and requesters. The weights $\alpha_j, \forall j$, are controlled by the crowdsensing platform operator. For example, the simple average scoring rule can be achieved via $\alpha_j = 1, \forall j$, and

$$\Theta_i = \frac{R_1 + \ldots + R_i}{i}. \tag{6}$$

Another possibility is to use a recency aware rating aggregation rule where $\alpha_j = j, \forall j$, and

$$\Theta_i = \frac{R_1 + 2R_2 + \ldots + iR_i}{1 + \ldots + i}. \tag{7}$$

Namely, a more recent rating is weighted more. Based on Θ_i, we first define the status of the worker w.

Definition 1. *The worker w is over rated if $\Theta_i > \theta$, under rated if $\Theta_i < \theta$, and correctly rated if $\Theta_i = \theta$.*

Note that the status of a worker (i.e., over, under or correctly rated) is unknown to the requesters and the crowdsensing platform operator because θ is unknown to them.

2.2 Workers' Task Selection Behavior Model

We model the cognitive bias [9] in task selection via two steps: (1) Model how historical ratings may influence workers' self-assessment of skills; (2) Model how skill self-assessment influences task selection. Then, we model how the task selection behavior influences the quality of sensor data submitted by a worker.

Skill Self-Assessment. We define the rating history of worker w up to the ith rating as

$$\mathcal{H}_i \triangleq \{R_1, \ldots, R_i\},\ \forall i \in \mathbb{N}_+, \tag{8}$$

and we define $\mathcal{H}_0 = \emptyset$ for presentation convenience. The historical ratings $\mathcal{H}_i$ serve as persuasion messages, and the message-based persuasion effect [14], [15] (a well-known psychological effect) implies that the historical ratings $\mathcal{H}_i$ may influence workers in assessing their own skills. Consider worker w, let

$$S_i \triangleq \tilde{g}(\mathcal{H}_i),\ \forall i \in \mathbb{N}_+, \tag{9}$$

denote her skill self-assessment from historical ratings $\mathcal{H}_i$, where $\tilde{g} : \mathcal{H} \to \mathcal{S}$ denotes the self-assessment function and

$$\mathcal{H} \triangleq \{\mathcal{H}_i : \forall \mathcal{H}_i, \forall i \in \mathbb{N}_+\} \tag{10}$$

denotes all possible rating histories. We define S_0 as a random variable to capture the noise in the initial skill assessment, and $\mathbb{E}[S_0] = s$ to capture that the initial assessment is statistically unbiased. Note that the self-assessed skill S_i is known to the worker w only and unknown to the requesters and the crowdsensing platform operator.

Definition 2. *The worker w over assesses her skill if $S_i > s$, under assesses if $S_i < s$, and correctly assesses if $S_i = s$.*

Our model also captures the special case that the worker w always correctly assesses her skill by $\tilde{g}(\mathcal{H}_i) = s, \forall \mathcal{H}_i$. In practice, the worker w may over or under assess her skill. For mathematical tractability of this practical scenario, we assume the following to capture dominating factors.

Assumption 1. *The self-assessment function $\tilde{g}$ satisfies that*

$$\tilde{g}(\mathcal{H}_i) = g(\Theta_i), \qquad \forall \mathcal{H}_i, \forall i \in \mathbb{N}_+, \tag{11}$$

where $g : \mathcal{M} \to \mathcal{S}$.

Assumption 1 captures that the influence of the historical ratings $\mathcal{H}_i$ on the self-assessment behavior is dominated by the reliability score Θ_i. We also call g the self-assessment function to simplify our presentation. The self-assessment function $g(\Theta_i)$ is non-decreasing in Θ_i and satisfies $g(\Theta_i) = s$ if $\Theta_i = \theta$. This property captures that worker w over/under/correctly assesses her skills when she is over/under/correctly rated. One possible example of self-assessment function g can be

$$g(\Theta_i) = s + s_{\max} \times \beta \times \frac{\Theta_i - \theta}{M}, \tag{12}$$

where β satisfies

$$0 \le \beta \le \min\left\{\frac{s}{s_{\max}}\frac{M}{\theta}, \frac{s_{\max} - s}{s_{\max}}\frac{M}{M - \theta}\right\}, \tag{13}$$

in order to guarantee that $g(\Theta_i)$ is well-defined, i.e., $g(\Theta_i) \in \mathcal{S}$. The β models the sensitivity of self-assessment to historical ratings. A larger value of β implies that worker w's self-assessment is more sensitive to historical ratings. Formation of the self-assessment is a complex psychological behavior. Worker skill s and reliability θ are two latent factors that affect this psychological behavior implicitly. Workers do not know exact values of them but they know their self-assessment, i.e., the value of $g(\Theta_i)$. This is similar to generative modeling in recommendation applications, where users know their preferences quantified by ratings, but they do not known their latent feature vectors, which are used to model their ratings [16]. Equation (12) illustrates an example to approximate this psychological behavior. In Equation (12), the parameters s and θ quantify how different users form different self-assessments. This model enables us to simulate self-assessment behavior of workers and based on it we study the empirical performance of the feedback rating system.

Self-Selection Behavior. The self-selection behavior means that the worker w selects the tasks based on her self-assessed skill, i.e., S_i. Let $V_i \in \mathcal{S}$ denote the overall skill requirement of the tasks selected by the worker w whose self-assessed skill is S_i. Note that $V_i = v$ if the worker w accurately assessed her own skill, i.e., $S_i = s$. However, it may happen that the self-assessed skill is not accurate, i.e., $S_i \neq s$, leading to that the worker w selects tasks that are not appropriate for her skill, i.e., $V_i \neq v$.

Definition 3. *We define the cognitive bias as the case $V_i \neq v$.*

Namely, the cognitive bias means that the worker selects a task which may not be appropriate at her reliability level. We formally model the self-selection behavior via

$$V_i \triangleq h(S_i), \qquad \forall i \in \mathbb{N}, \tag{14}$$

where $h : \mathcal{S} \to \mathcal{S}$ denotes the self-selection function. We classify the self-selection behavior as follows.

Definition 4. *The worker over selects tasks if $V_i > v$, under selects tasks if $V_i < v$, and properly selects tasks if $V_i = v$.*

The self-selection function $h(S_i)$ is non-decreasing in S_i, and satisfies that $h(S_i) = v$ if $S_i = s$. This property captures that the worker w tends to over/under/properly select tasks if she over/under/correctly assesses her skills. One possible example self-selection function h is

$$h(S_i) = v + \gamma \times (S_i - s), \tag{15}$$

where γ satisfies

$$0 \le \gamma \le \min\left\{\frac{v}{s}, \frac{s_{\max} - v}{s_{\max} - s}\right\}, \tag{16}$$

to guarantee $h(S_i) \in \mathcal{S}$. The γ models the sensitivity of self-selection to skill assessment S_i, i.e., a larger value of γ models that the overall skill requirement V_i of the selected tasks is more sensitive to skill assessment S_i. With similar reasons as Equation (12), Equation (15) contains parameters s and v that are unknown to the worker and Equation (15) mainly enables us to simulate the self-selection behavior of workers. Furthermore, if $g(\Theta_i)$ satisfies Equation (12), we can have

$$V_i = v + \gamma\left(s + s_{\max}\beta\frac{\Theta_i - \theta}{M} - s\right) = v + \beta\gamma s_{\max}\frac{\Theta_i - \theta}{M}. \tag{17}$$

It implies that the worker w tends to over/under/properly select tasks to solve if she is over/under/correctly rated.

Data Quality Variations. Note that the worker w may over or under select tasks to solve. Therefore, the quality of the sensor data submitted by the worker w evolves dynamically. Given the overall skill requirement V_i of the selected tasks, we denote the corresponding data quality as $\tilde{\Theta}_i \in \mathcal{M}$. Without loss of generality, we formally quantify sensor data quality using the reliability metric, i.e.,

$$\tilde{\Theta}_i = f(s, V_i), \qquad \forall i \in \mathbb{N}. \tag{18}$$

Suppose Equations (4), (12), and (15) hold. One possible example of $\tilde{\Theta}_i$ can be derived as

$$\tilde{\Theta}_i = M \times \frac{s - V_i + s_{\max}}{2s_{\max}} = \theta + \frac{\beta\gamma}{2} \times (\theta - \Theta_i). \tag{19}$$

From this simple example, one can observe that the worker w behaves less reliably (i.e., $\tilde{\Theta}_i < \theta$) if she is over rated, and she behaves more reliably (i.e., $\tilde{\Theta}_i > \theta$) if she is under rated. One may suggest to under rate worker to push them to behave more reliably. However, workers may get discouraged if we keep underrating them and they may even leave a crowdsensing platform. Hence, for the ecosystem of a crowdsensing platform, honest rating is essential. Note that the data quality influences the rating provided by requesters, leading to rating variations.

Remark. For a newly registered worker, her reliability score set is empty. Namely, we do not impose any prior information on the reliability of a new worker. Since she has no feedback ratings, and she has no idea about her own status, i.e., over rated/under rated. She will learn or assess her skill level through working on tasks. In selecting the first task, a worker may over select or under select. And we model it as an uncertain behavior by inducing a statistical unbiased condition modeled in the next subsection.

2.3 Requesters' Rating Behavior Model

Instead of restricting to any instance of feedback rating assigning methods, we develop a model to abstract these methods via rating noise or biases and we study how to design robust rating aggregation rules. Our results can be directly applied to an instance of rating protocol if it satisfies these properties characterized by our model. To capture the noise in feedback rating, we model each rating R_i as a random variable. If the rating is unbiased, each rating is statistically accurate in reflecting the quality of sensor data, i.e., $\mathbb{E}[R_{i+1}] = \tilde{\Theta}_i$. For example, consider $R_i = M(1 - X_i)$. The unbiasness property $\mathbb{E}[R_{i+1}] = \tilde{\Theta}_i$ can be achieved by that X_i accurately reflects the normalized data error. This can be achieved in practice, but it may incur a high cost. For example, the requester can hire some professionals to produce an accurate estimate of the true information and then use it to compute the normalized data error. We thus consider a general scenario that the feedback rating can be biased but it may also be cheaper to implement. We consider the rating bias caused by the historical ratings of each worker. This corresponds to the scenario that requesters incorporate each workers' historical performance record in aggregating the sensor data, or estimating the quality of sensor data. Formally,

$$\mathbb{E}[R_{i+1}|\mathcal{H}_i] = q(\Theta_i, \tilde{\Theta}_i), \qquad \forall i \in \mathbb{N}_+, \tag{20}$$

where $q : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ denotes the rating bias function. One can use different instances of q to model different crowdsensing application scenarios. For example, the special instance $q(\Theta_i, \tilde{\Theta}_i) = \tilde{\Theta}_i, \forall \tilde{\Theta}_i \in \mathcal{M}$, models the unbiased scenario.

Assumption 2. *The rating bias function $q(\Theta_i, \tilde{\Theta}_i)$ is non-decreasing in Θ_i and increasing in $\tilde{\Theta}_i$. Furthermore, $q(\Theta_i, \tilde{\Theta}_i) = \Theta_i$ if $\tilde{\Theta}_i = \Theta_i$.*

Assumption 2 captures that a worker receives a higher rating (on average) from requesters if she submits a higher quality sensor data, or if she has a more positive historical performance record (i.e., the aggregate rating). If the aggregate rating Θ_i accurately reflects the data quality, i.e., $\Theta_i = \tilde{\Theta}_i$, the feedback rating becomes unbiased. Namely, we consider the rating bias caused by the mismatch between the aggregate rating and the sensor data quality. We set $\mathbb{E}[R_1|\mathcal{H}_0] = \tilde{\Theta}_0$ to capture that the first rating is statistically unbiased because the worker does not have any rating.

To illustrate, the special case $q(\Theta_i, \tilde{\Theta}_i) = \tilde{\Theta}_i$ (i.e., unbiased scenario) satisfies Assumption 2. Another possible case of q can be

$$q(\Theta_i, \tilde{\Theta}_i) = \eta\Theta_i + (1 - \eta)\tilde{\Theta}_i, \tag{21}$$

where $\eta \in [0, 1)$ models the strength of rating bias. Note that $\eta = 0$ models the unbiased scenario. A larger value of η implies a strong rating bias toward Θ_i, or a less accurate rating protocol.

2.4 Technical Questions

Based on our model, we seek to understand the efficiency of the feedback rating approach. Formally, we explore the following questions from three different perspectives:

- *The crowdsensing platform operator:* The crowdsensing platform operator deploys the feedback rating system and sets the parameter α_i for the feedback rating system. How to select α_i for the rating aggregation rule? Will the true reliability score θ be revealed, i.e., Θ_i converges to θ? What is the speed of revealing it?

TABLE 1
Main Notations

s, θ	true skill level and reliability level of worker w
$\mathcal{S}, \mathcal{M}$	skill level and reliability level set
$s_{\max}, M$	maximum possible skill level and reliability level
$f(s, v)$	reliability function
R_i, α_i	the ith rating and the weight of ith rating
Θ_i	an estimate of θ from $R_1, \ldots, R_i$
$\mathcal{H}_i$	the rating history of worker w up to ith rating
$\mathcal{H}$	a set of all possible rating history
S_i	w's skill self-assessment from historical ratings $\mathcal{H}_i$
$g(\cdot)$	self-assessment function
β	sensitivity parameter of linear self-assessment function
V_i	overall skill requirement of tasks corresponding to S_i
$h(\cdot)$	self-selection function
γ	sensitivity parameter of the linear self-selection function
$\tilde{\Theta}_i$	the data quality corresponding to V_i
$q(\cdot, \cdot)$	the rating bias function
η	strength of rating bias
$\tilde{\alpha}_i$	normalized rating weight $\tilde{\alpha}_{i+1} \triangleq \alpha_i / \sum_{j=1}^{i} \alpha_j$
W_i	the noise of rating R_i
$\tilde{q}(\cdot)$	the reliability of worker w under Θ_i
μ	parameter for the rating aggregation rule
ρ	parameter that quantifies the sensitivity of reliability to worker skill
ω	parameter that quantifies the sensitivity of sill assessment to aggregate rating
ν	sensitivity parameter of sill requirement of selected tasks to skill assessment
τ	parameter that quantifies the sensitivity of rating bias to Θ_i and $\tilde{\Theta}_i$

- *Worker:* Will cognitive bias be eliminated, i.e., $(S_i, V_i, \tilde{\Theta}_i)$ converges to (s, v, θ)? What is the speed of eliminating it?
- *Requester:* Will rating bias be eliminated, i.e., $q(\Theta_i, \tilde{\Theta}_i)$ converges to $\tilde{\Theta}_i$? What is the speed of eliminating it?

Table 1 summarizes key notations used this paper.

3 Theoretical Analysis

We derive sufficient conditions, under which the aggregate rating is asymptotically accurate in estimating worker reliability, via the stochastic approximation theory. These conditions enable us to identify a class of asymptotically accurate rating aggregation rules for crowdsensing applications. We further derive the minimum number of ratings needed to guarantee a given reliability estimation accuracy, via the martingale theory.

3.1 Implications From the Model

We first state a lemma, which provides a connection between $S_i, V_i, \tilde{\Theta}_i$ and Θ_i, and prove their monotone properties.

Lemma 1. *Under Assumption 1, we derive S_i, V_i and $\tilde{\Theta}_i$ as*

$$S_i = g(\Theta_i), \qquad V_i = h(g(\Theta_i)), \\ \tilde{\Theta}_i = f(s, h(g(\Theta_i))). \tag{22}$$

Furthermore, S_i is non-decreasing in Θ_i and $S_i = s$, V_i is non-decreasing in Θ_i and $V_i = v$ if $\Theta_i = \theta$, and $\tilde{\Theta}_i$ is non-increasing in Θ_i and $\tilde{\Theta}_i = \theta$ if $\Theta_i = \theta$.

We present proofs to lemmas and theorems in the supplementary file. Lemma 1 states that if worker w is correctly rated (i.e., $\Theta_i = \theta$), she can correctly assess her skill (i.e., $S_i = s$) and properly select tasks to solve (i.e., $V_i = v$), and the data quality matches the ground truth reliability (i.e., $\tilde{\Theta}_i = \theta$). Over/under rating worker w results in that she over/under assesses her skill and over/under selects tasks to solve as well as the data quality increases/decreases. Lemma 1 also states that S_i, V_i and $\tilde{\Theta}_i$ are essentially determined by Θ_i. This connection reduces the study of the convergence of S_i, V_i, $\tilde{\Theta}_i$ and Θ_i to the study of the convergence of Θ_i. For some concrete understandings of Lemma 1, one can refer to examples of f, g and h that are derived in Equations (4), (12), and (15) respectively.

Note that the aggregate rating Θ_i evolves dynamically. Therefore, it is important to characterize how workers' task selection behavior as well as the rating bias change against the variation of the aggregate rating Θ_i. We impose the following assumption to make this characterization quantitative.

Assumption 3. *There exists a constant $\varrho \in \mathbb{R}_+$ such that*

$$\left|\frac{\partial f}{\partial v}\right| < \varrho, \left|\frac{dg}{d\Theta_i}\right| < \varrho, \left|\frac{dh}{dS_i}\right| < \varrho, \left|\frac{\partial q}{\partial \Theta_i}\right| < \varrho, \left|\frac{\partial q}{\partial \tilde{\Theta}_i}\right| < \varrho. \tag{23}$$

Assumption 3 captures that a small variation in the *mean of skill requirement v, or aggregate rating Θ_i, or skill self-assessment S_i*does not drastically change the *reliability score θ, or skill self-assessment S_i, or mean of skill requirement $\tilde{\Theta}_i$*. Furthermore, a small variation in the aggregate rating or quality of sensor data does not drastically change the mean of feedback rating. For example, the reliability function f derived (4), the self-assessment function g derived in (12), the self-selection function h derived (15) and the rating bias function q derived in (21) satisfy Assumption 3.

Assumption 3 implies the following lemma, which states that workers' task selection behavior will not change drastically under a small variation in aggregate rating Θ_i.

Lemma 2. *Under Assumptions 1 and 3, we have*

$$\left|\frac{dS_i}{d\Theta_i}\right| < \varrho, \qquad \left|\frac{dV_i}{d\Theta_i}\right| < \varrho^2, \qquad \left|\frac{\partial \tilde{\Theta}_i}{\partial \Theta_i}\right| < \varrho^3. \tag{24}$$

Lemmas 1 and 2 serve as building blocks for later analysis.

3.2 Revealing Ground Truth Reliability

We study the convergence of the aggregate rating Θ_i. In particular, we are interested in whether Θ_i converges to θ, i.e., revealing the ground truth reliability score. In the following lemma, we characterize the updating dynamics of aggregate rating Θ_i.

Lemma 3. *Under Assumption 1, the updating dynamics of Θ_i can be characterized by the following dynamical system*

$$\Theta_{i+1} = (1-\tilde{\alpha}_{i+1})\Theta_i + \tilde{\alpha}_{i+1}(\tilde{q}(\Theta_i) + W_{i+1}), \tag{25}$$

where $\tilde{\alpha}_{i+1}$, $\tilde{q}(\Theta_i)$ *and* W_{i+1} *are defined as*

$$\tilde{\alpha}_{i+1} \triangleq \frac{\alpha_{i+1}}{\sum_{j=1}^{i+1}\alpha_j}, \tag{26}$$

$$W_{i+1} \triangleq R_{i+1} - \tilde{q}(\Theta_i), \tag{27}$$

$$\tilde{q}(\Theta_i) \triangleq q(\Theta_i, f(s, h(g(\Theta_i)))). \tag{28}$$

The W_{i+1} *satisfies* $\mathbb{E}[W_{i+1}|\mathcal{H}_i] = 0$ *and* $\mathbb{E}[W_{i+1}^2|\mathcal{H}_i] \leq M^2$.

Lemma 3 derives a dynamical system to characterize the updating dynamics of Θ_i. This dynamical system enables us to study the convergence of Θ_i from the *stochastic approximation* [17] perspective. In particular, the $\tilde{\alpha}_i$ is interpreted as the updating step size, W_{i+1} is interpreted as the stochastic noise and $q(\Theta_i)$ is interpreted as the objective function. This interpretation can reveal fundamental understandings on the role that workers, requesters and the crowdsensing platform operator play in the convergence of Θ_i.

- *The role of the crowdsensing platform operator.* Note that $\tilde{\alpha}_i$ corresponds to the weights of the rating aggregation rule. Therefore, the weights of the rating aggregation rule, i.e., $\alpha_i, \forall i$, only influence whether Θ_i converges or not and they have nothing to do with which value the Θ_i converges to. It suggests that the crowdsensing platform operator should carefully select the weights, i.e., $\alpha_i, \forall i$, so as to guarantee the convergence of Θ_i. One example of $\tilde{\alpha}_i$ is

$$\tilde{\alpha}_i = \frac{\alpha_i}{\sum_{j=1}^{i}\alpha_j} = \frac{1}{i}, \tag{29}$$

  where $\alpha_j = 1, \forall j$. This example corresponds to the average scoring rule.
- *The role of the worker and requester.*Under appropriate weights $\alpha_i, \forall i$, if Θ_i converges, it converges to a value determined by the function $\tilde{q}(\Theta_i)$. From the analytical expression of $\tilde{q}(\Theta_i)$, one can observe that the rating bias function q, the reliability function f, the self-assessment function g and the self-selection function h jointly determine which value the Θ_i converges to. To illustrate $\tilde{q}(\Theta_i)$, consider f, g, h and q derived in Equations (4), (12), (15), and (21) respectively. Then we have

$$\begin{aligned}\tilde{q}(\Theta_i) &= \eta\Theta_i + (1-\eta)\left(\theta + \frac{\beta\gamma}{2}(\theta - \Theta_i)\right)\\ &= \Theta_i + (1-\eta)\left(1+\frac{\beta\gamma}{2}\right)(\theta-\Theta_i).\end{aligned} \tag{30}$$

In the following lemma, we provide further understandings on the role of the worker and requester via the function $\tilde{q}(\Theta_i)$.

Lemma 4. *Under Assumptions 1 and 2, the function* $\tilde{q}(\Theta_i)$ *has a unique fixed point, i.e.,* $\tilde{q}(\Theta_i) = \Theta_i$ *if and only if* $\Theta_i = \theta$.

Lemma 4 states that under mild assumptions on workers' task selection behavior and requesters' rating behavior, the function $\tilde{q}(\Theta_i)$ has a unique fixed point. This attractive property of $\tilde{q}(\Theta_i)$ serves as a building block for us to establish the convergence of the aggregate rating Θ_i, because the dynamical system derived in (25) aims to locate a fixed point of $\tilde{q}(\Theta_i)$. For some concrete understandings (or simple verification) of Lemma 4, one can work on the simple example of $\tilde{q}(\Theta_i)$ derived in Equation (30). Based on Lemma 4, we characterize the role of the crowdsensing platform operator and the convergence of Θ_i in the following theorem.

Theorem 1. *Suppose Assumptions 1-3 hold. Assume the weights* $\alpha_i, \forall i$, *of the rating aggregation rule satisfy*

$$\sum_{i=1}^{\infty}\tilde{\alpha}_i = \infty, \qquad \sum_{i=1}^{\infty}\tilde{\alpha}_i^2 < \infty. \tag{31}$$

We have $\mathbb{P}[\lim_{i\to\infty}\Theta_i = \theta] = 1$.

Theorem 1 states sufficient conditions, under which the aggregate rating Θ_i converges to θ. In other words, Θ_i is asymptotically accurate in estimating the ground truth reliability. Condition (31) characterizes a large class of asymptotically accurate rating aggregation rules for crowdsensing applications. For example, consider the average scoring rule, i.e., $\alpha_j = 1, \forall j$. Then we have $\tilde{\alpha}_i$ derived in Equation (29). Then it follows that

$$\sum_{i=1}^{\infty}\tilde{\alpha}_i \approx \lim_{i\to\infty}\ln i = \infty, \qquad \sum_{i=1}^{\infty}\tilde{\alpha}_i^2 = \sum_{i=1}^{\infty}\frac{1}{i^2} < 2. \tag{32}$$

This means that the average scoring rule is asymptotically accurate. Now consider a recency aware rating aggregation rule, i.e., $\alpha_j = j$. Then we have

$$\tilde{\alpha}_i = \frac{\alpha_i}{\sum_{j=1}^{i}\alpha_j} = \frac{i}{\sum_{j=1}^{i} j} = \frac{2}{i+1}. \tag{33}$$

With a similar derivation as Equation (32), we conclude that the rating aggregation rule with weights $\alpha_j = j, \forall j$, satisfies Condition (31). Now let us consider a more general class of rating aggregation rules, $\alpha_j = j^\mu$, where $\mu \in \mathbb{R}_+$. Then one can have

$$\sum_{j=1}^{i}\alpha_j = \sum_{j=1}^{i} j^\mu \approx \int_0^i x^\mu dx = \frac{1}{\mu+1}x^{\mu+1}\big|_0^i = \frac{i^{\mu+1}}{\mu+1}. \tag{34}$$

Then it follows that

$$\tilde{\alpha}_i = \frac{\alpha_i}{\sum_{j=1}^{i}\alpha_j} \approx i^\mu\frac{\mu+1}{i^{\mu+1}} = \frac{\mu+1}{i}. \tag{35}$$

With a similar derivation as Equation (32), we conclude that the rating aggregation rule with $\alpha_j = j^\mu, \forall j$, satisfies Condition (31).

Now, we apply Lemma 1 and Theorem 1 to characterize workers' task selection behavior and the requesters' rating behavior in an asymptotic sense.

Corollary 1. *Under Assumptions 1-3 and Condition (31), it holds that*

$$\mathbb{P}\left[\lim_{i\to\infty}(S_i, V_i, \tilde{\Theta}_i) = (s, v, \theta)\right] = 1, \tag{36}$$

$$\mathbb{P}\left[\lim_{i\to\infty} q(\Theta_i, \tilde{\Theta}_i) = \theta\right] = 1. \tag{37}$$

Corollary 1 states that in the long run, workers assess their skills correctly, select tasks properly and the quality of their data matches the ground truth reliability. Namely, the cognitive bias will be eliminated. On the requester side, the rating bias will eventually be eliminated.

3.3 Speed of Revealing Ground Truth Reliability

Now, we further study the speed of revealing the ground truth reliability. Examining the dynamical system derived in (25), one can observe that characterizing the convergence rate of Θ_i analytically is challenging in general. We first consider linear instances of f, g, h and q derived in Equations (4), (12), (15) and (21) respectively. We will complement these linear instances by extensive numerical studies on general non-linear instances in Section 4. Formally, when f, g, h and q are linear, the $\tilde{q}(\Theta_i)$ has the linear form derived in Equation (30). This linear form of $\tilde{q}(\Theta_i)$ enables us to identify a "martingale" structure, which governs the updating dynamics of Θ_i. We state it in the following lemma.

Lemma 5. *Suppose f, g, h and q satisfy Equation (4), (12), (15) and (21) respectively. Let*

$$k \triangleq \begin{cases} 0, & \text{if } \mathcal{K} = \emptyset, \\ \max \mathcal{K}, & \text{otherwise,} \end{cases} \quad \widehat{\Theta}_i \triangleq \begin{cases} \frac{\Theta_i - \theta}{\xi_i}, & \text{if i > k,} \\ 0, & \text{if i = k,} \end{cases} \tag{38}$$

where $\mathcal{K} \triangleq \arg_i\{\tilde{\alpha}_{i+1}(1-\eta)(1+\beta\gamma/2) = 1\}$ *and*

$$\xi_i \triangleq \begin{cases} \prod_{j=k+2}^{i}\left[1-\tilde{\alpha}_j(1-\eta)\left(1+\frac{\beta\gamma}{2}\right)\right], & \text{if } i \geq k+2, \\ 1, & \text{if } i = k+1. \end{cases} \tag{39}$$

Then, $\widehat{\Theta}_i$ forms a martingale with respect to the filtration $\mathcal{H}_i$, where $i \geq k$.

Lemma 5 identifies a martingale, which enables us to study the speed of revealing the ground truth reliability analytically. We like to remark that in the martingale statement, $\mathcal{H}_i$ represents the σ field generated by random variables $R_1, \ldots, R_i$. In the following theorem, we derive a metric to quantify the speed of revealing the ground truth reliability.

Theorem 2. *Suppose f, g, h and q satisfy Equations (4), (12), (15), and (21) respectively. Define ζ_i and $n(\epsilon, \delta)$ as*

$$\zeta_i \triangleq 2/\left(\xi_i^2 \sum_{j=k+1}^{i} \frac{\tilde{\alpha}_j^2}{\xi_j^2}\right), \qquad \forall i > k, \tag{40}$$

$$n(\epsilon,\delta) \triangleq \max_{i>k}\left\{\zeta_i \leq \frac{1}{\epsilon^2}\ln\frac{2}{\delta}\right\}, \qquad \forall \epsilon \in [0,1], \delta \in [0,1]. \tag{41}$$

If the number of ratings i satisfy $i \geq n(\epsilon, \delta)$, then Θ_i is ϵ-accurate, i.e., $|\Theta_i - \theta|/M \leq \epsilon$, with probability at least $1-\delta$.

Theorem 2 derives the minimum number of ratings $n(\epsilon, \delta)$ needed to guarantee Θ_i is ϵ-accurate with high probability (i.e., at least $1-\delta$). For each given (ϵ, δ) pair, smaller $n(\epsilon, \delta)$ reflects a faster speed in revealing the ground truth reliability, implying a higher efficiency of the feedback rating approach.

We next apply Theorem 2 to study the speed of eliminating the cognitive bias in workers' task selection behavior and the rating bias.

Corollary 2. *Suppose f, g, h and q satisfy Equation (4), (12), (15) and (21) respectively. We have*

$$i > n\left(\frac{\epsilon}{\beta}, \delta\right) \Rightarrow \mathbb{P}\left[\frac{|S_i - s|}{s_{\max}} < \epsilon\right] \geq 1-\delta, \tag{42}$$

$$i > n\left(\frac{\epsilon}{\beta\gamma}, \delta\right) \Rightarrow \mathbb{P}\left[\frac{|V_i - v|}{s_{\max}} < \epsilon\right] \geq 1-\delta, \tag{43}$$

$$i > n\left(\frac{2\epsilon}{\beta\gamma}, \delta\right) \Rightarrow \mathbb{P}\left[\frac{|\tilde{\Theta}_i - \theta|}{M} < \epsilon\right] \geq 1-\delta. \tag{44}$$

If $i > n(\epsilon/|1-(1-\eta)(1+0.5\beta\gamma)|, \delta)$, we have

$$\mathbb{P}\left[\frac{|q(\Theta_i, \tilde{\Theta}_i) - \theta|}{M} < \epsilon\right] \geq 1-\delta. \tag{45}$$

3.4 Robustness Against Misbehaving Ratings

Without loss of generality, we focus on one worker, i.e., worker w, to study the robustness of our feedback rating approach against misbehaving ratings from requesters. From worker w's perspective, the feedback rating itself matters instead of requesters who assign feedback ratings. Hence, we omit the requester ID in misbehaving ratings. Denote a set of $L \in \mathbb{N}_+$ misbehaving ratings as

$$\mathcal{R}_{\text{Misb}} \triangleq \left\{\left(i_\ell, \widetilde{R}_{i_\ell}\right) : \ell = 1, \ldots, L\right\}, \tag{46}$$

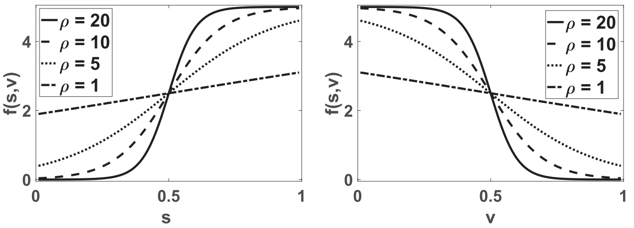
where $\widetilde{R}_{i_\ell} \in \mathcal{M}$ denotes the ℓth misbehaving rating and $i_\ell \in \mathbb{N}_+$ denotes the corresponding index. Note that $\mathcal{R}_{\text{Misb}}$ is a strong misbehaving rating attack, as we do not restrict both the arrival pattern and the value of misbehaving ratings. The following theorem characterizes the robustness of our feedback rating approach against $\mathcal{R}_{\text{Misb}}$.

Theorem 3. *Suppose Assumptions 1, 2, and 3 hold and $|\mathcal{R}_{\text{Misb}}| < \infty$. Assume the weights $\alpha_i, \forall i$, of the rating aggregation rule satisfy (31). We have $\mathbb{P}[\lim_{i\to\infty}\Theta_i = \theta] = 1$.*

Theorem 3 states that the aggregate rating Θ_i can still converges to θ if the number of misbehaving ratings is finite. Namely, our feedback rating approach is robust to a finite number of misbehaving ratings.

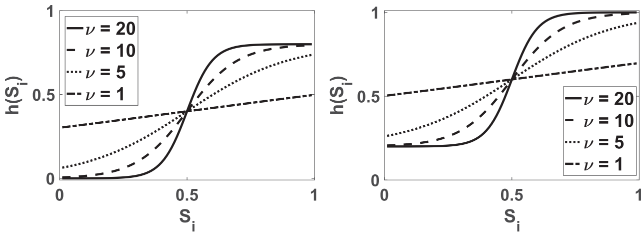
4 Experiments

We conduct experiments to study the impact of various model parameters on the convergence of Θ_i as well as the speed of Θ_i in revealing the ground truth reliability. We also apply our feedback rating approach to estimate AQI and show that it can improve estimation accuracy by as



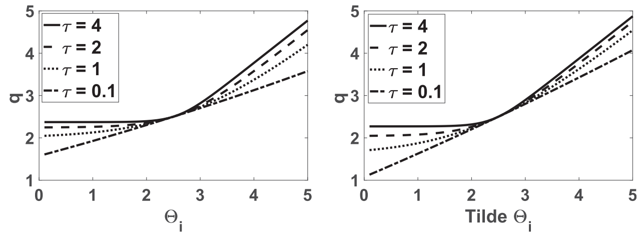
(a) $v = 0.5$ (b) $s = 0.5$

Fig. 1. Illustrating $f(s, v)$.



(a) $v = 0.4$ (b) $v = 0.6$

Fig. 3. Illustrating $h(S_i)$, where $s = 0.5$.



(a) $\tilde{\Theta}_i = 2.5$ (b) $\Theta_i = 2.5$

Fig. 4. Illustrating $q(\Theta_i, \tilde{\Theta}_i)$, where $\eta = 0.4$.

high as 50 percent over the URP algorithm [3], especially when the sensor data is noisy.

4.1 Experiment Settings

Without loss of generality, we set $s_{\max} = 1$ and $M = 5$ in our experiments. We consider the following non-linear form of reliability function f:

$$f(s, v) = M / \left[1 + \exp\left(-\rho \frac{s - v}{s_{\max}} \right) \right], \tag{47}$$

where $\rho \in \mathbb{R}_+$ models the sensitivity of the reliability θ to workers' skill and the skill requirement of tasks. A larger value of ρ models a higher degree of sensitivity. Fig. 1 illustrates $f(s, v)$ under different selections of ρ. One can observe that when $\rho = 1$ the curve of $f(s, v)$ is almost linear in both s and v, i.e., the linear form of f can be captured as a special case of $\rho = 1$.

We consider the following self-assessment function g:

$$g(\Theta_i) = \begin{cases} \frac{2s}{1+\exp\left(-\omega \frac{\Theta_i - \theta}{M}\right)}, & \text{if s} \leq \frac{s_{\max}}{2}, \\ 2s - s_{\max} + \frac{2(s_{\max} - s)}{1+\exp\left(-\omega \frac{\Theta_i - \theta}{M}\right)}, & \text{if } s > \frac{s_{\max}}{2}. \end{cases} \tag{48}$$

where $\omega \in \mathbb{R}_+$ models the sensitivity of skill assessment to the aggregate rating Θ_i. A larger value of ω models a higher degree of sensitivity. Fig. 2 illustrates the curve of $g(\Theta_i)$ under different selections of ω. Again, one can observe that we capture the linear form of g as a special case of $\omega = 1$.

We consider the following self-selection function h:

$$h(S_i) = \begin{cases} \frac{2v}{1+\exp\left(-\nu \frac{S_i - s}{s_{\max}}\right)}, & \text{if } v \leq \frac{s_{\max}}{2}, \\ 2v - s_{\max} + \frac{2(s_{\max} - v)}{1+\exp\left(-\nu \frac{S_i - s}{s_{\max}}\right)}, & \text{if } v > \frac{s_{\max}}{2}, \end{cases} \tag{49}$$

where $\nu \in \mathbb{R}_+$ models the sensitivity of the skill requirement of the selected tasks to skill self-assessment. A larger value of ν models a higher degree of sensitivity. Fig. 3 illustrates the curve of $h(S_i)$ under different selections of ν. Again, one can observe that we capture the linear form of h as a special case of $\nu = 1$.

We consider the following rating bias function q:

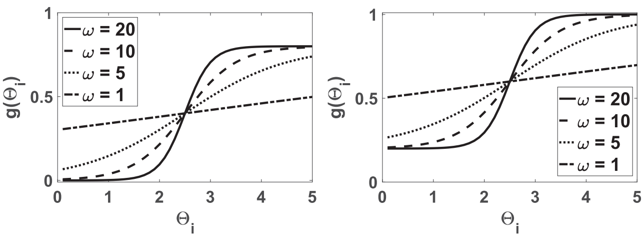
$$q(\Theta_i, \tilde{\Theta}_i) = \frac{1}{\tau} \ln\Big(\eta \, \exp(\tau \Theta_i) + (1 - \eta) \exp(\tau \tilde{\Theta}_i) \Big), \tag{50}$$

where $\tau \in \mathbb{R}_+$ models the sensitivity of rating bias to Θ_i and $\tilde{\Theta}_i$. A larger value of τ models a higher degree of sensitivity. Fig. 4 illustrates the curve of $q(\Theta_i, \tilde{\Theta}_i)$ under different selections of τ. Again, one can observe that we capture the linear form of q as a special case of $\tau = 0.1$.

The rating R_{i+1} follows a scaled beta distribution. More concretely, $R_{i+1} = MX$, where X follows a standard beta distribution, with mean $q(\Theta_i, \tilde{\Theta}_i)/M$ and variance $1/M^2$.

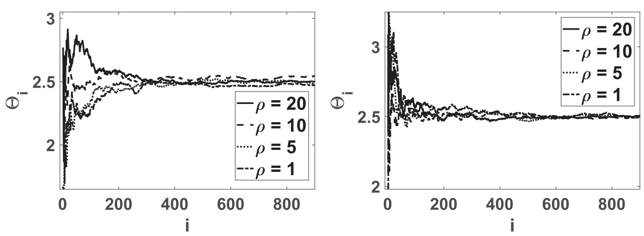
4.2 Impact of Non-Linearity on Convergence

Now we study the impact of non-linearity on the convergence of Θ_i. Throughout this section, we set $s = v = 0.5$ and $\eta = 0.5$. Fig. 5 shows the curve of Θ_i under different selections of ρ. One can observe that as ρ varies from 1 to 20, Θ_i converges after a similar number of ratings i. Recall that Fig. 1 illustrates that $\rho = 1$ is nearly the same as a linear form of f. This implies that the non-linearity of f influences the convergence speed of Θ_i only slightly compared with its linear form.



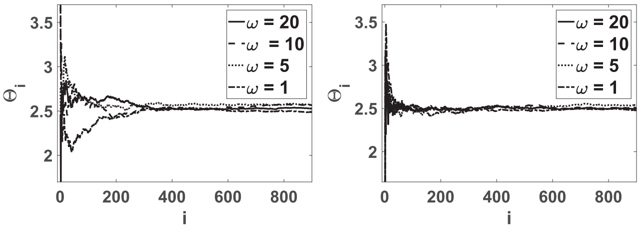
(a) $s = 0.4$ (b) $s = 0.6$

Fig. 2. Illustrating $g(\Theta_i)$, where $\theta = 2.5$.



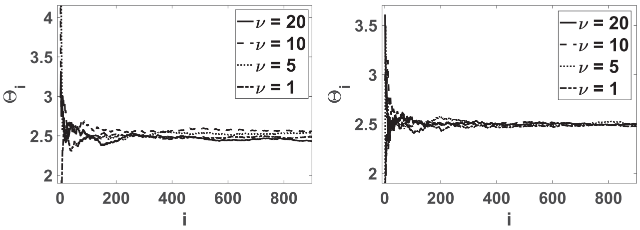
(a) $(\omega, \nu, \tau) = (1, 1, 0.1)$ (b) $(\omega, \nu, \tau) = (10, 10, 2)$

Fig. 5. Impact of ρ on the convergence, where $s = v = 0.5, \eta = 0.5$.



(a) $(\rho, \nu, \tau) = (1, 1, 0.1)$ (b) $(\rho, \nu, \tau) = (10, 10, 2)$

Fig. 6. Impact of ω on the convergence, where $s = v = 0.5, \eta = 0.5$.



(a) $(\rho, \omega, \tau) = (1, 1, 0.1)$ (b) $(\rho, \omega, \tau) = (10, 10, 2)$

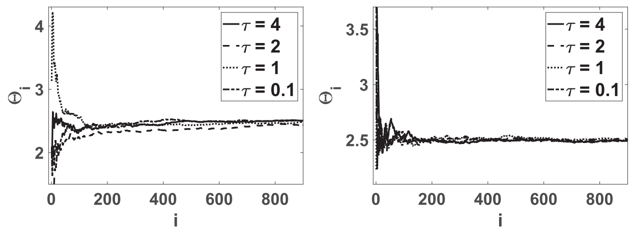
Fig. 7. Impact of ν on the convergence, where $s = v = 0.5, \eta = 0.5$.

Fig. 6 shows the impact of ω (i.e., the parameter that quantifies the sensitivity of the skill requirement of the selected tasks to skill self-assessment) on the convergence speed of Θ_i. From Fig. 6, one can observe that as ω varies from 1 to 20, Θ_i converges after a similar number of ratings i, implying that the non-linearity of g influences the convergence speed of Θ_i only slightly compared with its linear form.

Fig. 7 shows the impact of ν (i.e., the parameter that quantifies the sensitivity of the skill requirement of the selected tasks to skill self-assessment) on the convergence speed of Θ_i. From Fig. 7, one can observe that as ν varies from 1 to 20, the Θ_i converges after a similar number of ratings i, implying that the non-linearity of h influences the convergence speed of Θ_i only slightly as compared with its linear form.

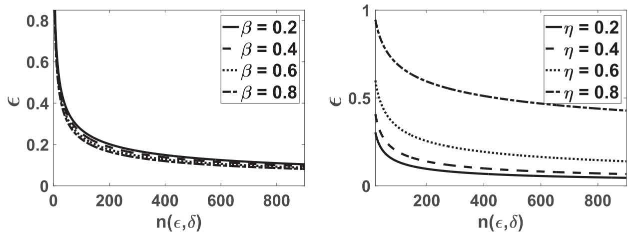
Fig. 8 shows the impact of τ (i.e., the parameter that quantifies the sensitivity of rating bias to Θ_i and $\tilde{\Theta}_i$) on the convergence speed of Θ_i. From Fig. 8, one can observe that as τ varies from 0.1 to 4, Θ_i converges after a similar number of ratings i, implying that the non-linearity of q influences the convergence speed of Θ_i only slightly as compared with its linear form.

Lessons Learned. The non-linearity of f, g, h and q influences the convergence speed of Θ_i only slightly as compared with their linear form.



(a) $(\rho, \omega, \nu) = (1, 1, 1)$ (b) $(\rho, \omega, \nu) = (10, 10, 10)$

Fig. 8. Impact of ν on the convergence, where $s = v = 0.5, \eta = 0.5$.



(a) $(\gamma, \eta) = (0.5, 0.5)$ (b) $(\beta, \gamma) = (0.5, 0.5)$

Fig. 9. Impact of β and η on the convergence rate, where $s = v = 0.5$.

4.3 Speed of Revealing Ground Truth Reliability

We further study the speed of revealing the ground truth reliability, via selecting linear forms of f, g, h and q derived in Equations (4), (12), (15), and (21) respectively.

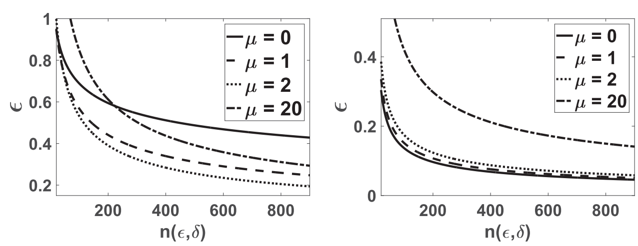
Fig. 9a shows the impact β on minimum number of ratings $n(\epsilon, \delta)$. One can observe that as β, i.e., the parameter that quantifies the sensitivity of skill self-assessment to aggregate rating, varies from 0.2 to 0.8, the minimum number of ratings decreases slightly. This statement also holds for γ, i.e., the parameter that quantifies the sensitivity of skill requirement of selected tasks to skill self-assessment, because γ and β are symmetric as illustrated in Equation (19). Fig. 9b shows that as the strength of rating bias η increases, on the minimum number of ratings $n(\epsilon, \delta)$ increases significantly.

Fig. 10 shows the impact of rating aggregation rules, i.e., μ, on the minimum number of ratings. From Fig. 10a, one can observe that the minimum number of ratings under $\mu = 2$ is smaller than that under $\mu = 0$. This implies that the averaging scoring rule is *not optimal*, but instead a recency aware rule can improve it. Furthermore, one can also observe that the improvement is significant, especially when the rating bias is strong and the accuracy parameter ϵ is small. Fig. 10 shows that when the strength of rating bias η is small, the average scoring rule has nearly the highest accuracy under the same number of ratings.

Lessons Learned. The minimum number of ratings varies slightly (or significantly) as the cognitive bias of workers (or the rating bias) varies from small to large. The unweighted averaging scoring rule is sub-optimal in terms of estimation accuracy, while a recency aware rating aggregation rule can significantly improve the accuracy under the same number of ratings.

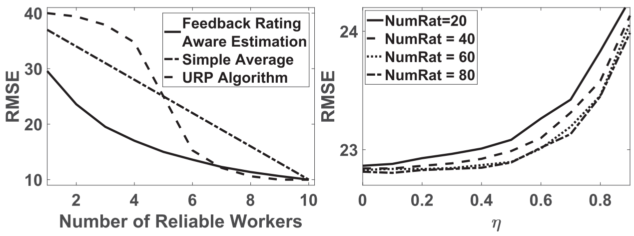
4.4 Applications to AQI Estimation

Application Settings. We now apply our feedback rating approach to AQI (air quality index) estimation. The AQI is



(a) $\eta = 0.8$ (b) $\eta = 0.2$

Fig. 10. Impact of rating aggregation rules on the convergence rate, where $s = v = 0.5, \beta = 0.5, \gamma = 0.5$.



(a) $\eta = 0.8, NumRat = 50$ (b) Two reliable workers

Fig. 11. RMSE of three estimation schemes.

measured in the range $[0, 100]$ and consider a task with a ground truth AQI of $x^* = 40$. A set of $\mathcal{W} \triangleq \{1, \ldots, W\}$ workers report AQI data. Specifically, let $X_w \in [0, 100]$ denote the AQI reported by worker w, where $X_w/100$ follows the standard beta distribution. The worker are classified into two types: reliable worker having $(\mathbb{E}[X_w], \sqrt{Var[X_w]}) = (50, 10)$ and unreliable worker having $(\mathbb{E}[X_w], \sqrt{Var[X_w]}) = (80, 10)$. Let Θ_{w,i_w} denote the aggregate rating of worker w, where i_w denotes the number of ratings of w. Let $\widehat{X}$ denote an estimate of the ground truth AQI. We use a *feedback rating aware estimation* scheme

$$\widehat{X} = \frac{\sum_{w \in \mathcal{W}} \Theta^5_{w,i_w} X_w}{\sum_{w \in \mathcal{W}} \Theta^5_{w,i_w}}.$$

We compare our estimation scheme with two baselines: (1) the *simple average* scheme $\widehat{X} = (\sum_{w \in \mathcal{W}} X_w)/W$ and (2) the state-of-the-art truth estimation scheme called *URP algorithm* [3]. We use the root mean square error as the performance metric

$$RMSE \triangleq \sqrt{\mathbb{E}[(\widehat{X} - x^*)^2]}.$$

We compute this metric via Monte Carlo simulation with 1000 rounds of repetitions. We consider $W = 10$ workers and for simplicity of discussion, we set $i_w = NumRat, \forall w \in \mathcal{W}$. We aim to demonstrate the benefit of feedback ratings in improving the AQI estimation. To achieve this, we start from the stage that the AQI sensing task has been assigned to some reliable workers and unreliable workers, i.e., we omit the process of task assignment or selection. We vary the fraction of unreliable workers who report the AQI from small to large. Furthermore, the historical ratings of each participating worker are generated by our model under the same settings of Section 4.3, where we set $(s, v) = (0.7, 0.7)$ for reliable workers and $(s, v) = (0.3, 0.7)$ for unreliable workers, and the variance of ratings is set to $(2/M)^2$. We aim to provide fundamental understandings on the rating bias by varying η.

AQI Estimation Accuracy. Fig. 11a shows the RMSE of three ground truth estimation schemes, where we vary the number of reliable workers from 1, representing that only 10 percent of the AQI are reported by reliable workers, to 10, representing that all AQI are reported from reliable workers. One can observe that when the number of reliable workers is no more than 4, i.e., at most 40 percent AQI are reported reliable workers, our feedback rating aware estimation scheme reduces the RMSE of URP algorithm by as high as $(34.7 - 17.01)/34.7 = 51\%$ (achieved when the number of reliable workers is 4). Namely, our feedback rating aware estimation scheme significantly outperforms the URP algorithm in the applications scenario that only a small fraction of data is reliable. When the number of reliable workers is no less than 8, i.e., at least 80 percent AQI are reported reliable workers, the URP algorithm reduces the RMSE of our feedback rating aware estimation scheme by at most $(11.37 - 10.60)/11.37 = 7\%$ (achieved when the number of reliable workers is 8). This implies that the URP algorithm can slightly outperform our feedback rating aware estimation scheme in the applications scenario in which a large fraction of data is reliable. Furthermore, our feedback rating aware estimation scheme always outperform the simple average scheme. Note that our feedback rating aware estimation scheme achieves the above improvement under a large rating bias $\eta = 0.8$. Fig. 11b shows that the RMSE of our feedback rating estimation scheme decreases when the rating bias η decreases. Namely, our feedback rating aware estimation scheme is more accurate if the crowdsensing platform operator can deploy a feedback rating system with small rating bias η. From Fig. 11b, one can observe that as we increase the number of ratings from 20 to 80, we decrease the RMSE slightly, implying that 20 ratings can already guarantee a small RMSE, showing that our scheme is highly efficient and robust.

Lessons Learned. Our feedback rating aware estimation scheme significantly outperforms (i.e., reduce the RMSE by as high as 50 percent) the URP algorithm in the applications scenario that only a small fraction of data is reliable. Our feedback rating aware estimation scheme is highly robust and efficient.

5 Related Work

Worker reliability has been applied in many aspects of crowdsensing applications. One important aspect is incentive mechanism design [6], [18], [19], [20], [21]. A variety of incentive mechanisms take the reliability of workers as input parameters [7], [12], [13]. Our work complements them by providing an approach to track the reliability scores of workers. Another important aspect is task matching or allocation [3], [22]. Yang *et al.* [3] showed that incorporating worker reliability into task matching can improve the matching accuracy significantly. Halabi *et al.* [4] designed a task assignment mechanism for vehicular crowdsensing applications, which assigns tasks based on vehicles' reliability. Our work supports task matching or assignment by providing explicit reliability scores on workers.

Worker reliability is essential for estimate the true information (called truth discovery) in crowdsensing applications. A number of works focused on estimating the true information and worker reliability jointly from the data submitted by workers. One class of algorithms can estimate worker reliability with proven guarantee on the estimation error [23], [24], [25]. These algorithms are batch learning algorithms, and Wang *et al.* [26] proposed a recursive expectation-maximization algorithm to accommodate streaming data scenario. In these works [23], [24], [25], [26], the data collected by workers is assumed to be either boolean or discrete and only of one dimension. Extending them to continuous or multiple dimensional data setting is technically non-trivial. Our feedback rating approach only requires requesters to assign ratings to assess the quality of data. Our work can be applied to discrete, continuous or even multi-dimensional data, as there are various sensor quality estimation algorithms [5], [6] that can be applied to discrete, continuous or even multi-dimensional data. Another class of approaches were inspired by the truth discovery framework [2], [3], [5], [27], [28], [29]. The intuition is that the data collected by reliable workers should be close to the true information and the worker whose data is close to the true information should have a high reliability score, or the graphical model [30]. This class of approaches have no theoretical guarantee on the reliability estimation error, while our work has. Furthermore, all these approaches are centralized and require one to input the data of multiple tasks in a batch manner. Our approach can handle the data of each task distributedly and the data of different tasks can even be allowed to arrive in a stream manner. Lastly, our approach updated worker reliability dynamically, capturing that worker may make errors regardless of reliable ones or unreliable ones, while previous works treat worker reliability estimation as a one shot problem.

Two notions that are closely related to worker reliability are worker performance and worker consistency. Huang *et al.* [31] conducted experiments to study the accuracy of peer consistency evaluation method in evaluating worker performance in human computation systems such as crowdsourcing. Qiu *et al.* [32] proposed to combine gold standard evaluation and peer consistency evaluation to measure worker performance. Williams *et al.* [10] proposed a metric to quantify the consistency of workers, i.e., the ability that a worker output the same result when repeating the same task. They proposed a mechanism that generates tasks with consistency probes to estimate worker consistency. Alqershi *et al.* [11] proposed a metric that evaluates worker consistency via pairwise comparisons. These worker performance methods and worker consistency methods are all one shot. They have the potential to be applied to estimate sensor data quality. Our approach aims to evaluate the long term or average reliability of workers.

A number of works designed rating mechanisms to address incentive issues in crowdsensing applications. Qiu *et al.* [33] designed a rating mechanism to incentivize requesters to allocate tasks to less skilled workers. Through this, less skilled workers can have chance to solve tasks and at the same time improve their skills, which in turn improves sustainability of crowdsourcing platforms. Lu *et al.* [34], [35] and Xie *et al.* [8] designed rating mechanisms to incentivize participation from workers and high quality contribution from workers. Different from these works, this paper focuses on designing rating systems to reveal worker reliability.

A variety of works studied sensor data quality estimation. Yang *et al.* [5] proposed a clustering based approach to estimate sensor data quality. Peng *et al.* [6] designed an expectation-maximization algorithm to estimate sensor data quality. These works provide alternatives to implement the rating protocol in our approach. Note that our work does not restrict to any specific rating protocols. We develop a general model to capture the impact of data quality estimation accuracy on the feedback ratings and our model allows errors in the estimation.

6 Conclusion

This paper utilizes a feedback rating approach to estimate worker reliability. We develop a mathematical model to characterize the cognitive bias in task selection and the rating bias. We derive sufficient conditions, under which the aggregate rating is asymptotically accurate in estimating worker reliability, and identify a class of asymptotically accurate rating aggregation rules. We further derive the minimum number of ratings needed to guarantee a given reliability estimation accuracy. We conduct experiments and found that: (1) A recency aware rating aggregation rule can significantly improve the accuracy of the average scoring rule in estimating worker reliability; (2) To achieve the same reliability estimation accuracy, the minimum number of ratings needed *increases significantly* as the degree of rating bias increases. (3) Our feedback rating approach improves AQI estimation accuracy by as high as 50 percent over the URP algorithm[3] when the sensor data is noisy.

Acknowledgments

The work of Hong Xie was supported in part by the National Natural Science Foundation of China under Grant 61902042 and in part by Chongqing Natural Science Foundation under Grant cstc2020jcyj-msxmX0652. The work of John C.S. Lui was supported by RGC's RIF under Grant R4032-18.

References

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

[2] C. Meng *et al.*, "Truth discovery on crowd sensing of correlated entities," in *Proc. ACM Conf Embedded Networked Sensor Syst.*, 2015, pp. 169–182.

[3] S. Yang, K. Han, Z. Zheng, S. Tang, and F. Wu, "Towards personalized task matching in mobile crowdsensing via fine-grained user profiling," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2018, pp. 2411–2419.

[4] T. Halabi and M. Zulkernine, "Reliability-driven task assignment in vehicular crowdsourcing: A matching game," in *Proc. Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. Workshops.*, 2019, pp. 78–85.

[5] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, and G. Chen, "On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 832–847, Apr. 2017.

[6] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 177–186.

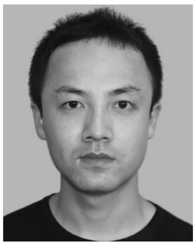
[7] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2016, pp. 341–350.

[8] H. Xie and J. C. Lui, "Incentive mechanism and rating system design for crowdsourcing systems: Analysis, tradeoffs and inference," *IEEE Trans. Serv. Comput.*, vol. 11, no. 1, pp. 90–102, Jan./Feb. 2016.

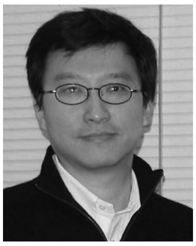
[9] C. Eickhoff, "Cognitive biases in crowdsourcing," in *Proc. Eleventh ACM Int. Conf. Web Search Data Mining*, 2018, pp. 162–170.

[10] A. Williams *et al.*, "Deja vu: Characterizing worker reliability using task consistency," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, 2017, pp. 197–205.

[11] F. Alqershi, M. Al-Qurishi, M. S. Aksoy, M. Alrubaian, and M. Imran, "A robust consistency model of crowd workers in text labeling tasks," *IEEE Access*, vol. 8, pp. 168381–168393, 2020.
[12] H. Jin, L. Su, and K. Nahrstedt, "Centurion: Incentivizing multi-requester mobile crowd sensing," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2017, pp. 1–9.
[13] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2016, pp. 344–353.
[14] C. I. Hovland, I. L. Janis, and H. H. Kelley, "Communication and persuasion; psychological studies of opinion change," London, U.K.: Yale Univ. Press, 1953.
[15] W. Wood, "Attitude change: Persuasion and social influence," *Annu. Rev. Psychol.*, vol. 51, no. 1, pp. 539–570, 2000.
[16] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, 2009, Art. no. 4.
[17] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. New York, NY, USA: Springer, 2003.
[18] K. Han, H. Huang, and J. Luo, "Posted pricing for robust crowdsensing," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2016, pp. 261–270.
[19] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 167–176.
[20] H. Zhang, B. Liu, H. Susanto, G. Xue, and T. Sun, "Incentive mechanism for proximity-based mobile crowd service systems," in *Proc. IEEE INFOCOM Annu. Int. Conf. Comput. Commun.*, 2016, pp. 1–9.
[21] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2014, pp. 1213–1221.
[22] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *Proc. IEEE INFOCOM Conf. Comput. Commun.* 2014, pp. 745–753.
[23] T. Bonald and R. Combes, "A minimax optimal algorithm for crowdsourcing," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2017, pp. 4355—4363
[24] Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet EM: A provably optimal algorithm for crowdsourcing," *J Mach. Learn. Res.*, vol. 17, no. 102, pp. 1–44, 2016.
[25] C. Gao and D. Zhou, "Minimax optimal convergence rates for estimating ground truth from crowdsourced labels," 2013, *arXiv:1310.5764*.
[26] D. Wang, T. Abdelzaher, L. Kaplan, and C. C. Aggarwal, "Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2013, pp. 530–539.
[27] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: A maximum likelihood estimation approach," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, 2012, pp. 233–244.
[28] L. Su *et al.*, "Generalized decision aggregation in distributed sensing systems," in *Proc. IEEE Real-Time Syst. Symp.*, 2014, pp. 1–10.
[29] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2014, pp. 1187–1198.
[30] F. Ma *et al.*, "Faitcrowd: Fine grained truth discovery for crowdsourced data aggregation," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2015, pp. 745–754.
[31] S.-W. Huang and W.-T. Fu, "Enhancing reliability using peer consistency evaluation in human computation," in *Proc. Conf. Comput. Supported Cooperative Work*, 2013, pp. 639–648.
[32] C. Qiu, A. Squicciarini, D. R. Khare, B. Carminati, and J. Caverlee, "Crowdeval: A cost-efficient strategy to evaluate crowdsourced worker's reliability," in *Proc. Int. Conf. Auton. Agents MultiAgent Syst.*, 2018, pp. 1486–1494.
[33] C. Qiu, A. Squicciarini, and S. Rajtmajer, "Rating mechanisms for sustainability of crowdsourcing platforms," in *Proc. ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2003–2012.
[34] J. Lu, Y. Xin, Z. Zhang, F. Wu, and J. Han, "Online rating protocol using endogenous and incremental learning design for mobile crowdsensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3190–3201, Mar. 2020.
[35] J. Lu *et al.*, "Multi-level two-sided rating protocol design for service exchange contest dilemma in crowdsensing," *IEEE Access*, vol. 7, pp. 78391–78405, 2019.



Hong Xie (Member, IEEE) received the BEng degree from the School of Computer Science and Technology, University of Science and Technology of China, in 2010, and the PhD degree, under the supervision of Prof. John. C.S. Lui, from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, in 2015. He is currently a research professor with the College of Computer Science, Chongqing University. His research interests include online learning, network economics, and data science. He is a member of ACM.



John C.S. Lui (Fellow, IEEE) received the PhD degree in computer science from the University of California at Los Angeles. He is currently the Choh-Ming Li chair professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong. His research interests include machine learning, online learning (e.g., multiarmed bandit and reinforcement learning), network science, future Internet architectures and protocols, network economics, network/system security, large scale storage systems. He is an elected member of the IFIP WG 7.3, fellow of ACM, senior research fellow of the Croucher Foundation, and was the past chair of the ACM SIGMETRICS in 2011–2015). He was the recipient of various departmental teaching awards and the CUHK Vice-Chancellor's Exemplary Teaching Award. He was the co-recipient of the Best Paper Award in the IFIP WG 7.3 Performance 2005, IEEE/IFIP NOMS 2006, SIMPLEX 2013, and ACM RecSys 2017.