Multi-object Classification via Crowdsourcing with a Reject Option

Qunwei Li, *Student Member, IEEE*, Aditya Vempaty, *Member, IEEE*, Lav R. Varshney, *Senior Member, IEEE*, and Pramod K. Varshney, *Fellow, IEEE*

Abstract—Consider designing an effective crowdsourcing system for M-ary classification where crowd workers complete simple binary microtasks which are aggregated to give the final result. We consider the novel scenario where workers have a reject option so they may skip microtasks they are unable or unwilling to do. For example in mismatched speech transcription, workers who do not know the language may be unable to respond to microtasks in phonological dimensions outside their categorical perception. We present an aggregation approach using a weighted majority voting rule, where each worker's response is assigned an optimized weight to maximize the crowd's classification performance. We evaluate system performance in both exact and asymptotic forms. Further, we consider the setting where there may be a set of greedy workers that complete microtasks even when they are unable to perform it reliably. We consider an oblivious and an expurgation strategy to deal with greedy workers, developing an algorithm to adaptively switch between the two based on the estimated fraction of greedy workers in the anonymous crowd. Simulation results show improved performance compared with conventional majority voting.

Index Terms—Classification, crowdsourcing, distributed inference, information fusion, reject option

I. INTRODUCTION

OPTIMIZATION problems in communication, computation, and sensing systems have driven signal processing advances over the last several decades. Now, engineered social systems such as crowdsensing, crowdsourcing, and social production are becoming prevalent for all these informational tasks. Yet, they are often constructed in an ad hoc manner. Advances in signal processing theory and methods to optimize these novel human-oriented approaches to signal acquisition and processing are needed [1].

Crowdsourcing has particularly attracted intense interest [2]–[8] as a new paradigm for signal processing tasks such as handwriting recognition, paraphrase acquisition, speech transcription, image quality assessment, and photo tagging [9]–[16], which are all essentially inference problems of M-ary classification. Unfortunately, the low quality of crowdsourced output remains a key challenge [17]–[19].

Q. Li and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (e-mail: qli33@syr.edu; varshney@syr.edu).

A. Vempaty was with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: avempat@us.ibm.com).

L. R. Varshney is with the Department of Electrical and Computer Engineering and with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: varshney@illinois.edu).
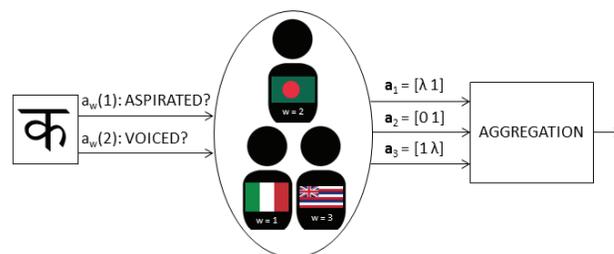
Fig. 1. An illustrative example of the proposed crowdsourcing framework.

Low-quality work may arise not only because workers are insufficiently motivated to perform well [20], [21], but also because workers lack the skills to perform the task that is posed to them [22]. Decomposing larger tasks into smaller subtasks for later aggregation allows workers lacking certain skills to contribute useful information [22], [23], by polarizing the lack of skill into a subset of subtasks.

As an illustrative example of lack of skill, consider the problem of mismatched crowdsourcing for speech transcription, which has garnered interest in the signal processing community [24]–[30]. The basic idea is to use crowd workers to transcribe a language they do not speak, into nonsense text in their own native language orthography. Certain phonological dimensions, such as aspiration or voicing, are used to differentiate phonemes in one language but not others [26]. Moreover due to categorical perception acquired in childhood, workers lacking a given phonological dimension in their native language may be unable to make relevant distinctions. That is, they lack the skill for the task.

Fig. 1 depicts the task of language transcription of Hindi. Suppose the four possibilities for a velar stop consonant to transcribe are $R = \{$ क, ख, ग, घ $\}$. The binary query of "aspirated or unaspirated" differentiates between $\{$ ख, घ $\}$ and $\{$ क, ग $\}$, whereas the query of "voice or unvoiced" differentiates between $\{$ ग, घ $\}$ and $\{$ क, ख $\}$. Note the two queries are independent. Now suppose the first worker is a native Italian speaker. Since Italian does not use aspiration, this worker will be unable to differentiate between क and ख, or between ग and घ. It would be useful if this worker specified the inability to perform the task through a special symbol λ, rather than guessing randomly. Suppose the second worker is a native Bengali speaker; since this language makes a four-way distinction among velar stops, such a worker will probably answer both questions without a λ. Now suppose the third worker is a native Hawaiian speaker; since this language

does not use voicing, such a worker will not be able to differentiate between क and ग, or between ख and घ. Hence, it would be useful if this worker answered λ for the question of differentiating among these two subchoices.

The present paper allows workers to not respond, i.e. allowing a *reject option*, as in the example. Research in psychology suggests a greater tendency to select the reject option when the choice set offers several attractive alternatives but none that can be easily justified as the best, resulting in less arbitrary decisions [31]. The reject option has previously been considered in the machine learning and signal processing literatures [32]–[36], but we specifically consider worker behavior and aggregation rules for crowdsourcing with a reject option. To characterize performance, we derive a closed-form expression for the probability of a microtask being correct, together with the asymptotic performance when the crowd size is large.

Several methods have been proposed to deal with noisy crowd work when crowd workers are required to provide a definitive yes/no response [21]–[23], [37]–[41], rather than allowing a reject option. Without the reject option, noisy responses to tasks cannot be tagged before aggregation so appropriate weights cannot be assigned [37]. For instance, the popular majority voting rule weights all answers equally [42], though new weighted aggregation rules have also been developed [37], [39]. We employed error-control codes and decoding algorithms to design reliable crowdsourcing systems with unreliable workers [22]. A group control mechanism where worker reputation is used to partition the crowd into groups is presented in [40], [41]; comparing group control to majority voting indicates majority voting is more cost-effective on less complex tasks [21].

Note that under typical crowdsourcing incentive schemes based on work volume, workers may respond to tasks for which they are not sufficiently skilled, even when a reject option is available. We therefore also consider the case where a greedy fraction of the anonymous crowd workers complete all microtasks with random guesses to maximize their rewards.

The main contributions of this paper are threefold. First, we study a new crowdsourcing paradigm where similarly-difficult microtasks are performed by workers having a reject option so they can skip microtasks if they cannot respond reliably. We assign a weight for every worker's response based on the number of skipped microtasks, yielding significant performance improvement when compared to when all workers must respond. We also provide asymptotic performance analysis as crowd size increases (worker pools are typically large in crowdsourcing). Next, we consider the setting where a fraction of greedy crowd workers respond to all tasks, first continuing to employ the oblivious strategy of weighting by response number and investigating the error performance as a function of the fraction of greedy workers in the crowd. We then study another strategy, an expurgation strategy, where responses of workers that respond to all microtasks are discarded and the remaining workers' responses are assigned optimized weights. Finally, an adaptive algorithm to switch between oblivious and expurgation strategies is derived, based on estimating several crowd parameters such as fraction of greedy workers.

Although the contributions listed above are stated for the crowdsourcing paradigm, our results hold for other signal classification tasks when decisions are made using signals that are quite uncertain. This is known as classification with a reject option [43] and has been focus of several recent studies in signal processing research including pattern recognition, image, and speech classification [33], [34], [36], [44], [45].

II. CLASSIFICATION TASK FOR CROWDS

Consider W workers taking part in an M-ary classification task. Each worker is asked N simple binary questions, termed microtasks, which eventually lead to a decision among the M classes. We assume it is possible to construct $N = \lceil \log_2 M \rceil$ independent microtasks of equal difficulty. Such microtasks can be designed using taxonomy and dichotomous keys [46]. The workers submit results that are combined to give the final decision. A worker's answer to a single microtask is represented by either "1" (Yes) or "0" (No) [22] and so the wth worker's ordered answers $\mathbf{a}_w(i)$, $i \in \{1, 2, \ldots, N\}$ to all microtasks form an N-bit word, denoted $\mathbf{a}_w$.

Each worker has the reject option of skipping microtasks; we denote a skipped answer as λ, whereas "1/0" (Yes/No) responses are termed *definitive answers*. Due to variability in worker expertise, the probability of submitting definitive answers is different for each worker. Let $p_{w,i}$ represent the probability the wth worker submits λ for the ith microtask. Similarly, let $r_{w,i}$ be the probability $\mathbf{a}_w(i)$, the ith answer of the wth worker, is correct given a definitive answer has been submitted. Due to worker anonymity, we study performance when $p_{w,i}$ and $r_{w,i}$ are realizations of certain probability distributions, denoted $F_P(p)$ and $F_R(r)$ with corresponding means m and μ, respectively. Let H_0 and H_1 be hypotheses of bits "0" and "1" for a single microtask, respectively. For simplicity of performance analysis, "0" and "1" are assumed equiprobable for every microtask. The crowdsourcing platform or fusion center (FC) collects the N-bit words from W workers and aggregates results, as discussed next.

A. Weighted Majority Voting

We first investigate weighted majority voting as the fusion rule for classification. Consider all object classes as elements in the set $D = \{e_j\}$, $j = 1, \ldots, M$, where e_j represents the jth class. As indicated earlier, a worker's definitive responses to the microtasks determine a subset in the original set D, consisting of multiple elements/classes. If all responses from the crowd are definitive, the final subsets are singletons and a single class is identified as the object class. Since some microtasks may be answered with a response λ, the resulting subsets may not be singletons and each element of the same corresponding subsets will be chosen equiprobably as the classification decision. Let us denote the subset determined by the definitive answers of the wth worker as $D_w \in D$. The task manager assigns the same weight to all elements in D_w based on the wth worker's answer. With the submitted answers from W workers, we determine the overall weight assigned to any jth element of D as

$$\mathbb{W}(e_j) = \sum_{w=1}^{W} W_w I_{D_w} \langle e_j \rangle, j = 1, 2, \ldots, M, \quad D_w \in D, \quad (1)$$

where W_w is the weight[1] assigned to D_w, and $I_{D_w} \langle e_j \rangle$ is an indicator function which equals 1 if $e_j \in D_w$ and 0 otherwise. Then the element e_D with the highest weight is selected as the final class, and the classification rule is stated as

$$e_D = \arg \max_{e_j \in D} \left\{ \mathbb{W} \left(e_j \right) \right\}. \tag{2}$$

For tie-breaking, randomly choose from the classes with the same weight. Notice that conventional majority voting requires full completion of all microtasks without the reject option and has identical W_w for each worker's decision.

Next, we show how the problem formulated in (2) can be further decomposed.

Proposition 1: Classification rule (2) is equivalent to bit-by-bit decision since the ith bit, $i = 1, \ldots, N$, is decided by

$$\sum_{w=1}^{W} W_w I_1 \langle i, w \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \sum_{w=1}^{W} W_w I_0 \langle i, w \rangle, \tag{3}$$

where $I_s \langle i, w \rangle$, $s \in \{0, 1\}$, is the indicator function which is 1 if the wth worker's answer to the ith bit is "s", otherwise $I_s \langle i, w \rangle = 0$. For tie-breaking, randomly choose 0 or 1.

Proof: The class e_D corresponds to a unique N-bit word. Thus, if the ith bit of the N-bit word corresponding to the class e_D is equal to s, s has the same weight as assigned to e_D, which is greater than or equal to the symbol $1 - s$. ∎

B. Optimal Bit-by-bit Bayesian Aggregation

Let $\mathcal{A}(i) = [\mathbf{a}_1(i), \mathbf{a}_2(i), \ldots, \mathbf{a}_W(i)]$ denote all the answers to ith microtask collected from the crowd. For the binary hypothesis testing problem corresponding to the ith bit of the N-bit word, the log-likelihood ratio test is

$$\log \frac{P\left(H_1 | \mathcal{A}(i)\right)}{P\left(H_0 | \mathcal{A}(i)\right)} \underset{H_0}{\overset{H_1}{\gtrless}} 0, \tag{4}$$

where $P\left(\cdot\right)$ denotes the probability density function. We can express the likelihood ratio as

$$\frac{P\left(H_1 | \mathcal{A}(i)\right)}{P\left(H_0 | \mathcal{A}(i)\right)} = \frac{\prod\limits_{w=1}^{W} P\left(\mathbf{a}_w(i) | H_1\right)}{\prod\limits_{w=1}^{W} P\left(\mathbf{a}_w(i) | H_0\right)}$$
$$= \frac{\prod\limits_{S_1}(1 - p_{w,i}) r_{w,i} \prod\limits_{S_0}(1 - p_{w,i})(1 - r_{w,i}) \prod\limits_{S_\lambda} p_{w,i}}{\prod\limits_{S_1}(1 - p_{w,i})(1 - r_{w,i}) \prod\limits_{S_0}(1 - p_{w,i}) r_{w,i} \prod\limits_{S_\lambda} p_{w,i}}, \tag{5}$$

where S_1 is the set of w such that $\mathbf{a}_w(i) = 1$, S_0 is the set of w such that $\mathbf{a}_w(i) = 0$ and S_λ is the set of w such that $\mathbf{a}_w(i) = \lambda$, respectively. Then, it is straightforward to show that the test for the decision on the ith bit is

$$\sum_{w \in S_1} \log \frac{r_{w,i}}{1 - r_{w,i}} \underset{H_0}{\overset{H_1}{\gtrless}} \sum_{w \in S_0} \log \frac{r_{w,i}}{1 - r_{w,i}}. \tag{6}$$

Note that the optimal Bayesian criterion can also be viewed as the general weighted majority voting rule in (3) with weight $W_w = \log \frac{r_{w,i}}{1 - r_{w,i}}$, also called the Chair-Varshney rule [47]. Note that (3) represents majority voting when $W_w = 1$.

[1]The assignment of these weights will be discussed later in the paper.

However, this optimal Bayesian criterion can only be used if $r_{w,i}$ for every worker is known *a priori*, which is usually difficult to estimate for anonymous crowds just from submitted answers. The difficulty in obtaining prior information makes the simple majority voting scheme very effective and therefore widely used [42]. We show later that our proposed method—which need not estimate $r_{w,i}$ but only its mean μ—can outperform conventional majority voting.

C. Class-based Aggregation Rule

For the general weighted majority voting scheme where e_C is the correct class, the probability of misclassification is

$$P_m = \Pr\left(e_D \neq e_C\right)$$
$$= \Pr\left(\arg \max_{e_j \in D} \left\{ \sum_{w=1}^{W} W_w I_{D_w} \langle e_j \rangle \right\} \neq e_C \right)$$
$$= 1 - \Pr\left(\arg \max_{e_j \in D} \left\{ \sum_{w=1}^{W} W_w I_{D_w} \langle e_j \rangle \right\} = e_C \right), \tag{7}$$

where $\Pr(\mathcal{E})$ is the occurrence probability of event $\mathcal{E}$.

A closed-form expression for the error probability P_m cannot be derived without an explicit expression for W_w; hence it is difficult to determine the optimal weights to minimize P_m.

Consequently, we consider an optimization problem based on a different objective function and propose a novel weighted majority voting method that outperforms simple majority voting. Note that e_D is chosen as the decision for classification such that e_D has the maximum overall weight collected from all the workers. Thus, we maximize the average overall weight assigned to the correct class while the overall weight collected by all the elements remains the same as the other existing methods such as majority voting. We state the optimization problem over the weights as

$$\begin{aligned} \text{maximize} \quad & E_C\left[\mathbb{W}\right] \\ \text{subject to} \quad & E_O\left[\mathbb{W}\right] = K \end{aligned} \tag{8}$$

where $E_C\left[\mathbb{W}\right]$ is the crowd's average weight contribution to the correct class and $E_O\left[\mathbb{W}\right]$ is the average weight contribution to all possible classes. K is set to a constant so we are looking for a maximized portion of weight contribution to the correct class while the weight contribution to all classes remains fixed. This procedure ensures one cannot obtain greater $E_C\left[\mathbb{W}\right]$ by simply increasing the weight for each worker's answer, while K results in a normalized weight assignment scheme. If two weight assignment schemes share the same value of $E_O\left[\mathbb{W}\right]$, one can expect better performance by the scheme with higher $E_C\left[\mathbb{W}\right]$. Thus, K facilitates a relatively easier performance comparison of different weight assignment schemes.

III. System With Honest Crowd Workers

We first consider the case where the crowd is entirely composed of honest workers—workers that are not greedy and honestly observe, think, and answer microtasks posed to them, while skipping queries they are not confident about. The wth worker responds with a λ to the ith microtask with probability $p_{w,i}$. Next, we derive the optimal weight W_w for the wth worker in this case.

Proposition 2: To maximize the normalized average weight assigned to the correct classification element, the weight for wth worker's answer is given by $W_w = \mu^{-n}$, where $\mu = E[r_{w,i}]$ and n is the number of definitive answers that the wth worker submits.

Proof: See Appendix A. ∎

Note that when workers are forced to make a hard decision for every single bit, the weights derived above become identical. In general, the weight depends on the number of questions answered by a worker: if more questions are answered, the weight assigned to the corresponding worker's answer is larger. Assuming a worker's correct probability is greater than half if he/she gives a definitive answer, a larger number of definitive answers increases the chance the quality of the worker is higher than others. Increased weight can thereby emphasize the contribution of high-quality workers and improve overall classification performance.

A. Estimation of μ

Before the proposed aggregation rule can be used, note that μ has to be estimated to assign the weight for every worker's answers. Here, we give two approaches to estimate μ.

1) Training-based: In addition to the N microtasks, the task manager inserts additional questions to estimate the crowd's μ value. The answers to such "gold standard" questions are known to the manager [48], [49]. By checking the crowd worker's answers, μ can be estimated. Suppose the first T questions are training questions, and let $\bar{\mathbf{B}}$ be the T-bit correct answers to them. First, we calculate the ratio $r(w)$ as

$$r(w) = \sum_{i=1}^{T} \frac{I_{\bar{\mathbf{B}}(i)}\langle \mathbf{a}_w(i)\rangle}{I(w)}, \tag{9}$$

where $I_x\langle y\rangle$ is the indicator function which is 1 if $x = y$ and 0 otherwise, and $I(w) = \sum_{i=1}^{T}(I_1\langle i, w\rangle + I_0\langle i, w\rangle)$. In order to avoid the cases where some workers submit λ for all the training questions, we estimate μ as follows

$$\hat{\mu} = \frac{1}{W - \epsilon}\sum_{w=1}^{W} r(w), \tag{10}$$

where ϵ is the number of workers that submit all λ for the training questions and the corresponding $r(w)$ is set to 0.

2) Majority-voting based: We use majority voting to obtain the initial aggregation result and set it as the benchmark to estimate μ. First, all the answers $\mathbf{a}_w(i)$ are collected to obtain the benchmark $\mathcal{B}(i)$ by traditional majority voting, where $i = 1, \ldots, N$. Note that $\mathcal{B}(i)$ may contain λ since it is possible that all answers $\mathbf{a}_w(i)$ have λ at the same position. Then, for the wth worker, we calculate the ratio $r(w)$ as

$$r(w) = \sum_{i=1}^{N} \frac{I_{\mathcal{B}(i)}\langle \mathbf{a}_w(i)\rangle}{I(w)}, \tag{11}$$

where $I_\lambda\langle\lambda\rangle = 0$, and $I(w) = \sum_{i=1}^{N}(I_1\langle i, w\rangle + I_0\langle i, w\rangle)$. As before, we estimate μ as in (10), but where ϵ is the number of workers that submit λ for all microtasks.

B. Performance Analysis

In this subsection, we characterize performance of the proposed classification framework in terms of probability of correct classification P_c. Note that we have overall correct classification only when all the bits are classified correctly,[2] which also offers the lower bound in the general case where the microtasks are not completely independent of each other.

First, we restate the bit decision criterion in (3) as

$$\sum_{w=1}^{W} T_w \overset{H_1}{\underset{H_0}{\gtrless}} 0 \tag{12}$$

with $T_w = W_w(I_1\langle i, w\rangle - I_0\langle i, w\rangle)$, where the resulting $T_w \in \{-\mu^{-N}, -\mu^{-N+1}, \ldots, -\mu^{-1}, 0, \mu^1, \ldots, \mu^{N-1}, \mu^N\}$.

Proposition 3: For the ith bit, the probability mass function of T_w under hypothesis H_s, $\Pr(T_w|H_s)$, for $s \in \{0, 1\}$, is:

$$\Pr\left(T_w = I(-1)^{t+1}\mu^{-n}|H_s\right)$$
$$= \begin{cases} r_{w,i}^{1-|s-t|}(1 - r_{w,i})^{|s-t|}\varphi_n(w), & I = 1 \\ p_{w,i}, & I = 0 \end{cases},$$
$$t \in \{0, 1\}, n \in \{1, \ldots, N\}, \tag{13}$$

where $I = I_1\langle i, w\rangle + I_0\langle i, w\rangle$, $\varphi_n(w) = (1 - p_{w,i})\sum_C \prod_{\substack{j=1 \\ j\neq i}}^{N} p_{w,j}^{k_j}(1 - p_{w,j})^{1-k_j}$ and C is the set

$$C = \left\{\{k_1, k_2, \ldots, k_{i-1}, k_{i+1}, \ldots, k_N\} : \sum_{\substack{j=1 \\ j\neq i}}^{N} k_j = N - n\right\}$$

with $k_j \in \{0, 1\}$.

Proof: See Appendix B. ∎

Since hypotheses H_0 and H_1 are assumed equiprobable, the correct classification probability for the ith bit $P_{c,i}$ is

$$P_{c,i} = \frac{1 + P_{d,i} - P_{f,i}}{2}, \tag{14}$$

where $P_{d,i}$ is the probability of deciding the ith bit as "1" when the true bit is "1" and $P_{f,i}$ is the probability of deciding the ith bit as "1" when the true bit is "0".

Proposition 4: The probability of correct classification for the ith bit $P_{c,i}$ is

$$P_{c,i} = \frac{1}{2} + \frac{1}{2}\sum_S \binom{W}{\mathbb{Q}}(F_i(\mathbb{Q}) - F_i'(\mathbb{Q}))$$
$$+ \frac{1}{4}\sum_{S'}\binom{W}{\mathbb{Q}}(F_i(\mathbb{Q}) - F_i'(\mathbb{Q})) \tag{15}$$

with

$$F_i(\mathbb{Q}) = \prod_{w\in G_\lambda} p_{w,i}\prod_{w\in G_0}(1 - r_{w,i})\varphi_n(w)\prod_{w\in G_1} r_{w,i}\varphi_n(w)$$

and

$$F_i'(\mathbb{Q}) = \prod_{w\in G_\lambda} p_{w,i}\prod_{w\in G_1}(1 - r_{w,i})\varphi_n(w)\prod_{w\in G_0} r_{w,i}\varphi_n(w),$$

---

[2] When $N > \log_2 M$, the N-bit answer after aggregation may correspond to a class that does not exist; this is also misclassification.

where

$$\mathbb{Q} = \left\{ (q_{-N}, q_{-N+1}, \ldots q_N) : \sum_{n=-N}^{N} q_n = W \right\} \qquad (16)$$

with natural numbers q_n and q_0, G_0 denotes the worker group that submits "0" for ith microtask, G_1 the group that submits "1" and G_λ the group that submits λ, and

$$S = \left\{ \mathbb{Q} : \sum_{n=1}^{N} \mu^{-n} (q_n - q_{-n}) > 0 \right\}, \qquad (17)$$

$$S' = \left\{ \mathbb{Q} : \sum_{n=1}^{N} \mu^{-n} (q_n - q_{-n}) = 0 \right\}, \qquad (18)$$

and $\binom{W}{\mathbb{Q}} = W! / \prod_{n=-N}^{N} q_n!$.

Proof: See Appendix C. ∎

Proposition 5: The probability of correct classification P_c in the crowdsourcing system is

$$P_c = \left[\frac{1}{2} + \frac{1}{2} \sum_{S} \binom{W}{\mathbb{Q}} (F(\mathbb{Q}) - F'(\mathbb{Q})) \right. $$
$$\left. + \frac{1}{4} \sum_{S'} \binom{W}{\mathbb{Q}} (F(\mathbb{Q}) - F'(\mathbb{Q})) \right]^N, \qquad (19)$$

where

$$F(\mathbb{Q}) = m^{q_0} \prod_{n=1}^{N} (1-\mu)^{q_{-n}} \mu^{q_n} \left(C_{N-1}^{n-1} (1-m)^n m^{N-n} \right)^{q_{-n}+q_n} \qquad (20)$$

and

$$F'(\mathbb{Q}) = m^{q_0} \prod_{n=1}^{N} (1-\mu)^{q_n} \mu^{q_{-n}} \left(C_{N-1}^{n-1} (1-m)^n m^{N-n} \right)^{q_{-n}+q_n}, \qquad (21)$$

Proof: See Appendix D. ∎

In practice, the number of workers for crowdsourcing tasks is large (hundreds). Thus it is useful to investigate asymptotic system performance when W increases without bound.

Proposition 6: As the number of workers W approaches infinity, the probability of correct classification P_c is

$$P_c = \left[Q \left(-\frac{M}{\sqrt{V}} \right) \right]^N, \qquad (22)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{\frac{-t^2}{2}} dt$, and M and V are given as

$$M = \frac{W(2\mu-1)(1-m)}{\mu} \left(\frac{1}{\mu} - \left(\frac{1}{\mu} - 1 \right) m \right)^{N-1}, \qquad (23)$$

and

$$V = \frac{W(1-m)}{\mu^2} \left(\frac{1}{\mu^2} - \left(\frac{1}{\mu^2} - 1 \right) m \right)^{N-1} - \frac{M^2}{W}. \qquad (24)$$

Proof: See Appendix E. ∎

For large but finite crowds, the asymptotic result (22) is a good characterization of actual performance. Let us therefore consider (22) in more detail. First, we rewrite (22) as

$$P_c = \left[Q \left(-\sqrt{\frac{W}{\frac{1}{f(\mu,m)} - 1}} \right) \right]^N, \qquad (25)$$

where

$$f(\mu, m) = (1-m)(2\mu-1)^2 (g(\mu,m))^{N-1}, \qquad (26)$$

and

$$g(\mu, m) = \frac{(1-(1-\mu)m)^2}{1-(1-\mu^2)m}. \qquad (27)$$

Theorem 1: The correct classification probability of the system increases with increasing crowd size W.

Proof: Follows from (25) as the probability of correct classification increases monotonically with respect to W. ∎

Theorem 2: The correct classification of the system increases with increasing μ.

Proof: We take the partial derivative of $g(\mu, m)$ with respect to μ and obtain

$$\frac{\partial g}{\partial \mu} = \frac{2m(1-\mu)(1-m) \mathbf{A}}{\mathbf{B}^2}, \qquad (28)$$

where $\mathbf{A} = m\mu - m + 1$, and $\mathbf{B} = m\mu^2 - m + 1$.

Clearly $\frac{\partial g}{\partial \mu} > 0$. Recall (25), (26), and (27): a larger P_c results as μ increases. Then, the classification performance of the task in the crowdsourcing system also increases. ∎

To obtain the relation between crowd's performance in terms of P_c and m, we take the partial derivative of $f(\mu, m)$ with respect to m and obtain

$$\frac{1}{(2\mu-1)^2} \frac{\partial f}{\partial m}$$
$$= (N-1)(1-m) \left(\frac{\mathbf{A}^2(\mu^2-1)}{\mathbf{B}^2} + \frac{2\mathbf{A}(1-\mu)}{\mathbf{B}} \right) \left(\frac{\mathbf{A}^2}{\mathbf{B}} \right)^{N-2} - \left(\frac{\mathbf{A}^2}{\mathbf{B}} \right)^{N-1}.$$

After some mathematical manipulations, we observe:

- When $m > \frac{1}{1+\mu}$, we can guarantee that $\frac{\partial f}{\partial m} < 0$, which means that the crowd performs better as P_c increases with decreasing m.
- When $m < \frac{1}{1+\mu}$ and $N \geq \frac{(m\mu-m+1)^2}{(1-m)(1-\mu)^2(m\mu+m-1)} + 1$, we can guarantee that $\frac{\partial f}{\partial m} > 0$, which means that the crowd performs better as P_c increases with increasing m.

These two observations indicate that a larger probability of the crowd responding to the ith microtask with λ does not necessarily degrade crowd's performance in terms of the detection of the ith microtask.

This counterintuitive result follows since even though the crowd skips more microtasks, the optimized weight takes advantage of the number of unanswered questions and extracts more information. For this to happen, the number of microtasks N has to be greater than a lower limit. Since a larger N induces more diversity in the number of unanswered questions, the existence of the lower limit means that this diversity can actually benefit the performance using the proposed scheme.

C. Simulation Results

In this subsection, we compare the performance of the proposed crowdsourcing system where crowd workers are allowed to skip microtasks with the conventional majority voting method in a hard-decision fashion, which means workers are forced to make a decision even if the workers believe

that no definitive answers could be provided. The number of equiprobable classes is set as $M = 8$.

Fig. 2 compares performance when $W = 20$ workers take part in the task. We consider here that workers have a fixed $p_{w,i}$ for each microtask and $r_{w,i} = 0.8$. We observe that performance degrades as $p_{w,i}$ gets larger, i.e. the workers have a higher probability of not submitting an answer to the microtask. A remarkable performance improvement associated with our proposed approach is observed. The two curves converge at $p_{w,i} = 1$ with P_c being equal to 0.125. At this point, with the majority-based approach, each worker gives random answers for each microtask whereas workers using the proposed scheme skip all the questions and the tie-breaking criterion is used to pick a random bit for every microtask. In Fig. 3, we fix $p_{w,i} = 0.5$ and vary $r_{w,i}$ to compare the resulting P_c. Notable performance improvement is also seen. The point at $r_{w,i} = 0.5$ indicates that the worker is making a random guess even if he/she believes that he/she can complete the corresponding microtask correctly. The performance improves as $r_{w,i}$ gets larger, which means that the crowd is able to give higher-quality definitive answers.

In Fig. 4, we compare performance with different number of workers, also showing the asymptotic performance characterization. Here, we consider different qualities of the individuals in the crowd which is represented by variable $p_{w,i}$ with uniform distribution $U(0,1)$ and $r_{w,i}$ with $U(0.6,1)$. First, it is observed that a larger crowd completes the classification task with higher quality. The asymptotic curves are derived under the assumption of a very large crowd, which are the bounds on the performance of the systems. It is not difficult to derive the asymptotic performance for conventional majority voting:

$$P_c = \left[Q\left(-\sqrt{\frac{W^2(2l-1)}{4l - 4l^2}} \right) \right]^N,$$

where $l = \mu + m(0.5 - \mu)$. Therefore, the asymptotic gap in the performance between conventional majority voting and the proposed method can also be obtained. We find that the asymptotic curves are quite a tight match to the actual performance. Again, we can see a significant improvement in P_c brought by the proposed approach.
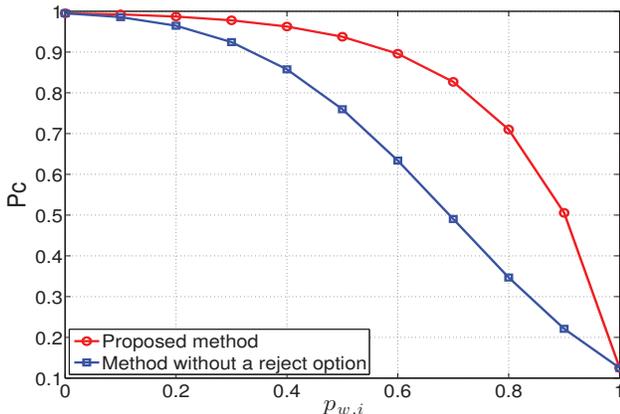

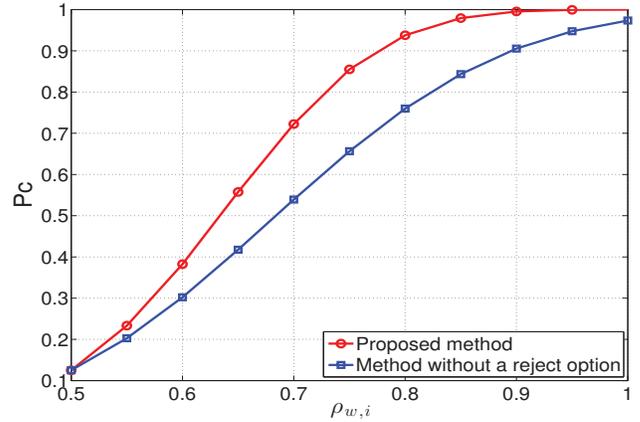Fig. 2. Proposed approach compared to majority voting at $r_{w,i} = 0.8$.

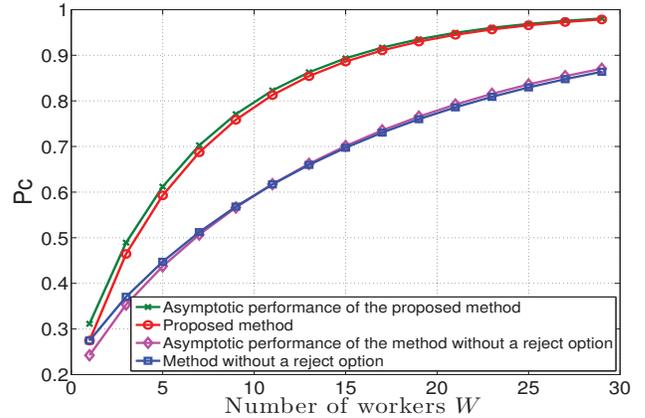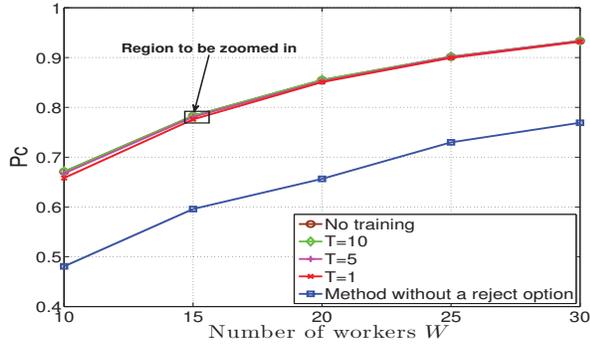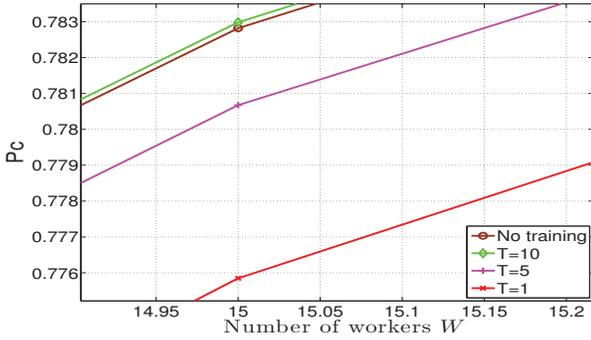Fig. 3. Proposed approach compared to majority voting at $p_{w,i} = 0.5$.

Fig. 4. Proposed approach compared to majority voting with various worker sizes at $p_{w,i} \sim U(0,1)$ and $r_{w,i} \sim U(0.6,1)$.

We now include the estimation of μ in Fig. 5 for weight assignment. Observe in Fig. 5(a) that the performance improves as the number of workers increases and that the proposed approach is significantly better than majority voting. Second, the performance of the proposed approach is significantly better than that of traditional majority voting, and it changes for different estimation settings. As is expected, a larger number of training questions result in better system performance as observed in Fig. 5(b). Another interesting finding is that training-based estimation can outperform majority voting only when a relatively large number of training questions are used. We see from the figure that the Training method with $T = 10$ slightly outperforms the Majority-voting method (without training). However, the number of microtasks N is only 3, which is much smaller than the training size. Much extra overhead beyond the classification task must be added if the training method is adopted. Hence, it is reasonable to employ the Majority-voting method together with the proposed approach for the classification using crowdsourcing.

Fig. 6 shows P_c performance as a function of overhead with different numbers of microtasks, to illustrate the performance gap between the two methods. Observe that the method without training exhibits remarkable advantage when a reasonable number of additional microtasks are inserted.

(a) Performance comparison.

(b) Zoomed-in version

Fig. 5. Proposed approach compared to majority voting with various worker sizes at $p_{w,i} \sim U(0,1)$ and $r \sim U(0.5,1)$. Two methods are used to estimate μ for weight assignment. One uses training to insert T additional microtasks for estimation, whereas the other one uses the decision results of majority voting as a benchmark to estimate μ. (a) provides the performance comparison while (b) is a zoomed-in region which is indicated in the box in (a).
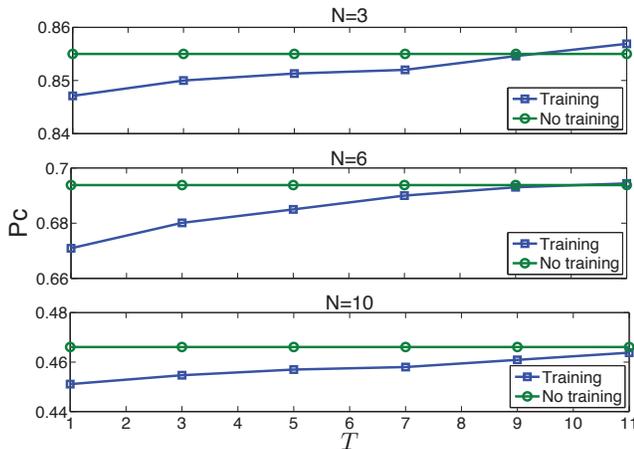

Fig. 6. Performance vs. overhead tradeoff. The crowd size is set as $W = 20$ and $N = 3$, 6, and 10, from top to bottom, respectively.

Improved performance of the training-based method is shown when T gets larger as this results in more accurate estimation. To have comparable performance with the method without training, the training-based method requires even more additional microtasks when the original number of microtasks N increases. With enough microtasks inserted, the training method can outperform the one without. Again, this result encourages employing the method without training in practice.

IV. SYSTEM WITH GREEDY CROWD WORKERS

In Sec. III, we considered conscientious crowd workers who respond only when having confidence in their ability to respond. In our formulation, the weight assigned to a worker's response increases with the number of definitive answers and this contributes to the selection of the correct class. In a reward-based system, such honest workers should be compensated and actually rewarded for completing as many tasks as possible. Typical crowdsourcing setups do in fact pay workers in proportion to the number of microtasks they complete [50]–[52]. However, if there are workers that try to get as much reward as possible without regard to the system goal of classification, such a reward mechanism can encourage these greedy workers to randomly complete all microtasks (without regard to the question being asked). These greedy workers who degrade system classification performance are often termed *spammers* and are known to exist in large numbers on crowdsourcing platforms [50]–[52]. Indeed, Mason and Watts observed that increasing financial incentives increases the quantity of work performed, but not necessarily the quality [53]. Note that the semi-greedy situation where some workers occasionally complete microtasks that they are not confident about can be considered as a special case of our model with $p_{w,i} = 0$ and $r_{w,i} = 1/2$. Thus our model to characterize the greedy behavior of workers is general.

In this section, we study system performance when a part of the crowd completes all microtasks with random guesses. In other words, these greedy workers submit N-bit codewords, termed as *full-length answers*. Note that the semi-greedy situation where some workers occasionally complete microtasks that they are not confident about can be characterized by small values of $p_{w,i}$ and $r_{w,i}$.

Insertion of a gold standard question set is a common method to address the issue of greedy workers, but comes at the cost of a large question set to avoid workers spotting recurrent questions [48]. Besides, this is wasteful since the fundamental reason for crowdsourcing is to collect classified labels that we do not have [49]. We therefore study two different strategies besides inserting a gold standard question set. The *Oblivious Strategy* continues to use the scheme from Sec. III, ignoring the existence of greedy workers. In the *Expurgation Strategy*, we discard the answers of workers who only give full-length answers, to reduce the impact of greedy workers on the overall system performance. Note that this strategy will also discard the responses of those honest workers that provided definitive answers to all microtasks. Also note that greedy workers are not being punished in any way here; only that their responses are being ignored in the aggregation strategy. Let α be the fraction of greedy workers in the crowd.

A. Oblivious Strategy

In this strategy, we continue to use the same weight allocation scheme as for honest workers: $W_w = \alpha_1 \mu^{-n}$, where the factor α_1 is introduced to satisfy the constraint $E_O[\mathbb{W}] = K$.

The average contribution from the crowd to the correct class and all the classes can be given respectively as

$$
\begin{aligned}
E_C[\mathbb{W}] &= \sum_{w=1}^{W\alpha} \alpha_1 \mu^{-N} \frac{1}{2^N} \\
&+ \sum_{w=W\alpha+1}^{W} \sum_{n=0}^{N-1} \alpha_1 \mu^{-n} C_N^n [(1-m)\mu]^n m^{N-n} \\
&= \frac{W\alpha\alpha_1}{(2\mu)^N} + \alpha_1 W (1-\alpha),
\end{aligned}
\tag{29}
$$

and

$$
\begin{aligned}
E_O[\mathbb{W}] &= \sum_{w=1}^{W\alpha} \alpha_1 \left(\frac{1}{\mu}\right)^N \\
&+ \sum_{w=W\alpha+1}^{W} \sum_{n=0}^{N-1} \alpha_1 \mu^{-n} 2^{N-n} C_N^n (1-m)^n m^{N-n} \\
&= W\alpha\alpha_1 \left(\frac{1}{\mu}\right)^N + (W - W\alpha)\alpha_1 \left(\frac{1-m}{\mu} + 2m\right)^N.
\end{aligned}
\tag{30}
$$

Therefore, we can calculate α_1 and obtain $E_C[\mathbb{W}]$ as:

$$
\alpha_1 = \frac{K}{W\alpha\left(\frac{1}{\mu}\right)^N + (W - W\alpha)\left(\frac{1-m}{\mu} + 2m\right)^N},
\tag{31}
$$

$$
E_C[\mathbb{W}] = \frac{K\alpha\left(\frac{1}{2\mu}\right)^N + K(1-\alpha)}{\alpha\left(\frac{1}{\mu}\right)^N + (1-\alpha)\left(\frac{1-m}{\mu} + 2m\right)^N}.
\tag{32}
$$

Proposition 7: The probability of correct classification P_c when the Oblivious Strategy is used is

$$
\begin{aligned}
P_c = \Bigg[\frac{1}{2} + \frac{1}{2} &\sum_{S_1} \binom{W}{\mathbb{Q}_1} (F(\mathbb{Q}_1) - F'(\mathbb{Q}_1)) \\
&+ \frac{1}{4} \sum_{S_1'} \binom{W}{\mathbb{Q}_1} (F(\mathbb{Q}_1) - F'(\mathbb{Q}_1)) \Bigg]^N
\end{aligned}
\tag{33}
$$

with

$$
F(\mathbb{Q}_1) = \frac{m^{q_0}}{2^{W\alpha}} \prod_{n=1}^{N} (1-\mu)^{q_{-n}} \mu^{q_n} \left(C_{N-1}^{n-1} (1-m)^n m^{N-n} \right)^{q_{-n}+q_n}
\tag{34}
$$

and

$$
\begin{aligned}
&F'(\mathbb{Q}_1) \\
&= \frac{m^{q_0}}{2^{W\alpha}} \prod_{n=1}^{N} (1-\mu)^{q_n} \mu^{q_{-n}} \left(C_{N-1}^{n-1} (1-m)^n m^{N-n} \right)^{q_{-n}+q_n}
\end{aligned}
\tag{35}
$$

where

$$
\mathbb{Q}_1 = \left\{ (q_{-N}, q_{-N+1}, \ldots q_N) : \sum_{n=-N}^{N} q_n = W - W\alpha \right\},
\tag{36}
$$

with natural numbers q_n,

$$
S_1 = \left\{ \mathbb{Q}_1 : \sum_{n=1}^{N} \mu^{-n} (q_n - q_{-n}) + \mu^{-N} W\alpha > 0 \right\},
\tag{37}
$$

$$
S_1' = \left\{ \mathbb{Q}_1 : \sum_{n=1}^{N} \mu^{-n} (q_n - q_{-n}) + \mu^{-N} W\alpha = 0 \right\},
\tag{38}
$$

and $\binom{W}{\mathbb{Q}_1} = W! / \prod_{n=-N}^{N} q_n!$.

Proof: See Appendix F. ∎

B. Expurgation Strategy

In this strategy, all definitive answers of length of N bits are discarded to avoid answers from greedy workers. The classification decision is made based on the answers with maximum number of bits equal to $N-1$. To proceed, we need the weight for every worker's answer in this case. We begin by restating the optimization problem

$$
\begin{aligned}
&\text{maximize} \quad E_C[\mathbb{W}] \\
&\text{subject to} \quad E_O[\mathbb{W}] = K
\end{aligned}
\tag{39}
$$

and we have

$$
\begin{aligned}
E_C[\mathbb{W}] &= \sum_{w=1}^{W-W\alpha} \sum_{n=0}^{N-1} W_w \binom{N}{n} [(1-m)\mu]^n m^{N-n} \\
&= \sum_{w=1}^{W-W\alpha} \sum_{n=0}^{N-1} W_w \mu^n x^{n-N} P_x(n),
\end{aligned}
\tag{40}
$$

where

$$
P_x(n) = \binom{N}{n} (1-m)^n (mx)^{N-n},
\tag{41}
$$

and x is such that

$$
\sum_{n=0}^{N-1} P_x(n) = 1.
\tag{42}
$$

Then, we can write

$$
E_C[\mathbb{W}] \le \sum_{w=1}^{W-W\alpha} \sqrt{\sum_{n=0}^{N-1} (W_w \mu^n x^{n-N})^2 P_x(n)} \sqrt{\sum_{n=0}^{N-1} P_x(n)},
\tag{43}
$$

which holds with equality only if

$$
W_w \mu^n x^{n-N} \sqrt{P_x(n)} = \alpha_2 \sqrt{P_x(n)},
\tag{44}
$$

where the factor α_2 is introduced to satisfy the constraint $E_O[\mathbb{W}] = K$. Hence, we have the maximum of $E_C[W_w]$ as

$$
E_C[\mathbb{W}] = W(1-\alpha)\alpha_2,
\tag{45}
$$

when

$$
W_w = \alpha_2 \mu^{-n} x^{N-n}.
\tag{46}
$$

To obtain the value of x, we rewrite (42) as:

$$
(1 - m + mx)^N - (1-m)^N = 1,
\tag{47}
$$

and x is given as

$$x = \frac{\left(1 + (1-m)^N\right)^{\frac{1}{N}} + m - 1}{m}. \quad (48)$$

For this strategy, the overall weight constraint is given as

$$E_O\left[\mathbb{W}\right] = \sum_{w=1}^{W-W\alpha} \sum_{n=0}^{N-1} \alpha_2 \mu^{-n} x^{N-n} 2^{N-n} \binom{N}{n}(1-m)^n m^{N-n}$$
$$= W(1-\alpha)\alpha_2\left[\left(\frac{1-m}{\mu} + 2mx\right)^N - \left(\frac{1-m}{\mu}\right)^N\right]. \quad (49)$$

By substituting this result back into (45), the maximum value of $E_C[\mathbb{W}]$ can be written as

$$E_C\left[\mathbb{W}\right] = \frac{K}{\left(\frac{1-m}{\mu} + 2mx\right)^N - \left(\frac{1-m}{\mu}\right)^N}. \quad (50)$$

Note that the weight could be $W_w = \mu^{-n} x^{-n}$ when the Expurgation Strategy is employed in practice, where x is given by (48).

Proposition 8: The probability of correct classification P_c when the Expurgation Strategy is used is

$$P_c = \left[\frac{1}{2} + \frac{1}{2}\sum_{S_2}\binom{W}{\mathbb{Q}_2}(F(\mathbb{Q}_2) - F'(\mathbb{Q}_2))\right.$$
$$\left. + \frac{1}{4}\sum_{S_2'}\binom{W}{\mathbb{Q}_2}(F(\mathbb{Q}_2) - F'(\mathbb{Q}_2))\right]^N \quad (51)$$

with

$$F(\mathbb{Q}_2) = m^{q_0}\prod_{n=1}^{N-1}(1-\mu)^{q_{-n}}\mu^{q_n}\left(C_{N-1}^{n-1}(1-m)^n m^{N-n}\right)^{q_{-n}+q_n}$$

and

$$F'(\mathbb{Q}_2) = m^{q_0}\prod_{n=1}^{N-1}(1-\mu)^{q_n}\mu^{q_{-n}}\left(C_{N-1}^{n-1}(1-m)^n m^{N-n}\right)^{q_{-n}+q_n},$$

where

$$\mathbb{Q}_2 = \left\{(q_{-N+1}, q_{-N+2}, \ldots q_{N-1}): \sum_{n=-N+1}^{N-1} q_n \leq W - W\alpha\right\}$$

with natural numbers q_n, and

$$S_2 = \left\{\mathbb{Q}_2 : \sum_{n=1}^{N-1}\mu^{-n}x^{-n}(q_n - q_{-n}) > 0\right\},$$

$$S_2' = \left\{\mathbb{Q}_2 : \sum_{n=1}^{N}\mu^{-n}x^{-n}(q_n - q_{-n}) = 0\right\}$$

and $\binom{W}{\mathbb{Q}_2} = W!/\prod_{n=-N+1}^{N-1} q_n!$.
Proof: See Appendix F, *mutatis mutandis.* ∎

C. Adaptive Algorithm

We now investigate the adaptive use of our two strategies to improve system performance. The goal is to find a threshold to distinguish when one strategy will outperform the other, so as to allow switching.

Note that the two strategies are associated with the same overall weight for all classes. Thus, we compare the crowd's total contribution to the correct class under this condition and derive the corresponding switching scheme. From (32) and (45), this can be expressed in (52), which simplifies to having the switching threshold of α as

$$\alpha\left(\left(\frac{1}{\mu}\right)^N - \gamma_1\left(\frac{1}{2\mu}\right)^N - \gamma_2 + \gamma_1\right) \underset{\text{Oblivious Strategy}}{\overset{\text{Expurgation Strategy}}{\gtrless}} \gamma_1 - \gamma_2, \quad (53)$$

where

$$\gamma_1 = \left(\frac{1-m}{\mu} + 2mx\right)^N - \left(\frac{1-m}{\mu}\right)^N, \quad (54)$$

and

$$\gamma_2 = \left(\frac{1-m}{\mu} + 2m\right)^N. \quad (55)$$

To obtain the threshold associated with α in the switching criterion (53), μ and m should be estimated first. The previous section established a simple and effective method to estimate μ based on majority voting. Therefore, we again use majority voting to get initial detection results, which are then set as the benchmark to estimate μ. Note that estimation of μ is based on the answers without the full-length ones to avoid degradation from the greedy workers.

The performance of this integrated scheme can be derived using Props. 7 and 8, and switching criterion (53).

D. Joint Estimation of m and α

The threshold on α is specified based on the estimated values of m and μ. Then, we estimate α and compare it with the corresponding threshold to switch the strategies adaptively. Though we discard full-length answers and use the rest to estimate m, it is an inaccurate estimate because the discarded answers also contain those from honest workers.

Several works have studied the estimation of α in crowd-sourcing systems [18], [21], [41], which can be divided into two categories: one studies the behavior of the workers in comparison to the honest control group [21]; the other one learns worker's reputation profile [18], [41], which is stored and updated over time to identify the greedy ones from the crowd. However, neither estimation method is suitable here due to the anonymous nature of crowd workers. The first category suffers from the difficulty in extracting the honest group from the anonymous crowd while the second requires identification of every worker.

Since the worker's quality is assumed to be i.i.d., we give a joint parametric estimation method of both m and α based on maximum likelihood estimation (MLE).

As defined earlier, out of W workers, $q_n + q_{-n}$ workers submit answers of n bits, $0 \leq n \leq N$. Thus, the probability mass function of the number of submitted answers given m and α

$$\frac{\alpha K \left(\frac{1}{2\mu}\right)^N + K(1-\alpha)}{\alpha \left(\frac{1}{\mu}\right)^N + (1-\alpha)\left(\frac{1-m}{\mu} + 2m\right)^N} \underset{\text{Expurgation Strategy}}{\overset{\text{Oblivious Strategy}}{\gtrless}} \frac{K}{\left(\frac{1-m}{\mu} + 2mx\right)^N - \left(\frac{1-m}{\mu}\right)^N}, \tag{52}$$

$$f\left(q_n + q_{-n}|m,\alpha\right) = \begin{cases} \binom{W-W\alpha}{q_n+q_{-n}} A_{N,n,m}{}^{q_n+q_{-n}}(1 - A_{N,n,m})^{W-W\alpha-q_n-q_{-n}}, 0 \le n < N \\ \binom{W-W\alpha}{q_N+q_{-N}-W\alpha}(1-m)^{N(q_N+q_{-N}-W\alpha)}\left(1-(1-m)^N\right), n = N \end{cases} \tag{56}$$

is obtained in (56), where $A_{N,n,m} = \binom{N}{n}(1-m)^n m^{N-n}$, is defined as the expectation of the probability of a single worker submitting n definitive answers.

Because of the independence of workers, we can form the likelihood statistic as

$$L(m,\alpha) = \sum_{n=0}^{N} \log f\left(q_n + q_{-n}|m,\alpha\right). \tag{57}$$

Therefore, the ML estimates of m and α, which are denoted by $\hat{m}$ and $\hat{\alpha}$, can be obtained as

$$\{\hat{m},\hat{\alpha}\} = \arg \max_{\{m,\alpha\}\in[0,1]} L(m,\alpha). \tag{58}$$

Once we have $\hat{\mu}$, $\hat{m}$ and $\hat{\alpha}$, we can adaptively switch to the suitable strategy using (53).

E. Simulation Results

Now, we present some simulation results to illustrate the performance of our proposed algorithm. First, the theoretical value of the threshold for adaptive switching between the strategies is obtained for different values of m and μ based on (53). We switch to the Expurgation Strategy if the fraction of greedy workers α is greater than the threshold. Otherwise we stick to the Oblivious Strategy. As we observe from Fig. 7, when m decreases and μ increases—when the quality of the crowd improves—the threshold increases. This implies a crowdsourcing system employing the Oblivious Strategy can tolerate a higher fraction of greedy workers in the crowd; instead of discarding all answers from the greedy workers and from those honest workers who submit full-length answers, it is better to keep them as long as the honest ones can perform well. The effect of greedy workers' answers can be compensated by the high-quality answers of the honest workers in the crowd.

Next, we give the estimation results for $\hat{\alpha}$ in Table I using the proposed MLE method. The crowd quality parameters $p_{w,i}$ and $r_{w,i}$ are drawn from distributions $U(0,1)$ and $U(0.5,1)$ respectively. The number of microtasks N and the number of workers W are set to 3 and 20, respectively.

TABLE I
ESTIMATION OF α

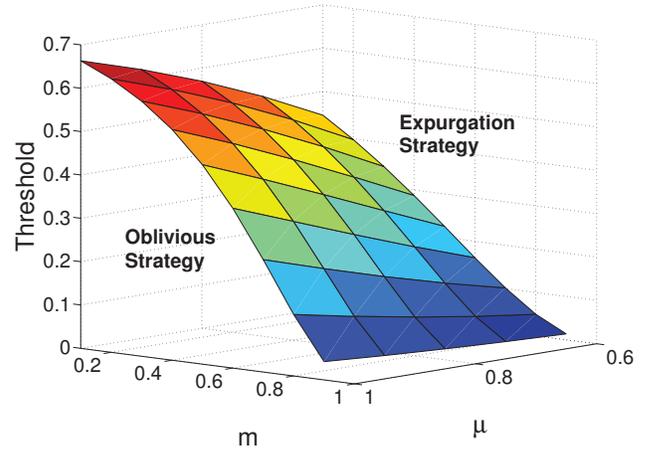| α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\alpha}$	0.11	0.26	0.36	0.48	0.58	0.67	0.79	0.87	0.96

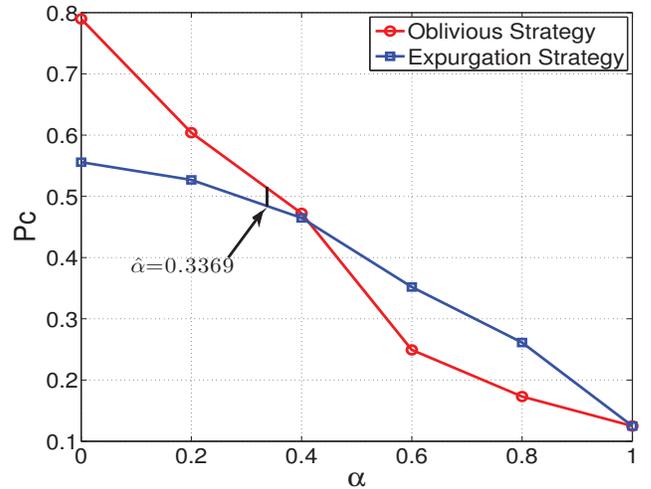Fig. 7. Threshold to switch between strategies.

Fig. 8. Performance of both the strategies with greedy workers.

Fig. 8 shows the performance of the proposed adaptive scheme. The system parameters are the same as in previous simulations except that the crowd size W is set to 15. The crowdsourcing system starts with the estimation of the parameters μ, m, and α. Once it has obtained $\hat{\mu}$ and $\hat{m}$, the system calculates the threshold value and compares it with $\hat{\alpha}$, and then decides the strategy to be used. Next, the system allocates weights to the answers for aggregation based on the strategy selected and makes the final classification decision. Fig. 8

presents the performance of both strategies and the estimated threshold for switching. The performance deterioration caused by greedy workers is quite obvious as the probability of correct classification P_c decreases for both strategies with increasing α. The intersection of curves illustrates the need for strategy switching. The estimated Pareto frontier $\hat{\alpha}$ for switching is 0.3369 in this system setting, which is indicated in the figure by a line segment and is very close to the intersection of the two curves. Therefore, the actual performance curve of the proposed algorithm consists of the curve with squares when $\alpha < \hat{\alpha}$ and curve with circles when $\alpha > \hat{\alpha}$.

V. CONCLUSION AND DISCUSSION

We have studied a novel framework for crowdsourcing of classification tasks that arise in human-based signal processing, where an individual worker has the reject option and can skip a microtask if he/she has no definitive answer. We presented an aggregation approach using a weighted majority voting rule, where each worker's response is assigned an optimized weight to maximize the crowd's classification performance. We have shown our proposed approach significantly outperforms traditional majority voting and provided asymptotic characterization as an upper bound on performance. Further, we considered greedy workers in the crowd. Oblivious and expurgation strategies were studied to deal with greed, and an algorithm to adaptively switch between them, based on the estimated fraction of greedy workers in the anonymous crowd, was developed to combat performance degradation.

We assumed in this paper that the worker qualities are identically distributed. In some settings, worker qualities are not identically distributed, which makes estimating the greedy fraction α difficult. The difficulty of microtasks might not be equal, which makes the microtask design quite challenging. Therefore, further research directions include the development of a general model to characterize the crowd quality, design of a robust method to estimate the fraction of greedy workers in the crowd, and binary microtask design with a reject option.

APPENDIX A

To solve problem (8), we need $E_C[\mathbb{W}]$ and $E_O[\mathbb{W}]$. First, the wth worker can have weight contribution to $E_C[\mathbb{W}]$ only if all his/her definitive answers are correct. Thus, we have the average weight assigned to the correct element as

$$E_C[\mathbb{W}] = E_{p,r}\left[\sum_{w=1}^{W}\sum_{n=0}^{N} W_w P(n, N-n)\right]$$
$$= \sum_{w=1}^{W} E_{p,r}\left[\sum_{n=0}^{N} W_w P(n, N-n)\right], \quad (59)$$

where $P(n, N-n)$ represents the probability of $N-n$ bits equal to λ and the rest of the n definitive answers in the N-bit word are correct.

Then, given the wth worker with $p_{w,i}$ known, we write

$$A_w(p_{w,i}) = E_r\left[\sum_{n=0}^{N} W_w P(n, N-n)|p_{w,i}\right]. \quad (60)$$

Let $P_\lambda(n)$ denote the probability of the wth worker submitting n definitive answers out of N microtasks, which only depends on $p_{w,i}$. Note that $\sum_{n=0}^{N} P_\lambda(n) = 1$, and then (60) is upper-bounded using the Cauchy-Schwarz inequality as:

$$A_w(p_{w,i}) = \sum_{n=0}^{N} E_r[W_w r(n)]\sqrt{P_\lambda(n)}\sqrt{P_\lambda(n)}$$
$$\leq \sqrt{\sum_{n=0}^{N} E_r^2[W_w r(n)]P_\lambda(n)}\sqrt{\sum_{n=0}^{N} P_\lambda(n)} \quad (61)$$
$$\triangleq \alpha_w(p_{w,i}), \quad (62)$$

where $r(n)$ is the product of any n out of N variables $r_{w,i}$ as $i = 1, \ldots, N$, and α_w is a positive quantity independent of n, which might be a function of $p_{w,i}$. Note that equality holds in (61) only if

$$E_r[W_w r(n)]\sqrt{P_\lambda(n)} = \alpha_w(p_{w,i})\sqrt{P_\lambda(n)}, \quad (63)$$

which results in (62) and

$$E_r[W_w r(n)] = \alpha_w(p_{w,i}). \quad (64)$$

Then we maximize the crowd's average weight corresponding to the correct class under the constraint $\int_{p_{w,i}} \Pr(p_{w,i} = x)dx = 1$, and the maximization problem is written as

$$A = E_p[A_w(p_{w,i})] = \int_{p_{w,i}} \alpha_w(p_{w,i}) \Pr(p_{w,i} = x)\,dx$$
$$\leq \sqrt{\int_{p_{w,i}} \alpha_w^2(p_{w,i}) \Pr(p_{w,i} = x)\,dx}\sqrt{\int_{p_{w,i}} \Pr(p_{w,i} = x)\,dx}$$
$$(65)$$
$$\triangleq \beta. \quad (66)$$

The equality (65) holds only if

$$\alpha_w(p_{w,i})\sqrt{\Pr(p_{w,i} = x)} = \beta\sqrt{\Pr(p_{w,i} = x)}, \quad (67)$$

with β is a positive constant independent of $p_{w,i}$, and we conclude that α_w is also a positive quantity independent of $p_{w,i}$. Then from (64), we have

$$E_r[W_w r(n)] = \beta. \quad (68)$$

Since $r(n)$ is the product of n variables, its distribution is not known *a priori*. A possible solution to weight assignment is a deterministic value given by $W_w E[r(n)] = \beta$ and, therefore, we can write the weight as $W_w = \beta/\mu^n$. Note that if $r_{w,i}$ is known *a priori* or can be estimated, the optimal weight assignment is simply $W_w = \beta/r(n)$.

Then, we can express the crowd's average weight contribution to all the classes defined in (8) as

$$E_O[\mathbb{W}] = \sum_{w=1}^{W} E_{p,r}\left[\sum_{n=0}^{N} \beta\mu^{-n} 2^{N-n} P_\lambda(n)\right]$$
$$= \sum_{w=1}^{W}\sum_{n=0}^{N} \beta\mu^{-n} 2^{N-n}\binom{N}{n}(1-m)^n m^{N-n}$$
$$= W\beta\left(\frac{1-m}{\mu} + 2m\right)^N = K. \quad (69)$$

Thus, β can be obtained from (69) and we get the weight by solving optimization problem (8) to get:

$$W_w = \frac{K}{W\mu^n\left(\frac{1-m}{\mu}+2m\right)^N}. \tag{70}$$

Note that the weight derived above has a term that is common for every worker. Since the voting scheme is based on comparison, we can ignore this factor and have the normalized weight as $W_w = \mu^{-n}$.

APPENDIX B

Note that

$$T_w \in \{-\mu^{-N}, -\mu^{-N+1}, \ldots, -\mu^{-1}, 0, \mu^{-1}, \ldots, \mu^{-N+1}, \mu^{-N}\}, \tag{71}$$

which can be written as

$$T_w = I(-1)^{t+1}\mu^{-n}, t \in \{0,1\}, I \in \{0,1\}, n \in \{1,\ldots,N\}, \tag{72}$$

and leads to

$$\Pr\left(T_w = I(-1)^{t+1}\mu^{-n}|H_s\right)$$
$$= \begin{cases} \Pr\left(T_w = \frac{(-1)^{t+1}}{\mu^n}|H_s\right), & I=1 \\ \Pr\left(T_w = 0|H_s\right), & I=0 \end{cases}. \tag{73}$$

These two terms can be expressed as

$$\Pr\left(T_w = \frac{(-1)^{t+1}}{\mu^n}|H_s\right)$$
$$= \Pr\left(\mathbf{a}_w(i) = t|H_s\right) \cdot P_\lambda\left(n|\mathbf{a}_w(i)=t,H_s\right)$$
$$= r_{w,i}^{1-|s-t|}(1-r_{w,i})^{|s-t|}\sum_C \prod_{\substack{j=1\\j\neq i}}^N p_{w,j}^{k_j}(1-p_{w,j})^{1-k_j}, \tag{74}$$

and

$$\Pr\left(T_w = 0|H_s\right) = p_{w,i}. \tag{75}$$

APPENDIX C

Let $q_n, -N \leq n \leq N$, represent the number of workers that submit $|n|$ total definitive answers to all the microtasks. Specifically, $n < 0$ indicates the group of workers that submit "0" for the ith bit while $n > 0$ indicates "1". For $n = 0$, q_0 represents the number of workers that submit λ for the ith bit. Since the workers independently complete the microtasks, recalling the results in (13), the probabilities of the crowd's answer profile for the ith bit $\{G_0, G_1, G_\lambda\}$ can be obtained under H_1 and H_0 given $p_{w,i}$ and $r_{w,i}$ are expressed by $F_i(\mathbb{Q})$ and $F_i'(\mathbb{Q})$, respectively. Thus, $P_{d,i}$ given $p_{w,i}$ and $r_{w,i}$ can be expressed as

$$P_{d,i} = \sum_S \binom{W}{\mathbb{Q}}F_i(\mathbb{Q}) + \frac{1}{2}\sum_{S'}\binom{W}{\mathbb{Q}}F_i(\mathbb{Q}), \tag{76}$$

where the first term on the right-hand side corresponds to the case where the aggregation rule gives a result of "1" and the second term indicates the case where "1" is given due to the tie-breaking of the aggregation rule.

Similarly, we can obtain P_f given $p_{w,i}$ and $r_{w,i}$ as

$$P_{f,i} = \sum_S \binom{W}{\mathbb{Q}}F_i'(\mathbb{Q}) + \frac{1}{2}\sum_{S'}\binom{W}{\mathbb{Q}}F_i'(\mathbb{Q}). \tag{77}$$

Then, it is straightforward to obtain the desired result.

APPENDIX D

We can have a correct classification if and only if all the bits are classified correctly. Thus, the expected probability of correct classification is given as

$$P_c = E\left[\prod_{i=1}^N P_{c,i}\right], \tag{78}$$

which can be expressed, due to the independence of the microtasks, as $P_c = \prod_{i=1}^N E[P_{c,i}]$. Recall $P_{c,i}$ from Prop. 4, and we can obtain:

$$E[P_{c,i}] = \frac{1}{2} + \frac{1}{2}\sum_S\binom{W}{\mathbb{Q}}(F(\mathbb{Q})-F'(\mathbb{Q}))$$
$$+ \frac{1}{4}\sum_{S'}\binom{W}{\mathbb{Q}}(F(\mathbb{Q})-F'(\mathbb{Q})) \tag{79}$$

with $F(\mathbb{Q})$ and $F'(\mathbb{Q})$ defined in (20) and (21). Thus we have the desired result.

APPENDIX E

When W goes to infinity, we show that $E[P_{c,i}] = Q\left(-\frac{M}{\sqrt{V}}\right)$ and the desired result can be obtained. Based on the Central Limit Theorem [54], the test statistic in (12) is approximately Gaussian if $W \to \infty$:

$$\sum_{w=1}^W T_w \sim \begin{cases} \mathcal{N}(M_1, V_1), & H_1 \\ \mathcal{N}(M_0, V_0), & H_0 \end{cases}, \tag{80}$$

where M_s and V_s are the means and variances of the statistic $\sum_{w=1}^W T_w$ under hypotheses H_s, respectively.

For the wth worker, we have the expectation of T_w as

$$M_{H_1} = \sum_{t=0}^1\sum_{n=1}^N (-1)^{t+1}\mu^{-n}(r_{w,i})^t(1-r_{w,i})^{1-t}\varphi_n(w). \tag{81}$$

We define M_1 as

$$M_1 \triangleq WE[M_{H_1}]$$
$$= W\sum_{n=1}^N \mu^{-n}(2\mu-1)\binom{N-1}{n-1}(1-m)^n m^{N-n}$$
$$= \frac{W(2\mu-1)(1-m)}{\mu}\left(\frac{1}{\mu}-\left(\frac{1}{\mu}-1\right)m\right)^{N-1}. \tag{82}$$

Likewise, we define V_1 as

$$V_1 \triangleq W\left(E[T_w^2]-E^2[M_{H_1}]\right)$$
$$= WE\left[\sum_{t=0}^1\sum_{n=1}^N\mu^{-2n}(r_{w,i})^t(1-r_{w,i})^{1-t}\varphi_n(w)\right]-\frac{M_1^2}{W}$$
$$= \frac{W(1-m)}{\mu^2}\left(\frac{1}{\mu^2}-\left(\frac{1}{\mu^2}-1\right)m\right)^{N-1}-\frac{M_1^2}{W}. \tag{83}$$

Similarly, we can derive $M \triangleq M_1 = -M_0$ and $V \triangleq V_1 = V_0$. By looking back at the decision criterion for the ith bit (12), we obtain the desired result.

APPENDIX F

Since the workers complete the microtasks independently, recall the results in (13) and we have

$$E[P_{d,i}] = \sum_{S_1} \binom{W}{\mathbb{Q}_1} F(\mathbb{Q}_1) + \frac{1}{2} \sum_{S_1'} \binom{W}{\mathbb{Q}_1} F(\mathbb{Q}_1), \quad (84)$$

and

$$E[P_{f,i}] = \sum_{S_1} \binom{W}{\mathbb{Q}_1} F'(\mathbb{Q}_1) + \frac{1}{2} \sum_{S_1'} \binom{W}{\mathbb{Q}_1} F'(\mathbb{Q}_1), \quad (85)$$

with $F(\mathbb{Q}_1)$ and $F'(\mathbb{Q}_1)$ as given above. Then, it is straightforward to obtain the desired result.

REFERENCES

[1] H. V. Poor, K.-C. Chen, V. Krishnamurthy, D. Shah, and P. J. Wolfe, "Introduction to the issue on signal processing for social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 511–513, Aug. 2014.

[2] D. Tapscott and A. D. Williams, *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio Penguin, 2006.

[3] J. Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York: Crown Business, 2008.

[4] D. Tapscott and A. D. Williams, *Macrowikinomics: Rebooting Business and the World*. New York: Portfolio Penguin, 2010.

[5] M.-C. Yuen, I. King, and K.-S. Leung, "A survey of crowdsourcing systems," in *Proc. IEEE 3rd Int. Conf. Social Comput. (SOCIALCOM 2011)*, 2011, pp. 766–773.

[6] D. Bollier, *The Future of Work*. Washington, DC: The Aspen Institute, 2011.

[7] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," *IEEE Trans. Multimedia*, vol. 16, pp. 541–558, Jan. 2014.

[8] M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "The four pillars of crowdsourcing: A reference model," in *Proc. IEEE 8th Int. Conf. Research Challenges in Inf. Sci. (RCIS 2014)*, 2014, pp. 1–12.

[9] P. Paritosh, P. Ipeirotis, M. Cooper, and S. Suri, "The computer is the new sewing machine: Benefits and perils of crowdsourcing," in *Proc. 20th Int. Conf. World Wide Web (WWW'11)*, Mar.–Apr. 2011, pp. 325–326.

[10] E. Kamar, S. Hacker, and E. Horvitz, "Combining human and machine intelligence in large-scale crowdsourcing," in *Proc. 11th Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Jun. 2012, pp. 467–474.

[11] S. Burrows, M. Potthast, and B. Stein, "Paraphrase acquisition via crowdsourcing and machine learning," *ACM Trans. Intell. Syst. Technol.*, vol. 4, p. 43, Jul. 2013.

[12] J. Fan, M. Zhang, S. Kok, M. Lu, and B. C. Ooi, "CrowdOp: Query optimization for declarative crowdsourcing systems," *IEEE Trans. Knowl. Data Eng.*, vol. 27, pp. 2078–2092, Aug. 2015.

[13] R. Sprugnoli, G. Moretti, M. Fuoli, D. Giuliani, L. Bentivogli, E. Pianta, R. Gretter, and F. Brugnara, "Comparing two methods for crowdsourcing speech transcription," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2013)*, May 2013, pp. 8116–8120.

[14] T. Hoßfeld, M. Hirth, P. Korshunov, P. Hanhart, B. Gardlo, C. Keimel, and C. Timmerer, "Survey of web-based crowdsourcing frameworks for subjective quality assessment," in *Proc. 2014 IEEE 16th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2014, pp. 1–6.

[15] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in *Proc. 2011 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 3097–3100.

[16] F. Ribeiro, D. Florencio, C. Zhang, and M. Seltzer, "CROWDMOS: An approach for crowdsourcing mean opinion score studies," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2011)*, May 2011, pp. 2416–2419.

[17] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on Amazon Mechanical Turk," in *Proc. ACM SIGKDD Workshop Human Comput. (HCOMP'10)*, Jul. 2010, pp. 64–67.

[18] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality control in crowdsourcing systems: Issues and directions," *IEEE Internet Comput.*, pp. 76–81, Mar. 2013.

[19] K. Mo, E. Zhong, and Q. Yang, "Cross-task crowdsourcing," in *Proc. ACM Int. Conf. Knowl Discovery Data Mining*, Aug. 2013, pp. 677–685.

[20] L. R. Varshney, "Participation in crowd systems," in *Proc. 50th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2012, pp. 996–1001.

[21] M. Hirth, T. Hoßfeld, and P. Tran-Gia, "Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms," *Math. Comput. Model.*, vol. 57, pp. 2918–2932, Jul. 2013.

[22] A. Vempaty, L. R. Varshney, and P. K. Varshney, "Reliable crowdsourcing for multi-class labeling using coding theory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 667–679, Aug. 2014.

[23] D. R. Karger, S. Oh, and D. Shah, "Iterative learning for reliable crowdsourcing systems," in *Advances in Neural Information Processing Systems (NIPS) 24*. Cambridge, MA: MIT Press, 2011, pp. 1953–1961.

[24] M. Hasegawa-Johnson, J. Cole, P. Jyothi, and L. R. Varshney, "Models of dataset size, question design, and cross-language speech perception for speech crowdsourcing applications," *Laboratory Phonology*, vol. 6, pp. 381–431, Oct. 2015.

[25] P. Jyothi and M. Hasegawa-Johnson, "Acquiring speech transcriptions using mismatched crowdsourcing," in *Proc. 29th AAAI Conf. Artificial Intelligence (AAAI'15)*, Nov. 2015.

[26] L. R. Varshney, P. Jyothi, and M. Hasegawa-Johnson, "Language coverage for mismatched crowdsourcing," in *Proc. 2016 Inf. Theory Appl. Workshop*, Feb. 2016.

[27] C. Liu, P. Jyothi, H. Tang, V. Manohar, R. Sloan, T. Kekona, M. Hasegawa-Johnson, and S. Khudanpur, "Adapting ASR for under-resourced languages using mismatched transcriptions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2016)*, Mar. 2016.

[28] P. Jyothi and M. Hasegawa-Johnson, "Transcribing continuous speech using mismatched crowdsourcing," in *Proc. Int. Conf. Spok. Lang. Process. (INTERSPEECH 2015)*, Sep. 2015.

[29] W. Chen, M. Hasegawa-Johnson, and N. F. Chen, "Mismatched crowdsourcing based language perception for under-resourced languages," *Procedia Computer Science*, vol. 81, pp. 23–29, 2016.

[30] X. Kong, P. Jyothi, and M. Hasegawa-Johnson, "Performance improvement of probabilistic transcriptions with language-specific constraints," *Procedia Computer Science*, vol. 81, pp. 30–36, 2016.

[31] R. Dhar, "Consumer preference for a no-choice option," *J. Consumer Research*, vol. 24, pp. 215–231, Sep. 1997.

[32] H. J. Jung, Y. Park, and M. Lease, "Predicting next label quality: A time-series model of crowdwork," in *Proc. 2nd AAAI Conf. Hum. Compt. Crowdsourcing (HCOMP'14)*, Nov. 2014, pp. 87–95.

[33] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, pp. 41–46, Jan. 1970.

[34] P. L. Bartlett and M. H. Wegkamp, "Classification with a reject option using a hinge loss," *J. Mach. Learn. Res.*, vol. 9, pp. 1823–1840, Jun. 2008.

[35] J. B. Predd, D. N. Osherson, S. R. Kulkarni, and H. V. Poor, "Aggregating probabilistic forecasts from incoherent and abstaining experts," *Decision Analysis*, vol. 5, pp. 177–189, Jul. 2008.

[36] K. R. Varshney, "A risk bound for ensemble classification with a reject option," in *Proc. IEEE Workshop Stat. Signal Process. (SSP 2011)*, Jun. 2011, pp. 769–772.

[37] D. Yue, G. Yu, D. Shen, and X. Yu, "A weighted aggregation rule in crowdsourcing systems for high result accuracy," in *Proc. IEEE 12th Int. Conf. Depend., Auton. Secure Comput. (DASC)*, Aug. 2014, pp. 265–270.

[38] D. R. Karger, S. Oh, and D. Shah, "Budget-optimal crowdsourcing using low-rank matrix approximations," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 284–291.

[39] D. Sanchez-Charles, J. Nin, M. Sole, and V. Muntes-Mulero, "Worker ranking determination in crowdsourcing platforms using aggregation functions," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ 2014)*, Jul. 2014, pp. 1801–1808.

[40] A. J. Quinn and B. B. Bederson, "Human computation: a survey and taxonomy of a growing field," in *Proc. 2011 Annu. Conf. Hum. Factors Comput. Syst. (CHI 2011)*, May 2011, pp. 1403–1412.

[41] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. 31st IEEE Conf. Computer Commun. (INFOCOM 2012)*, Mar. 2012, pp. 2140–2148.

[42] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Inform. Fusion*, vol. 6, pp. 63–81, Mar. 2005.

[43] R. Herbei and M. H. Wegkamp, "Classification with reject option," *Canadian J. Stat.*, vol. 34, pp. 709–721, Dec. 2006.

[44] F. Condessa, J. Bioucas-Dias, C. A. Castro, J. A. Ozolek, and J. Kovacevic, "Classification with reject option using contextual information," in *2013 IEEE 10th Int. Symp. Biomedical Imag.*, Apr. 2013, pp. 1340–1343.

[45] I. Pillai, G. Fumera, and F. Roli, "A classification approach with a reject option for multi-label problems," in *Proc. 2011 16th Int. Conf. Image Anal. Process. (ICIAP)*, Sep. 2011, pp. 98–107.

[46] J. Rocker, C. M. Yauch, S. Yenduri, L. Perkins, and F. Zand, "Paper-based dichotomous key to computer based application for biological indentification," *J. Comput. Sci. Coll.*, vol. 22, pp. 30–38, May 2007.

[47] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, pp. 98–101, Jan. 1986.

[48] D. E. Difallah, G. Demartini, and P. Cudré-Mauroux, "Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms." in *CrowdSearch*, Apr. 2012, pp. 26–30.

[49] H. J. Jung and M. Lease, "Modeling temporal crowd work quality with limited supervision," in *Proc. AAAI Workshop Human Comput. (HCOMP'15)*, Nov. 2015, pp. 83–91.

[50] P. Wais, S. Lingamneni, D. Cook, J. Fennell, B. Goldenberg, D. Lubarov, D. Marin, and H. Simons, "Towards building a high-quality workforce with mechanical turk," *NIPS workshop on computational social science and the wisdom of crowds*, pp. 1–5, 2010.

[51] J. Vuurens, A. P. de Vries, and C. Eickhoff, "How much spam can you take? an analysis of crowdsourcing results to increase accuracy," in *Proc. ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIR?11)*, 2011, pp. 21–26.

[52] S. Jagabathula, L. Subramanian, and A. Venkataraman, "Reputation-based worker filtering in crowdsourcing," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 2492–2500.

[53] W. Mason and D. J. Watts, "Financial incentives and the "performance of crowds"," in *Proc. ACM SIGKDD Workshop Human Comput. (HCOMP'09)*, Jun. 2009, pp. 77–85.

[54] A. de Moivre, *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*, 3rd ed. London: A. Millar, 1756.

Lav R. Varshney (S'00–M'10–SM'15) received the B.S. degree (*magna cum laude*) in electrical and computer engineering with honors from Cornell University, Ithaca, New York, in 2004. He received the S.M., E.E., and Ph.D. degrees, all in electrical engineering and computer science, from the Massachusetts Institute of Technology, Cambridge, in 2006, 2008, and 2010, where his theses received the E. A. Guillemin Thesis Award and the J.-A. Kong Award Honorable Mention.

He is an assistant professor in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute at the University of Illinois at Urbana-Champaign. During 2010–2013, he was a research staff member at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York. His research interests include information and coding theory; limits of nanoscale, social, and neural computing; human decision making and collective intelligence; and creativity.

Dr. Varshney is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi. He received the IBM Faculty Award and was a finalist for the Bell Labs Prize, both in 2014. He and his students have won several best paper awards including the 2015 Data for Good Exchange Paper Award. His work appears in the anthology, *The Best Writing on Mathematics 2014* (Princeton University Press). He is a founding member of the IEEE Special Interest Group on Big Data in Signal Processing and currently serves on the Shannon Centenary Committee of the IEEE Information Theory Society. He also currently serves on the advisory board of the AI XPRIZE.

Qunwei Li (S'16) received the B.S. and M.S. degrees in electrical engineering with honors from Xidian University, Xi'an, China, in 2011 and 2014. He has been pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science, Syracuse University since 2014.

He received the Syracuse University Graduate Fellowship award in 2014. His research interests include statistical signal processing, crowdsourcing, machine learning, and optimization.

Aditya Vempaty (S'12-M'15) was born in Hyderabad, India, on August 3, 1989. He received the B. Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Kanpur, in 2011, with academic excellence awards for consecutive years, and the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY in 2015.

He is a Postdoctoral Researcher at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. He was a Graduate Research Intern in the Data Systems Group at Intel Corporation in Santa Clara, CA, in summer 2013 and a Visiting Student at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, in summer 2014. His research interests include human-machine inference networks, data analytics and statistical signal processing, and network security.

Dr. Vempaty is a member of Phi Kappa Phi and Golden Key International Honor Society. He received the Syracuse University Graduate Fellowship award in 2013 and the All University Doctoral Prize 2016 by Syracuse University for superior achievement in completed dissertations.

Pramod K. Varshney (S'72-M'77-SM'82-F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively.

From 1972 to 1976, he held teaching and research assistantships with the University of Illinois. Since 1976, he has been with Syracuse University, Syracuse, NY, where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department from 1993 to 1996. He is also an Adjunct Professor of Radiology at Upstate Medical University, Syracuse. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He has published extensively. He is the author of Distributed Detection and Data Fusion (New York: Springer-Verlag, 1997). He has served as a consultant to several major companies.

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while at the University of Illinois. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the Guest Editor of the Special Issue on Data Fusion of the IEEE PROCEEDINGS January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor?s Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He is on the Editorial Boards of the Journal on Advances in Information Fusion and IEEE Signal Processing Magazine. He was the President of International Society of Information Fusion during 2001.