

Prospect Theory Based Crowdsourcing for Classification in the Presence of Spammers

Baocheng Geng^{ID}, Qunwei Li, and Pramod K. Varshney^{ID}, *Life Fellow, IEEE*

Abstract—We consider the M -ary classification problem via crowdsourcing, where crowd workers respond to simple binary questions and the answers are aggregated via decision fusion. The workers have a reject option to skip answering a question when they do not have the expertise, or when the confidence of answering that question correctly is low. We further consider that there are spammers in the crowd who respond to the questions with random guesses. Under the payment mechanism that encourages the reject option, we study the behavior of honest workers and spammers, whose objectives are to maximize their monetary rewards. To accurately characterize human behavioral aspects, we employ prospect theory to model the rationality of the crowd workers, whose perception of costs and probabilities are distorted based on some value and weight functions, respectively. Moreover, we estimate the number of spammers and employ a weighted majority voting decision rule, where we assign an optimal weight for every worker to maximize the system performance. The probability of correct classification and asymptotic system performance are derived. We also provide simulation results to demonstrate the effectiveness of our approach.

Index Terms—Classification, crowdsourcing, spammers, human behavioral analysis, distributed inference, information fusion, prospect theory.

I. INTRODUCTION

CROWDSOURCING has attracted intense interest in recent years as a new paradigm for distributed inference. It harnesses the intelligence of the crowd, by exploiting the inexpensive and online labor markets in an effective manner [1]–[5]. Crowdsourcing enables a new framework to utilize distributed human wisdom to solve problems that machines cannot perform well, like handwriting recognition, anomaly detection, voice transcription, and image labelling [6]–[8]. While conventional group collaboration and cooperation frameworks rely heavily on a collection of experts in related fields, the crowd in crowdsourcing usually consists of non-experts. Therefore, the responses obtained from the crowd have diverse quality levels, which makes decision fusion in the problem of classification via crowdsourcing quite challenging.

Manuscript received August 28, 2019; revised May 6, 2020; accepted June 23, 2020. Date of publication July 3, 2020; date of current version July 24, 2020. This work was supported by NSF under Grant ENG 1609916. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bo Chen. (*Corresponding author: Qunwei Li.*)

Baocheng Geng and Pramod K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13205 USA (e-mail: bageng@syr.edu; varshney@syr.edu).

Qunwei Li is with the Ant Financial Services Group, Hangzhou 310099, China (e-mail: qunwei.qw@antfin.com).

Digital Object Identifier 10.1109/TSP.2020.3006754

Although crowdsourcing has been applied in many applications, the quality of the aggregated result is relatively low [9]–[11] due to the following reasons. First, the worker pool is anonymous in nature, which may result in an unskilled and unreliable crowd [12]. Second, the assumption that the workers are sufficiently motivated, extrinsically or intrinsically, to take part seriously in the crowdsourcing task, is highly questionable [13], [14]. Third, for the non-expert crowd to successfully complete the crowdsourcing work, some tasks are specifically designed to be composed of easy but tedious microtasks [15], which might cause boredom and result in low-quality work. Finally, noisy and unreliable responses to the tasks cannot be detected and tagged before aggregation so that appropriate weights could be assigned to responses [16]. For this reason, simple majority voting is widely used as the aggregation rule and it takes all of the answers (including the noisy and low quality ones) into account with the same weight [17].

Different methods have been developed to deal with the above problems [12], [14]–[16], [18]–[22]. In [18], the authors decompose a complex task into simple binary questions that are easy for the workers in the crowd to accomplish. It is expected that very little knowledge would be needed to complete the microtasks, and typically common sense or observation is good enough for such microtasks. The authors in [15] employ taxonomy and dichotomous keys in the design of the simple binary questions and the optimal question ordering problem in crowdsourcing is considered in [23]. These schemes that break hard questions into simple ones lower the chance for the workers to make mistakes in responding to each of the questions. Different decision fusion rules are developed in order to deal with the unreliability of the crowd and increase the classification accuracy [16], [20]. We have proposed the use of coding and decoding algorithms for reliable classification with unreliable crowd workers [12], [15]. The comparison between group control and majority voting techniques are presented in [14], which suggests that majority voting is more cost-efficient on simple binary tasks.

In the past literature, crowd workers could only submit a definitive yes/no answer in responding to a binary microtask/question. However, research in psychology [24] indicates a frequent tendency to select the reject option (no choice) when the choice set offers several attractive alternatives but none that can be easily justified as the best. In such cases, the workers may be unsure in answering some of the questions because of their lack of expertise. For instance, phonemes in some languages are very hard to distinguish, especially for foreigners [25]. To avoid requiring workers to respond to microtasks beyond their

expertise resulting in random guesses, we considered the optimal design of the aggregation rule in crowdsourcing systems where the workers are not forced to make a binary choice when they are unsure of their response and can choose not to respond [26]. As shown in [27], the quality of label prediction can be improved by adopting a decision rejection option to avoid results with low confidence. The reject option has also been considered in machine learning and signal processing literatures [28]–[31]. With a reject option, the payment mechanism is investigated in crowdsourcing systems where the workers can also report their confidence about the submitted answers [32].

Additionally, in crowdsourcing systems, there can be greedy crowd workers, also known as spammers, who aim to earn more monetary rewards by answering as many questions as possible. They often submit random guesses independent of the questions being asked. The presence of spammers degrades the system performance and has posed a threat to many crowdsourcing applications [33]. Currently, there are two categories of anti-spammer techniques in crowdsourcing: a priori reputation system [34], [35] and a posteriori quality control [36]–[38]. The first method aims to manage a pool of honest workers with high reputation, so as to ensure the reliability of their answers. However, since the crowd is usually large, anonymous and transient, it is impractical to keep track of the workers' answers and build up a trust relationship. In quality control schemes, several verifiable (golden standard) questions are inserted for the workers to answer and those who do not perform well on these questions are identified as spammers. However, due to the heterogeneous expertise levels of the crowd workers and the subjective criterion used to decide on a spammer, it is easy to mistakenly identify an honest worker to be a spammer, which discourages the worker from participating in answering the questions next time. According to the study in [39], it is better to treat all the crowd workers to be honest, than risking to identify honest workers as spammers.

We study the presence of spammers in the context of a scenario where crowd workers answer questions with a reject option. The payment mechanism proposed in [32], which encourages the workers to skip/reject answering a question below some confidence threshold, is employed. In this scenario, both the honest workers and the spammers choose either to 'answer a question'¹ or 'skip a question' to maximize their monetary rewards. In [32], the authors assume that the workers are rational decision makers in the sense that they can perceive the expected payoff of taking each decision without biases. However, since the crowd workers/spammers are humans, they are subject to cognitive biases in decision making and have disparate behavioral properties. The Nobel Prize winning prospect theory (PT) proposed by Kahneman and Tversky [40] provides a psychologically accurate description of human cognitive biases that includes diminishing marginal utility, risk seeking/aversion and asymmetric valuation of gains and losses. In realistic crowdsourcing applications, these behavioral properties can not be ignored while developing

effective decision making strategies of the honest crowd workers and the spammers.

We employ PT to model the rationality of the crowd workers, and study their behavior while answering or skipping the microtasks/questions² in crowdsourcing that has a reject option. Based on the behavioral difference between the honest workers and the spammers, we design the optimal aggregation rule at the fusion center (FC) to combat the effects of spammers. The contributions of our work are two fold:

- By applying PT to model human cognitive biases, we study the optimal behavior of the honest workers and the spammers based on the payment mechanism proposed in [32]. This payment mechanism has been proved to be the only mechanism that satisfies the "no-free-lunch" rule and exhibits incentive compatibility. We find that the spammers should either complete or skip all the microtasks in order to get the maximal reward. The statistical behavioral properties of the crowd determine whether the spammers should complete or skip all the microtasks.
- We provide methods for estimating the number of spammers that is used for weight assignment. We also design an optimal aggregation rule where the workers are assigned appropriate weights.³ Probability of correct classification and asymptotic performance of our method are derived.

It should be noted that our approach only requires the workers to respond to several microtasks in one session without identifying themselves. Hence, our proposed method can be employed in many applications where the workers remain anonymous and, therefore, it is not required to keep track of the workers' profiles. On the other hand, instead of detecting the spammers, we estimate and employ the number of spammers while designing the optimal counter-measure to ameliorate their effects. The spammers still get paid according to the payment mechanism. As a result, the risk of declaring honest workers as spammers and preventing them from further participation is avoided.

The rest of the paper is organized as follows. In Section II, we introduce the problem of multi-object classification via crowdsourcing with a reject option. In Section III, by employing PT, we study the optimal behavior of the honest crowd workers when their objective is to maximize their monetary rewards. In Section IV, we consider that there are spammers in the crowd and develop a weight assignment scheme for optimal decision aggregation at the FC. Asymptotic performance analysis of our proposed method is presented in Section V. We conduct simulation results in Section VI and conclude our work in Section VII.

II. CLASSIFICATION via CROWDSOURCING WITH A REJECT OPTION

We formulate the classification problem via crowdsourcing with a reject option in this section. Assume that we have W workers participating in an M -ary classification task. There are $N = \lceil \log_2 M \rceil$ simple binary questions to be answered by each worker, where we consider that the binary questions are

¹In this case, honest workers submit their true answers and the spammers submit random guesses.

²The terms 'microtask' and 'question' are used interchangeably in the paper.

³A brief discussion on the weight assignment was presented in [41] In this work, we elaborate on details and explanations related to this model.

independent of each other and are of the same difficulty. For each of the questions, the worker can either provide a definitive answer “1” (Yes) / “0” (No) [15], [42], or has a reject option to skip the question, where a skipped answer is denoted as λ . Let \mathbf{a}_w represent the N -bit word that contains the w th worker’s ordered answers to all the microtasks, where $\mathbf{a}_w(i) \in \{1, 0, \lambda\}$ for $i = 1, \dots, N$. We assume the following statistical properties for the honest workers in the crowd: let $p_{w,i}$ be the probability that the w th worker submits λ to the i th question, i.e., $\mathbf{a}_w(i) = \lambda$, and let $r_{w,i}$ be the probability that $\mathbf{a}_w(i)$ is the correct answer to the i th question, given that the worker has provided definitive answers “1” or “0”. Since the workers in the crowd are anonymous and have diverse expertise levels, we consider that $p_{w,i}$ and $r_{w,i}$ are random and follow certain probability density functions (PDFs) $f_p(p)$ and $f_r(r)$, respectively. The expected values of $p_{w,i}$ and $r_{w,i}$, namely, the average probability that a worker submits λ to a question, and the average probability that an answer is correct given a definitive answer has been submitted, are denoted by m and μ , respectively.

After the N -bit words regarding an object from all the workers are collected by the FC, the object needs to be classified to a class $d_j \in D$, $j = 1, \dots, M$, where D is the set of all the object classes and d_j is the j th class. From the N -bit word $\mathbf{a}_w(i)$ submitted by the w th worker, one can infer the classification decision of the w th worker namely the subset of classes D_w ⁴ to which the object belongs to. At the FC, each class d_j inside D_w increments its candidate score by the weight assigned to the w th worker W_w . After incorporating the responses from all the W workers, the FC determines the class with the highest overall candidate score to be the final classification result:

$$d^* = \arg \max_{d_j \in D} \left\{ \sum_{w=1}^W W_w I_{D_w} \langle d_j \rangle \right\}, j = 1, \dots, M, \quad (1)$$

where $I_{D_w} \langle d_j \rangle = 1$ if $d_j \in D_w$ and $I_{D_w} \langle d_j \rangle = 0$ otherwise. The objective is to find the appropriate weight assignment W_w for every worker in the crowd, so that the best classification performance can be achieved. One approach is to split the M -ary classification task into N binary hypothesis testing problems, each of which determines a bit in the N -bit word. For each hypothesis testing problem, the Chair-Varshney rule gives the optimal weight as $W_w = \log \frac{r_{w,i}}{1-r_{w,i}}$ [43]. However, this requires the prior knowledge regarding $r_{w,i}$ for every worker, which is not available in practice. One may also look into the minimization of the misclassification probability, for which a closed-form expression for W_w cannot be derived due to the lack of prior knowledge of $p_{w,i}$ and $r_{w,i}$.

In [26], we developed a weight assignment scheme to optimize the crowd workers’ weights:

$$\begin{aligned} & \text{maximize } E_C[\mathbb{W}] \\ & \text{subject to } E_O[\mathbb{W}] = K \end{aligned} \quad (2)$$

where $E_C[\mathbb{W}]$ denotes the crowd’s average weight contribution to the correct class and $E_O[\mathbb{W}]$ denotes the average weight

⁴If all the responses from the w th worker are definitive, D_w is a singleton. Otherwise, D_w contains multiple classes.

contribution to all possible classes. We set K to be a constant so that the portion of weight contribution to the correct class is maximized while the weight contribution to all the classes remains fixed. By assuming that there are no spammers in the crowd, we showed that the weight assigned to the w th worker is $W_w = \mu^{-n}$, where n represents the number of definitive answers the worker submits in total. This method significantly outperforms the conventional majority voting approach.

In this following sections, we investigate the impact of spammers on system performance. Based on the payment mechanism that encourages the workers to skip the questions about which they are not sure, we characterize the behavior of both honest workers and spammers in realistic environments where they are subject to cognitive biases while decision making. With this information, we estimate the number of spammers in the crowd and design the weight assignment strategy for every worker to ameliorate the impacts of the spammers and maximize the system performance.

III. BEHAVIOR OF THE HONEST CROWDWORKERS

In this section, we consider that there are no spammers in the crowd and explore the workers’ behavior in answering or skipping a microtask. We adopt the payment mechanism proposed in [32], which encourages the use of the reject option when the confidence of answering a question is low. The mechanism was proved to be the only incentive-compatible mechanism that satisfies the “no-free-lunch” axiom (“no-free-lunch” axiom requires that the payment is minimum possible if all the answers attempted by the worker in the gold standard questions are wrong). Under this payment mechanism, we discuss the optimal behavior for the honest workers when they are rational decision makers and want to maximize their monetary rewards. Next, considering that the honest workers are human decision makers that are subject to cognitive biases in practice, we employ prospect theory to model their behavioral property and analyze their decision making strategy in realistic environments. Compared to the case where the workers are assumed to behave rationally, it is shown that the behavioral factors captured in PT may cause the humans to act quite differently.

A. Payment Mechanism and Optimal Behavior of Rational Crowd Workers

The payment to the worker is based on the answers that the worker gives to the G gold standard questions (which are not known to the crowdworkers in advance). The goal of the mechanism is to incentivize the worker to skip the questions for which its confidence is lower than a threshold T , where confidence about an answer is the probability of this answer being correct. The value of $T \in [0, 1]$ is chosen *a priori* based on factors such as the targeted performance quality. A larger value of T leads to a higher probability that a question is skipped (or equivalently, a lower probability that this question is answered). When T is large, the answer has a higher probability of being correct given that a definitive answer has been submitted. Let f denote the payment rule, which is proposed in [32] and is written

as

$$f(x_1, \dots, x_G) = \kappa \prod_{i=1}^G \alpha_{x_i} + \mu_{\min} \quad (3)$$

where $x_j \in \{-1, \lambda, +1\}$, $1 \leq j \leq G$, are the responses to the gold standard questions. “−1” denotes that the worker attempted to answer the microtask and the answer was incorrect, “ λ ” denotes that the worker skipped the microtask, and “+1” denotes that the worker attempted to answer the microtask and the answer was correct. Set $\alpha_{-1} = 0$, $\alpha_{\lambda} = 1$, $\alpha_{+1} = \frac{1}{T}$, and $\kappa = (\mu_{\max} - \mu_{\min})T^G$ with budget parameters μ_{\max} and μ_{\min} denoting the maximum and minimum payments respectively. Note that μ_{\min} is a constant that represents the fixed reward, and $\kappa \prod_{i=1}^G \alpha_{x_i}$ represents the variable reward that is determined by the worker’s answers. According to the payment mechanism, the variable reward is multiplied by a factor of $\frac{1}{T}$ when the worker answers the microtask and the answer is correct, and the variable reward reduces to 0 if the answer is wrong. Therefore, only when the worker’s confidence towards a microtask is higher than T , the expected payoff is positive and it is beneficial for the worker to answer the microtask. Otherwise, there is a loss in expected reward if the answer to the microtask is incorrect and in this case the worker should use the skip option. Note that when the confidence is exactly equal to T , the worker can choose to either answer or skip the microtask. Under this payment mechanism, the workers are encouraged to use the reject option when they are not sure of their answer, i.e., when the confidence regarding the question is smaller than T . Next, we employ PT to model the rationality of the crowd workers in deciding to answer or skip a microtask.

B. Behavior of Crowdworkers Predicted by Prospect Theory

In this subsection, we analyze the behavior of the honest workers while considering their cognitive biases. We start with a brief introduction of PT. According to PT, human cognitive biases have the following two characteristics: (i) a human decision maker is loss averse in that he/she strongly prefers avoiding losses than achieving gains, (ii) humans are risk seeking in events with small probabilities and are risk averse in events with large probabilities, which can be interpreted as people overweighting small probabilities and underweighting large probabilities. Under PT, humans are loss averse and have asymmetric valuation towards gains and losses through a value function:

$$v(x) = \begin{cases} x^\delta & x \geq 0 \\ -\beta(-x)^\delta & x < 0 \end{cases} \quad (4)$$

where $v(x)$ assigns a subjective utility to an outcome x . A positive x represents a gain and a negative x represents a loss. As β changes, $v(x)$ reflects different loss aversion attitudes of humans. Under PT, gains and losses are perceived with respect to a reference point, which is usually the current wealth and is set to be 0 in (4). δ is the coefficient of diminishing marginal utility and it suggests that as more gains are added (or more losses are subtracted) to the present wealth, the subjective utility of an additional gain (or loss) becomes more insignificant. The value function $v(x)$ is illustrated in Fig. 1(a).

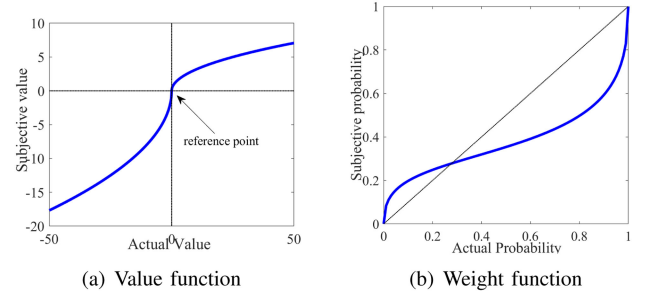


Fig. 1. Illustration of the value function and the weight function in prospect theory.

As shown in Fig. 1(b), the probability weighting function shows humans’ distorted perception of probability with which an event occurs:

$$w(x) = \frac{x^\alpha}{(x^\alpha + (1-x)^\alpha)^{1/\alpha}} \quad (5)$$

where x is the actual probability and α is the probability distortion coefficient. A smaller value of α means more severe distortion of the actual probability x . For rational decision makers, the PT behavioral parameters $\alpha = \beta = \delta = 1$. In practice, these parameters can be estimated through real experimental data and Tversky *et al.* showed that the means of α , β and δ are 0.69, 2.25 and 0.88, respectively [44].

According to the expected utility theory (EUT) [45], when a decision maker selects an action from a set of alternative choices, the one that results in a higher expected payoff is always preferred. In realistic environments, we employ PT to model the rationality of the crowdworkers and study their strategy in responding to the microtasks. In our set up, the crowdworker has to decide whether to answer or skip a particular microtask based on the confidence t , i.e., the probability of correctly answering the microtask. Under PT, a behaviorally biased crowdworker makes decisions according to the following theorem to maximize the subjective payoff.

Theorem 1: If the confidence of a crowdworker with behavioral parameters α , β and δ towards a question is t , then he/she decides to answer or skip a question according to the following rule:

$$\frac{t}{1-t} \underset{\text{skip}}{\overset{\text{answer}}{\gtrless}} \left(\frac{\beta T}{1-T} \right)^{\frac{\delta}{\alpha}} \triangleq \eta. \quad (6)$$

Proof: See Appendix A

Note that (6) can be written as:

$$t \underset{\text{skip}}{\overset{\text{answer}}{\gtrless}} t^* \triangleq \frac{\eta}{1+\eta}. \quad (7)$$

For a rational decision maker with $\alpha = \beta = \delta = 1$, $t^* = T$ and the decision rule (6) suggests answering the question if $t > T$, and skip the question otherwise. We find that when $\beta T \geq 1 - T$, t^* becomes larger, i.e., the worker is more likely to skip the question, as β , δ increase and α decreases. Otherwise, t^* becomes smaller as δ increases and α , β decrease.

We consider that all the binary questions are equally difficult and the w th crowd worker has average confidence t of answering a question correctly. Let t follow the PDF $f_t^w(t)$. According to the behavior characterized in (7), the probability that the w th crowd worker skips a question can be expressed as $p_w = \int_0^{t^*} f_t^w(t)dt$. The probability that the answer is correct, given the w th worker has submitted a definitive answer, can be expressed as $r_w = \int_{t^*}^1 f_t^w(t)dt$. Statistically, p_w determines the number of definitive answers n of the w th worker and r_w determines μ . Recall that the weight assignment scheme in [26] is $W_w(n) = \mu^{-n}$. The variations of p_w and r_w in the crowd may lead to different weights assigned to the workers. Given the confidence PDFs of all the workers $F_t = \{f_t^1(t), \dots, f_t^W(t)\}$ and a priori T , the classification performance obtained assuming that the workers are rational is not accurate in predicting the system performance in realistic situations, where the crowd workers make decisions under cognitive biases. The first part of our simulations in Section VI provides comparisons between the system performance for a crowd with different behavioral properties.

IV. CLASSIFICATION IN THE PRESENCE OF SPAMMERS

Spammers are known to exist in large numbers on crowdsourcing platforms. They submit their answers randomly without being relevant to the question being asked, in the hope of earning some extra money. In this section, we determine the optimal behavior of the spammers that maximizes their expected monetary reward based on the payment mechanism (3). By optimal behavior, we mean the optimal number of questions to be skipped by the spammers. We assume that a spammer skips g out of G gold standard questions, and answers the remaining $G - g$ by random guesses.

A. Spammers are Rational

First, we assume that the spammers behave rationally, i.e., they do not have cognitive biases while calculating the expected payoffs of answering or skipping a question. We hereby define the spammers that skip all the microtasks as Type I spammers, and the spammers that complete all the microtasks as Type II spammers.

Proposition 1: To maximize the expected monetary reward, a rational spammer completes all the microtasks (Type II) if $T < \frac{1}{2}$, and skips all the microtasks (Type I) otherwise.

Proof: See Appendix B. ■

The above proposition gives the optimal strategy for the spammers to participate in the crowdsourcing task. Since a spammer can not distinguish the gold standard ones from the other questions, the result derived in Proposition 1 indicates that the spammers should either complete or skip all the questions according to the value of T to maximize their expected monetary reward. In realistic applications, the FC selects T ($T > 1/2$) to ensure the high quality of the workers' definitive answers (note that the workers' minimum possible value of the confidence regarding a question is $1/2$, when the answer is a random guess). In this case, according to Proposition 1, all the spammers are Type I and choose to skip all the microtasks. It is shown later in

the paper (Appendix C) that the answers from Type I spammers who respond λ to all the microtasks are not aggregated for decision fusion. In other words, the weights assigned to Type I spammers do not affect the aggregation result and in this sense, we can consider all Type I spammers to be honest. As a result, by assuming that spammers behave rationally, all the workers in the crowd can be treated as honest ones and the weight assignment scheme is $W_w(n) = \mu^{-n}$.

B. Spammers are Modeled by PT

As the spammers are also human decision makers, we employ PT to model their rationality and predict their behavior in completing the micro-tasks. From the result of Theorem 1, we have the following corollary.

Corollary 1: To maximize the subjective monetary reward under PT, a spammer with behavioral parameters α, β, δ completes all the microtasks (Type II) if $T > \frac{1}{\beta+1}$, and skips all the tasks (Type I) otherwise.

Proof: Since a spammer employs random guesses to respond to the microtasks, $t = 1/2$. From the result of Theorem 1 and note $\alpha > 0, \delta > 0$, the corollary follows. ■

Based on the above analysis of the spammers' behavior, we study the optimal weight assignment strategy at the FC and the classification performance in this section. As mentioned earlier, the workers/spammers in crowdsourcing systems have different backgrounds and are heterogeneous in their behavioral parameters α, β, δ . Considering that the spammers behave according to Corollary 1, we need the loss aversion coefficient β to predict whether the spammers answer or skip all the questions. Since the spammers remain anonymous in the crowd and certainly do not want to expose themselves, elicitation of parameter β for the spammers is not possible. Therefore, without loss of generality, we assume that among the crowd workers of size W , there are a total of $M = M_0 + M_N$ spammers, with M_0 Type I spammers skipping all the microtasks and M_N Type II spammers completing all the N microtasks.

The presence of spammers will significantly affect the classification performance of the crowdsourcing system, which may make it even worse when the spammers are starting to act strategically. To ameliorate the spammers' impact on system performance, we propose the aggregation rule for the FC by maximizing the candidate score assigned to the correct classification class as in (2). We denote our method as Amelioration of Spammers under PT (ASPT).

Proposition 2: In a crowd with M_0 Type I spammers and M_N Type II spammers, the weight for the w th worker's answer under formulation (2) is given by

$$W_w(n) = \left[(W - M) \mu^n + \frac{M_N}{2^N(1 - m)^N} \delta(n - N) \right]^{-1} \quad (8)$$

where n represents the number of definitive answers submitted by the w th worker, and $\delta(\cdot)$ is the Dirac delta function.

Proof: The detailed proof is presented in Appendix C of [46]. ■

Since the answers from Type I spammers who respond λ to all the microtasks are not aggregated for decision fusion,

M_0 does not appear in the expression (8). In our scheme, the honest workers whose answers are not all definitive employ weights equal to $W_w(n) = \frac{1}{(W-M)\mu^n}$. This is the same as the weight assignment $W_w(n) = \mu^{-n}$ developed in [26], where all the workers are assumed to be honest.⁵ For workers who submit all definitive answers, the weight is decreased to a smaller value by adding $\frac{M_N}{2^N(1-m)^N}$ to the denominator of $\frac{1}{(W-M)\mu^n}$. The weight assigned to a worker with all definitive answers, namely, $W_w(N)$ can not be large because it is likely that this worker is a spammer. On the other hand, since it is possible that this worker is honest, $W_w(N)$ can not be too small. Essentially maximization of the candidate score for the correct classification class gives the optimal value of $W_w(N)$, leading to the expression in (8). The larger M_N is, the more likely that a worker with all definitive answers is a spammer. Correspondingly, $W_w(N)$ is smaller.

V. PARAMETER ESTIMATION AND PERFORMANCE ANALYSIS OF ASPT

In this section, we present the parameter estimation technique used in our proposed method. Classification performance and asymptotic performance will also be examined.

A. Parameter Estimation

The FC needs to estimate the crowd parameters μ , m , M_N , M_0 before assigning weights to the workers. Following our previous work [26], either the *Training-based* or the *Majority-voting based* method is adopted to estimate μ . The estimate of m is given by the ratio of the sum of skipped questions and all the questions attempted by the crowd. Since m and μ represent statistical parameters for the honest workers in the crowd, the workers completing or skipping all the questions are not incorporated in the parameter estimation procedure to mitigate the impacts of spammers. The number of Type I and II spammers M_0 and M_N are jointly estimated using the maximum likelihood estimation (MLE) method. G gold standard questions are inserted into the N classification questions, so that a worker responds to a total of $N + G$ questions. After answers from all the workers are collected by the FC, we count the number of workers submitting $N + G$ definitive answers and skipping all the microtasks, denoted by W_{N+G} and W_0 , respectively. Given the numbers of Type I and II spammers M_0 and M_N , the joint PDF of W_{N+G} and W_0 , $f(W_{N+G}, W_0 | M_N, M_0)$, is expressed in (9) shown at the bottom of this page, where \hat{m} is the estimated m , and $\binom{a}{b} = \frac{a!}{(a-b)!b!}$.

Therefore, by the MLE method, the estimates of M_0 and M_N , which are denoted by \hat{M}_0 and \hat{M}_N respectively, can be

⁵ $W - M$ is a constant representing the number of honest workers in our scheme. Here, this constant acts as a scaling parameter.

obtained as

$$\{\hat{M}_N, \hat{M}_0\} = \arg \max_{\{M_N, M_0\} \geq 0} f(W_{N+G}, W_0 | M_N, M_0). \quad (10)$$

By writing $\mathbb{W} = W_0, \dots, W_i, \dots, W_{N+G}$ where W_i is the number of workers submitting i definitive answers for $i = 0, \dots, N + G$, M_N and M_0 can be more accurately estimated according to

$$\{\hat{M}_N, \hat{M}_0\} = \arg \max_{\{M_N, M_0\} \geq 0} f(\mathbb{W} | M_N, M_0). \quad (11)$$

However, the likelihood function $f(\mathbb{W} | M_N, M_0)$ becomes very complicated to compute. As we will see later in the simulation results section, the approach using (10) is sufficient to get a relatively accurate estimate of M_N and M_0 . After the estimates of $\hat{\mu}$, \hat{m} , \hat{M}_N , and \hat{M}_0 are obtained, the FC can assign the appropriate weight to each worker based on (8) and use the answers for aggregation.

B. Performance Analysis

In this section, we assume that there are M_0 Type I spammers and M_N Type II spammers in a crowd population of size W . The spammers attempt to maximize their monetary rewards under the PT model as presented in Section IV. The probability of correct classification P_c is investigated for the weight assignment scheme (8). For simplicity, we assume that the prior probabilities of the true answers for each microtask to be 0 or 1 are equal. Note that we have a correct overall classification only when all the N microtasks are correctly labeled.

Proposition 3: The probability of correct classification P_c in the crowdsourcing system is

$$P_c = \left[\frac{1}{2} + \frac{1}{2} \sum_S \binom{W, M}{\mathbb{G}} (F(\mathbb{G}) - F'(\mathbb{G})) + \frac{1}{4} \sum_{S'} \binom{W, M}{\mathbb{G}} (F(\mathbb{G}) - F'(\mathbb{G})) \right]^N \quad (12)$$

with

$$F(\mathbb{G}) = m^{q_0} \left(\frac{1}{2} \right)^{M_N N} \prod_{n=1}^N (1-\mu)^{q_n} \mu^{q_n} \left(\binom{N-1}{n-1} (1-m)^n m^{N-n} \right)^{q_n + q_n} \quad (13)$$

and

$$F'(\mathbb{G}) = m^{q_0} \left(\frac{1}{2} \right)^{M_N N} \prod_{n=1}^N (1-\mu)^{q_n} \mu^{q_n} \left(\binom{N-1}{n-1} (1-m)^n m^{N-n} \right)^{q_n + q_n} \quad (14)$$

$$f(W_{N+G}, W_0 | M_N, M_0) = \binom{W - M_0 - M_N}{W_0 - M_0} (\hat{m}^{N+G})^{W_0 - M_0} (1 - \hat{m}^{N+G})^{W - W_0 - M_N} \cdot \binom{W - W_0 - M_N}{W_{N+G} - M_N} (1 - \hat{m})^{(N+G)(W_{N+G} - M_N)} (1 - (1 - \hat{m})^{N+G})^{W - W_{N+G} - W_0} \quad (9)$$

where

$$\mathbb{G} = \left\{ (q_{-N}, q_{-N+1}, \dots, q_N, M'_N, M''_N) : \sum_{n=-N}^N q_n = W - M_N - M_0, M'_N + M''_N = M_N \right\}, \quad (15)$$

and q_n , M'_N , and M''_N take values from natural numbers $\{0, 1, \dots\}$,

$$S = \left\{ \mathbb{G} : \sum_{n=1}^N (q_n - q_{-n}) W_w(n) + (M'_N - M''_N) W_w(N) > 0 \right\}, \quad (16)$$

$$S' = \left\{ \mathbb{G} : \sum_{n=1}^N (q_n - q_{-n}) W_w(n) + (M'_N - M''_N) W_w(N) = 0 \right\}, \quad (17)$$

$$\text{and } \binom{W, M}{\mathbb{G}} = \frac{(W - M_0)!}{M'_N! M''_N! \prod_{n=-N}^N q_n!}.$$

Proof: The detailed proof is presented in Appendix D of [46].

C. Asymptotic Performance Analysis

In a practical situation, the number of workers for the crowdsourcing task is relatively large (normally in the hundreds). Then, it is of great value to investigate the asymptotic system performance when W approaches infinity. Here, we give the asymptotic performance characterization for a large crowd, i.e., for a large W .

Proposition 4: As the number of workers W approaches infinity, the probability of correct classification P_c can be expressed as

$$P_c = \left[Q \left(-\frac{\mathcal{M}}{\sqrt{\mathcal{V}}} \right) \right]^N, \quad (19)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$, and \mathcal{M} and \mathcal{V} are given in (18), shown at the bottom of this page, with $Z_M = 2^N (1 - m)^N$.

Proof: The detailed proof is presented in Appendix E of [46].

As stated above, the size of the crowd in practice can be fairly large and the asymptotic result derived in (19) is a good characterization of the actual performance. We further consider that the percentages of Type I and Type II spammers in the crowd are γ and ϵ respectively and give the analysis for the following two cases:

- *Case 1:* $\lim_{W \rightarrow \infty} \frac{M_0}{W} = \gamma > 0$, $\lim_{W \rightarrow \infty} \frac{M_N}{W} = \epsilon = 0$. In this situation, we have $\mathcal{M} = \frac{(2\mu-1)(1-m)}{\mu} \left(\frac{1-m}{\mu} + m \right) +$

$m)^{N-1} + \frac{(2\mu-1)(1-m)^N}{\mu^N}$, which is a constant given m , μ and N . $\mathcal{V} = \frac{1}{W(1-\gamma)} \left(\frac{1-m}{\mu^2} \left(\frac{1-m}{\mu} + m \right)^{N-1} + \frac{(1-m)^N}{\mu^N} - \mathcal{M}^2 \right)$. Note that $W(1-\gamma)$ represents the number of honest workers in the crowd. As there are more honest workers in the crowd, \mathcal{V} becomes smaller and the probability of correct classification becomes larger. The expressions of \mathcal{M} and \mathcal{V} analytically show that as long as the number of honest workers is fixed, the number of Type I spammers have no impact on the system performance.

- *Case 2:* $\lim_{W \rightarrow \infty} \frac{M_0}{W} = \gamma = 0$, $\lim_{W \rightarrow \infty} \frac{M_N}{W} = \epsilon > 0$. We have $\mathcal{M} = \frac{(2\mu-1)(1-m)}{\mu} \left(\frac{1-m}{\mu} + m \right)^{N-1} + \frac{(2\mu-1)(1-m)^N Z_M}{\mu^N Z_M + \frac{\epsilon}{1-\epsilon}}$. As the percentage of Type II spammers in the crowd ϵ increases, \mathcal{M} becomes smaller and the classification performance in terms of P_c deteriorates. In this scenario, it is not easy to show the monotonicity of \mathcal{V} with respect to ϵ and we rely on simulations to show that the probability of correct classification decreases as ϵ is larger.

VI. SIMULATION RESULTS

In the first part of this section, we consider that there are no spammers in the crowd. Simulations are provided to illustrate how PT affect the workers' behavior and system performance.

A. Crowdsourcing Without Spammers

We plot the actual confidence thresholds of cognitively biased crowdworkers t^* with respect to the pre-designed threshold T in Fig. 2 for humans with different behavioral properties. It can be observed that t^* becomes larger as T increases. Since we restrict that $T \geq 0.5$, $\beta T > (1 - T)$ is satisfied. As a result, in Fig. 2(a) we see that t^* becomes larger as β increases and in Fig. 2(b) we see that t^* becomes smaller when α increases. Note that in the green curve in the upper subplot, $\alpha = 0.69$, $\beta = 2.25$ and $\delta = 0.88$ are the mean values of behavioral parameters of the humans from the experiment in [44]. Hence, for this group of population, the green curve represents the average t^* employed by the nominal cognitively biased workers. Since $t^* > T$, we can see that the workers are more likely to use the skip option in practice.

Next, we assume that the confidence t of the crowd workers follows a uniform distribution $U(0.5, x)$ where x is uniformly distributed in $[0.7, 0.9]$. The size of the crowd is $W = 30$ and the confidence threshold T is set equal to 0.6. There are $N = 3$ microtasks and $G = 3$ gold standard questions. In Fig. 3, we plot the system performance in terms of probability of correct

$$\mathcal{M} = \frac{(2\mu-1)(1-m)}{\mu} \left(\frac{1-m}{\mu} + m \right)^{N-1} + \frac{(W-M)(2\mu-1)(1-m)^N Z_M}{(W-M)\mu^N Z_M + M_N}$$

$$\mathcal{V} = \frac{1-m}{(W-M)\mu^2} \left(\frac{1-m}{\mu^2} + m \right)^{N-1} + \frac{((W-M)(1-m)^N + M_N) Z_M^2}{((W-M)\mu^N Z_M + M_N)^2} - \frac{\mathcal{M}^2}{W-M}. \quad (18)$$

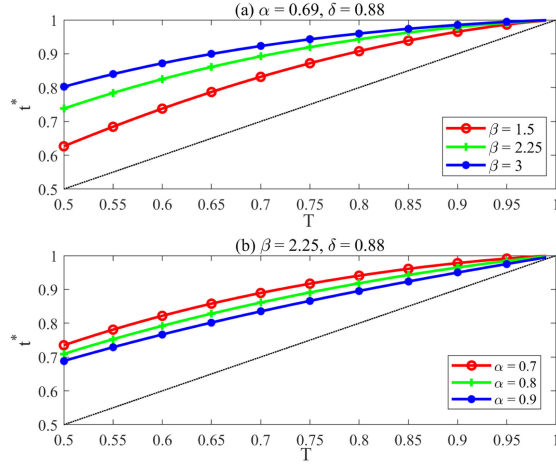


Fig. 2. Confidence thresholds of cognitively biased crowdworkers based on PT.

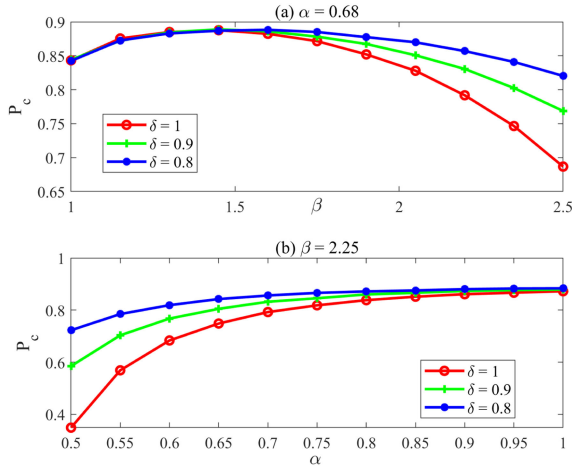


Fig. 3. Classification accuracy when crowd workers have different behavioral parameters.

classification P_c for crowds with different behavioral parameters. If the workers in the crowd are assumed to be rational, i.e., $\alpha = \beta = \delta = 1$, we obtain that $P_c = 0.8445$ and as the behavioral parameters change, P_c has different values. Basically, given the distribution of the workers' confidence t , different behavioral parameters lead to different confidence threshold t^* , which in turn causes variations of the statistical parameters m and μ of the crowd, leading to different classification performances. In the upper subplot where the probability distortion factor $\alpha = 0.68$, P_c first increases and then decreases as the loss aversion parameter β becomes larger. Besides, P_c is higher as the diminishing marginal utility parameter δ has a smaller value. In this case, appropriate values of β and δ counter the probability distortion factor α and improve the system performance. The behavioral parameters α , β and δ jointly determine the overall probability of correct classification P_c . Moreover, we can observe that the best achievable P_c in this subplot is higher than $P_c = 0.8445$ when the crowd workers are assumed to be rational. In the lower subplot where $\beta = 2.25$, P_c monotonically decreases as

TABLE I
ESTIMATES OF M_0 AND M_N

$M_N \backslash M_0$	1	3	5	7	9	11	13	15	17	19
1	1,0	1,3	1,5	1,8	2,9	2,12	1,14	2,15	2,17	2,20
3	3,1	3,2	3,5	4,7	4,9	3,11	3,14	3,15	3,18	3,20
5	5,2	5,3	5,6	6,7	5,9	6,11	6,14	5,17	6,18	5,19
7	7,0	8,4	7,5	8,8	7,10	8,12	7,13	7,17	7,17	8,20
9	9,1	9,4	9,5	10,7	9,9	11,11	9,13	10,15	11,17	9,20
11	11,1	11,5	11,5	12,8	11,6	12,11	11,13	11,16	11,17	12,19
13	13,2	13,6	13,5	14,8	13,9	13,11	14,13	13,16	13,17	14,19
15	15,1	15,3	16,6	16,7	15,9	17,11	15,13	15,15	15,17	15,19
17	17,1	18,4	17,5	17,8	17,9	17,12	18,13	17,16	18,17	18,19
19	20,2	19,2	19,5	19,8	19,9	19,11	19,13	19,16	20,17	21,19
21	21,2	21,3	22,5	21,7	21,9	22,12	21,13	21,15	21,17	21,19
23	23,1	24,3	25,5	23,9	24,9	24,11	25,13	23,16	23,17	23,19
25	26,1	26,3	25,6	25,7	26,9	26,12	25,13	25,15	26,17	25,20

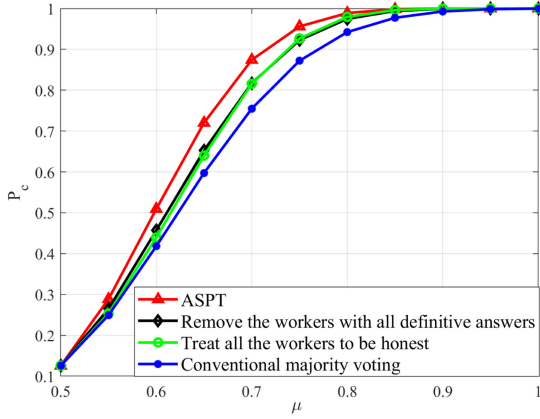
α decreases from 1 to 0.5. Same to the upper subplot, we have P_c become larger as δ has a smaller value.

B. Crowdsourcing in the Presence of Spammers

In this subsection, we present some simulation results to illustrate the advantage of our proposed method ASPT, in which PT is employed to characterize the behavior of spammers. $W = 50$ workers participate in a crowdsourcing task with $N = 3$ micro-tasks and $G = 3$ gold standard questions. $f_p(p)$ is chosen as a uniform distribution $U(0.2, 0.8)$, so that the average probability of a honest worker skipping a task is $m = 0.6$. Let $f_r(r)$ be a uniform distribution expressed as $U(x, 1)$ with $0 \leq x \leq 1$, and thus we can have μ varying from 0.5 to 1.

First, we show the efficiency of our methods for estimating the parameters M_0 and M_N . Table I shows the estimation results of M_0 and M_N , when the true numbers of spammers are $M_0 = \{1, 3, \dots, 19\}$ and $M_N = \{1, 3, \dots, 25\}$. Here, μ is set as 0.75. The estimation process is based on the distribution of the number of workers completing and skipping all the questions W_{N+G} and W_0 , and we can see from the table that most pairs of numbers M_0 and M_N can be exactly estimated, and the estimation errors are at most ± 1 .

We present the performance comparison between different aggregation rules in Fig. 4, where the quality of the crowd μ varies. For illustration, we assume that there are 14 spammers in a crowd of 50 workers, and we have 7 spammers completing all the questions and the other 7 skipping all the questions. As μ increases, we plot the probability of correct classification P_c of four different weight assignment methods. The first one is the ASPT developed in this paper, where we employ PT for modeling the behavior of the spammers. In the second approach, we exclude the workers who submit all definitive answers and treat the remaining workers to be honest. The weights assigned to the honest workers are given by $W_w = \mu^{-n}$ [26]. The third one is where we consider the existence of spammer without incorporating PT, where all the spammers are assumed to be Type I and the FC treats all the workers to be honest, i.e.,


 Fig. 4. Performance comparison with spammers as μ increases.

weight assignment rule is $W_w = \mu^{-n}$ no matter whether the workers submit all definitive answers or not. The last approach is conventional majority voting without a reject option, where all the workers are assigned the same weight. It can be seen in Fig. 4 that at $\mu = 0.5$, all the four curves merge to the same point. It is because when $\mu = 0.5$, even the honest workers are making random guesses like a spammer. In this case, the FC collects no useful information from the crowd and the choice of weight assignment schemes does not make a difference. As the quality of the crowd, μ , improves, the system performance also improves as expected. The proposed ASPT performs better than the method that excludes the workers with all definitive answers and the method that treats all the workers to be honest, which outperform the conventional majority voting approach that does not have a reject option. It should be noted that the second and the third methods have very similar performances. Compared to treating all the workers as honest ones, excluding the workers who submit all definitive answers has the advantage of removing the side effects of Type II spammers. At the same time, however, the second method may also remove the honest workers who submit all definitive answers for decision fusion, leading to a potential deterioration. Hence, this trade-off determines whether the second approach performs better than the third approach or not.

In Fig. 5, we plot the performance comparison when the number of spammers changes. For simplicity, we set that $M_0 = M_N$, and μ is fixed at 0.75. As the number of spammers increases, the classification performance degrades, where the ASPT method gives the best performance. Furthermore, there are two phenomena that need to be discussed:

1) When the number of spammers is small, the conventional majority voting method is outperformed by the one that treats all the workers as honest. However, this is not the case when the number of spammers is large. The reason is that with honest workers, the FC assigns a greater weight to the worker with a larger number of definitive answers. In the regime where M_N is large, which means that the number of spammers completing all the questions is large, the impact from the spammers is much more severe on the performance with such a weight assignment

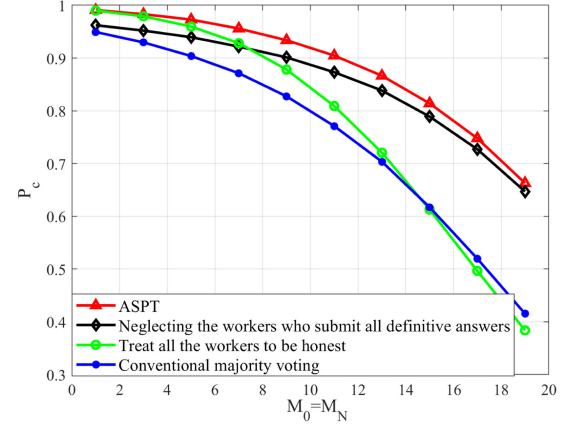
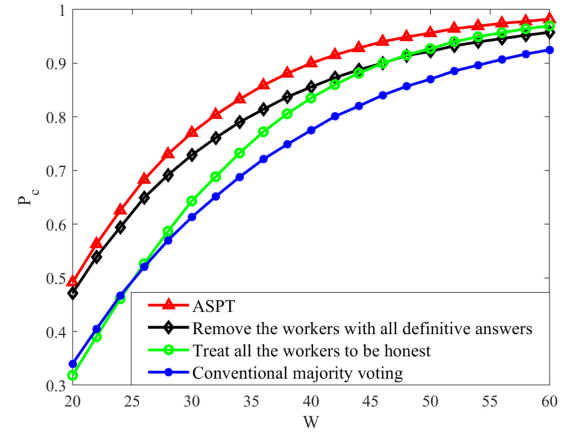


Fig. 5. Performance comparison with different numbers of spammers.


 Fig. 6. Performance comparison as W increases.

scheme. Thus, the corresponding performance degrades significantly.

2) When the number of spammers is small, the method that excludes the workers who submit all definitive answers performs better than the method that treats all the workers to be honest, and vice versa. It can be explained by the fact that in the crowd when the percentage of spammers is small, the second method (the one that excludes the workers who submit all definitive answers) has a smaller probability to remove Type II spammers and has a higher probability to remove honest workers. On the other hand, when the percentage of spammers is large, the probability of excluding Type II spammers is large and that of excluding honest workers is small.

Lastly, we keep the number of spammers in the crowd fixed, with $M_0 = M_N = 7$ and plot the asymptotic performance as the crowd population W increases in Fig. 6. We can see that our proposed approach has the best performance among all the weight assignment strategies. Analogous to the explanations provided for results presented in Fig. 5, we observe that when W is small, i.e., when the percentage of spammers is large, the method that excludes the workers with all definitive answers and the majority voting method outperform the method that treats all the workers to be honest, and vice versa.

VII. CONCLUSION

We have explored a novel framework of crowdsourcing systems to solve classification problems, where the crowd workers may skip a microtask if the confidence regarding the question being asked is lower than a threshold. Our scheme is extremely effective in dealing with spammers as it: (i) exploits the behavioral differences of honest workers and spammers in realistic situations, where the rationality of humans is modeled via PT; (ii) estimates the number of spammers in the crowd and (iii) designs the optimal weight for every worker in the weighted majority voting fusion rule. We provide analytical expressions for probability of correct classification and asymptotic system performance. Compared to rational decision makers, the honest crowd workers and spammers behave in a different manner if PT is incorporated to model their rationality. To accurately characterize the system performance, the behavioral property of the crowd must be taken into consideration. Simulation results illustrate the efficiency of our method compared to other weight assignment schemes that do not include the humans behavior using a prospect theoretic approach.

This work employed a psychologically accurate description of human behavior in crowdsourcing environments. We provided insights in designing strategies to ameliorate the side impacts of human spammers. Our study can also be applied to analyze and model sophisticated human behavior under different payment mechanisms in many applications. Our future work includes the study of task allocation in crowdsourcing considering the behavioral differences of the crowd. As the security issues of distributed inference systems are becoming increasingly important, we also plan to investigate the robustness of the proposed crowdsourcing algorithm against adversarial attacks.

APPENDIX A

Let the perception of the variable reward before answering the question be denoted as Z . According to the payment mechanism (3), $Z = (\mu_{\max} - \mu_{\min})T^G$ before answering any question. In case that the worker has already provided definitive answers to $i = 1, \dots, L$ questions, the expected variable reward is $Z = (\mu_{\max} - \mu_{\min})T^G \prod_{i=1}^L \frac{t_i}{T}$, where t_i is the worker's confidence on the i^{th} question. Hence, we have $Z > 0$.

If the worker decides to answer the question, there is a probability t that the answer is correct and Z is multiplied by $\frac{1}{T}$, and a probability $1 - t$ that the answer is wrong leading to $\bar{Z} = 0$. By using the current expected reward Z as a reference point, and applying the value function to the gains (losses) and the probability weighting function to the probabilities, the subjective payoff if the worker answers this question is expressed as:

$$SP(t) = w(t)v\left(Z\left(\frac{1}{T} - 1\right)\right) + w(1-t)v(-Z) \quad (20)$$

On the other hand, if the worker skips the question, the expected reward Z stays the same and hence, the subjective payoff is 0. The worker makes a decision by choosing the action which yields a higher subjective payoff: $SP \overset{\text{answer}}{\underset{\text{skip}}{\gtrless}} 0$, which yields the result of Theorem 1 after simplification.

APPENDIX B

If a spammer skips g out of G gold standard questions and answers the remaining $G - g$ with random guesses, the expected monetary reward E for the spammer is expressed as

$$\begin{aligned} E &= (\mu_{\max} - \mu_{\min})T^G \prod_{i=1}^G \alpha_{x_i} + \mu_{\min} \\ &= (\mu_{\max} - \mu_{\min})T^G \left(\frac{1}{2}\right)^{G-g} \left(\frac{1}{T}\right)^{G-g} + \mu_{\min} \\ &= (\mu_{\max} - \mu_{\min}) \left(\frac{1}{2}\right)^G (2T)^g + \mu_{\min}, \end{aligned} \quad (21)$$

where $X = \{x_1, \dots, x_G\}$ are the spammer's responses to the gold standard questions. Since $0 \leq g \leq G$, E is maximized as follows

$$\text{if } T < \frac{1}{2} \Rightarrow g = 0, \quad \text{if } T > \frac{1}{2} \Rightarrow g = G. \quad (22)$$

REFERENCES

- [1] D. Tapscott and A. D. Williams, *Wikinomics: How Mass Collaboration Changes Everything*. New York, NY, USA: Portfolio Penguin, 2006.
- [2] J. Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York, NY, USA: Crown Business, 2008.
- [3] D. Tapscott and A. D. Williams, *Macrowikinomics: Rebooting Business and the World*. New York, NY, USA: Portfolio Penguin, 2010.
- [4] M.-C. Yuen, I. King, and K.-S. Leung, "A survey of crowdsourcing systems," in *Proc. IEEE 3rd Int. Conf. Social Comput.*, 2011, pp. 766–773.
- [5] Q. Li and P. K. Varshney, "Does confidence reporting from the crowd benefit crowdsourcing performance?" in *Proc. 2nd Int. Workshop Social Sens.*, 2017, pp. 49–54.
- [6] 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*, Mar./Apr. 2011, pp. 325–326.
- [7] S. Burrows, M. Potthast, and B. Stein, "Paraphrase acquisition via crowdsourcing and machine learning," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, pp. 1–21, Jul. 2013.
- [8] 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, no. 8, pp. 2078–2092, Aug. 2015.
- [9] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on Amazon Mechanical Turk," in *Proc. ACM SIGKDD Workshop Human Comput.*, Jul. 2010, pp. 64–67.
- [10] 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.*, no. 2, pp. 76–81, Mar. 2013.
- [11] K. Mo, E. Zhong, and Q. Yang, "Cross-task crowdsourcing," in *Proc. ACM Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 677–685.
- [12] L. R. Varshney, A. Vempaty, and P. K. Varshney, "Assuring privacy and reliability in crowdsourcing with coding," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2014, pp. 1–6.
- [13] L. R. Varshney, "Participation in crowd systems," in *Proc. 50th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2012, pp. 996–1001.
- [14] M. Hirth, T. Hoffeld, and P. Tran-Gia, "Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms," *Math. Comput. Model.*, vol. 57, no. 11, pp. 2918–2932, Jul. 2013.
- [15] 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, no. 4, pp. 667–679, Aug. 2014.
- [16] 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.*, Aug. 2014, pp. 265–270.
- [17] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Inf. Fusion*, vol. 6, no. 1, pp. 63–81, Mar. 2005.
- [18] D. R. Karger, S. Oh, and D. Shah, "Iterative learning for reliable crowdsourcing systems," in *Proc. Advances Neural Inf. Process. Syst.* 24. Cambridge, MA, USA, Dec. 2011, pp. 1953–1961.

- [19] 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.
- [20] D. Sanchez-Charles, J. Nin, M. Sole, and V. Muntés-Mulero, "Worker ranking determination in crowdsourcing platforms using aggregation functions," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2014, pp. 1801–1808.
- [21] A. J. Quinn and B. B. Bederson, "Human computation: A survey and taxonomy of a growing field," in *Proc. Annu. Conf. Human Factors Comput. Syst.*, May 2011, pp. 1403–1412.
- [22] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. 31st IEEE Conf. Computer Commun.*, Mar. 2012, pp. 2140–2148.
- [23] B. Geng, Q. Li, and P. K. Varshney, "Decision tree design for classification in crowdsourcing systems," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, 2018, pp. 859–863.
- [24] R. Dhar, "Consumer preference for a no-choice option," *J. Consum. Res.*, vol. 24, no. 2, pp. 215–231, Sep. 1997.
- [25] L. R. Varshney, P. Jyothi, and M. Hasegawa-Johnson, "Language coverage for mismatched crowdsourcing," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2016, pp. 1–9.
- [26] Q. Li, A. Vempaty, L. R. Varshney, and P. K. Varshney, "Multi-object classification via crowdsourcing with a reject option," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 1068–1081, Feb. 2017.
- [27] H. J. Jung, Y. Park, and M. Lease, "Predicting next label quality: A time-series model of crowdwork," in *Proc. 2nd AAAI Conf. Human Comput. Crowdsourcing*, Nov. 2014, pp. 87–95.
- [28] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 1, pp. 41–46, Jan. 1970.
- [29] 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.
- [30] J. B. Predd, D. N. Osherson, S. R. Kulkarni, and H. V. Poor, "Aggregating probabilistic forecasts from incoherent and abstaining experts," *Decis. Anal.*, vol. 5, no. 4, pp. 177–189, Jul. 2008.
- [31] K. R. Varshney, "A risk bound for ensemble classification with a reject option," in *Proc. IEEE Workshop Stat. Signal Process.*, Jun. 2011, pp. 769–772.
- [32] N. B. Shah and D. Zhou, "Double or nothing: Multiplicative incentive mechanisms for crowdsourcing," in *Proc. Advances Neural Inf. Process. Syst.*, 2015, pp. 1–9.
- [33] G. Wang *et al.*, "Serf and turf: Crowdturfing for fun and profit," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 679–688.
- [34] G. Demartini, D. E. Difallah, and P. Cudré-Mauroux, "Zencrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 469–478.
- [35] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," *J. Mach. Learn. Res.*, vol. 13, pp. 491–518, 2012.
- [36] Q. Liu, A. T. Ihler, and M. Steyvers, "Scoring workers in crowdsourcing: How many control questions are enough?" in *Proc. Advances Neural Inf. Process. Syst.*, 2013, pp. 1914–1922.
- [37] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor, "Are your participants gaming the system?: Screening mechanical turk workers," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 2399–2402.
- [38] D. E. Difallah, G. Demartini, and P. Cudré-Mauroux, "Pick-a-crowd: Tell me what you like, and i'll tell you what to do," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 367–374.
- [39] O. Alonso and M. Lease, "Crowdsourcing 101: Putting the WSDM of crowds to work for you," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 1–2.
- [40] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica: J. Econometric Soc.*, vol. 47, pp. 263–291, 1979.
- [41] Q. Li and P. K. Varshney, "Optimal crowdsourced classification with a reject option in the presence of spammers," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 4109–4113.
- [42] J. Rocker, C. M. Yauch, S. Yenduri, L. Perkins, and F. Zand, "Paper-based dichotomous key to computer based application for biological identification," *J. Comput. Sci. Coll.*, vol. 22, no. 5, pp. 30–38, May 2007.
- [43] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp. 98–101, Jan. 1986.
- [44] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.
- [45] P. Mongin, "Expected utility theory," *Handbook of Economic Methodology*, vol. 342350. London, U.K.: Edward Elgar, 1997, pp. 234–350.
- [46] B. Geng, Q. Li, and P. K. Varshney, "Prospect theory based crowdsourcing for classification in the presence of spammers," 2019, *arXiv:1909.01463*.