Incentives for Truthful Reporting in Crowdsourcing

(Extended Abstract)

Ece Kamar
Microsoft Research
Redmond, WA 98052
eckamar@microsoft.com

Eric Horvitz
Microsoft Research
Redmond, WA 98052
horvitz@microsoft.com

Categories and Subject Descriptors

I.2 [**Distributed Artificial Intelligence**]: Intelligent agents

General Terms

Design, Economics

Keywords

crowdsourcing systems, peer-prediction rules

1. INTRODUCTION

A challenge with the programmatic access of human talent via crowdsourcing platforms is the specification of incentives and the checking of the quality of contributions. Methodologies for checking quality include providing a payment if the work is approved by the task owner and hiring additional workers to evaluate contributors' work. Both of these approaches place a burden on people and on the organizations commissioning tasks, and may be susceptible to manipulation by workers and task owners. Moreover, neither a task owner nor the task market may know the task well enough to be able to evaluate worker reports. Methodologies for incentivizing workers without external quality checking include rewards based on agreement with a peer worker or with the final output of the system. These approaches are vulnerable to strategic manipulations by workers. Recent experiments on Mechanical Turk have demonstrated the negative influence of manipulations by workers and task owners on crowdsourcing systems [3]. We address this central challenge by introducing incentive mechanisms that promote truthful reporting in crowdsourcing and discourage manipulation by workers and task owners without introducing additional overhead.

We focus on a large class of crowdsourcing tasks that we refer to as *consensus tasks*. Consensus tasks are aimed at determining a single correct answer or a set of correct answers to a question or challenge based on reports collected from workers. These tasks include numerous applications where multiple reports collected from people are used to make decisions. We adapt the *peer prediction rule* [4] to formulate a payment rule that incentivizes workers to contribute to consensus tasks truthfully in crowdsourcing. The rule pays workers depending on how well their report helps to predict another worker's report for the same task. To address several shortcomings of the peer prediction rule, we introduce a novel payment rule, called the *consensus prediction rule*. This payment rule couples payment computations with planning to generate a robust signal for evaluating worker reports. The consensus prediction rule rewards a worker based on how well her report can predict the consensus of other workers. It incentivizes truthful reporting, while providing better fairness than peer prediction rules.

A more detailed presentation of the ideas investigated in this work, including a comparison with existing payment rules, an investigation of considerations in applying payment rules in real-world applications, and a detailed empirical evaluation can be found in [2].

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

2. SOLVING CONSENSUS TASKS

A task is a consensus task if it has a correct answer, has access to a population of workers who are able to share assessments about the correct answer, and where a worker's inference is stochastically relevant to the assessment of a randomly selected worker. The goal of a consensus-centric crowdsourcing system is to deduce an accurate prediction of the correct answer of a task by making use of multiple worker reports.

Let us assume that a crowdsourcing system has access to inferential models that can be used to predict the correct answer, to make hiring decisions, and to calculate payments. These models include an *answer model* (M_A) and a *report model* (M_R). $M_A(a, f)$ is the probability of the correct answer being a given the feature set of the task (f). $M_R(r_i, a^*, f_i)$ is the probability of worker i reporting r_i, given that the correct answer of the task is a^* and the set of features relevant to the worker report is f_i. At each point during execution, the system makes a decision about whether to hire a worker randomly from the worker population, or to terminate the task. When the system decides to not hire additional workers, it provides a final consensus answer $\hat{a}$ based on aggregated worker reports and delivers this answer to the owner of the task. Let π be the policy for making hiring decisions. We define a function M_π such that for a given sequence of worker reports r and feature set f, $M_\pi(r, f)$ is $\emptyset$ if π does not terminate after receiving r, and is $\hat{a}$, the consensus answer, otherwise.

Detailed investigations of learning answer and report models and policies for consensus tasks have been presented separately [1].

3. PAYMENT RULES

We now present payment rules that ensure that truthful reporting is an equilibrium of a consensus task. We start by presenting definitions and assumptions that are needed to formalize payment rules for consensus tasks.

In consensus tasks, workers report on a task once and maximize their individual utilities for the current task. We make the assumptions that the probability assessments performed by models M_A and M_R are accurate and common knowledge. $\tau_i(r_i, r_{-i}) \rightarrow \bar{\mathbb{R}}$ denotes the system's payment to worker i, based on r_i, worker i's report, and r_{-i}, a sequence of reports collected for the same task excluding r_i. Ω_R is the domain of worker inferences and reports. Let s_i^t be a reporting strategy of worker i such that for all possible inferences c_i she can make for task t, $s_i^t(c_i \in \Omega_R) \rightarrow r_i \in \Omega_R$. A strategy s_i^t is truth-revealing if for all $c_i \in \Omega_R$, $s_i^t(c_i) = c_i$. $\mathcal{M} = (t, \pi, \tau)$, mechanism for task t with policy π and payment rule τ, is strict Bayesian-Nash incentive compatible if truth-revelation is a strict Bayesian-Nash equilibrium of the task setting induced by the mechanism.

S is a proper scoring rule for the forecast of a categorical random variable with domain Ω. It takes as input p, probability vector over Ω, and $\omega_i \in \Omega$, the realized outcome of the variable, and it outputs a reward in $\bar{\mathbb{R}}$. The expected reward is maximized if the reported forecast agrees with the true forecast.

3.1 Peer Prediction Rule

The *peer prediction rule* is an adaptation of the rule proposed by Miller et. al. to the domain of crowdsourcing. It rewards a worker based on how well her report can predict the report of another worker.

PROPOSITION 1. *For a given consensus task t and policy π, let r_j be the report of a random worker from I_{-i}. $M = (t, \pi, \tau^p)$ is strict Bayesian-Nash incentive compatible, where worker i's payment, τ_i^p, for reporting to task t is,*

$$\tau_i^p(r_i, r_j) = S(p^p, r_j), \text{ where}$$
$$\text{for all } r_k \in \Omega_R,\ p_k^p = Pr_f(C_j = r_k | C_i = r_i)$$

In the equilibrium when all workers report their true inference, $Pr_f(C_j|C_i)$ can be computed by applying Bayes rule and by making use of answer and report models presented in Section 2.

$$Pr_f(C_j = r_j | C_i = r_i) = \frac{\sum_{a \in A} M_A(a, f) M_R(r_i, a, f_i) M_R(r_j, a, f_j)}{\sum_{a \in A} M_A(a, f) M_R(r_i, a, f_i)}$$

3.2 Consensus Prediction Rule

Despite its incentive compatibility properties, the peer prediction payments may not be fair in the way it rewards workers. A worker reporting correctly may receive a low payment if paired with a worker reporting incorrectly. We now present a novel incentive compatible payment rule that provides higher levels of fairness. The *consensus prediction rule* rewards a worker according to how well her report can predict the outcome of the system (i.e., the consensus answer that will be decided by the system), *if she were not participating*. This payment rule forms a direct link between a worker's payment and the outcome of this system. Because the outcome of a successful system is more robust to erroneous reports than the signal used in peer prediction rules, this payment rule has better fairness properties.

Let $\hat{A}_{-i}$ be a random variable for the consensus answer decided by the system if the system runs without access to worker i. An inference of a worker provides evidence about the task, its correct answer, and other workers' inferences, which are used to predict a value for $\hat{A}_{-i}$. Thus, it is realistic to assume that a worker's inference is stochastically relevant for $\hat{A}_{-i}$, given feature set f.

PROPOSITION 2. *For a given consensus task t and policy π, let $\hat{a}_{-i}$ be the consensus answer predicted based on r_{-i}. $M = (t, \pi, \tau^c)$ is strict Bayesian-Nash incentive compatible for any worker i, where*

$$\tau_i^c(r_i, r_{-i}) = S(p^c, \hat{a}_{-i}), \text{ where}$$
$$\text{for all } a_k \in A,\ p_k^c = Pr_f(\hat{A}_{-i} = a_k | C_i = r_i)$$

Next, we demonstrate how payments can be calculated with the consensus prediction rule for consensus tasks in the equilibrium when all workers report their true inferences. The calculation of τ_i^c payments is a two-step process; generating a forecast about $\hat{A}_{-i}$ based on worker i's report, and calculating a value for $\hat{a}_{-i}$ based on r_{-i}.

To generate a forecast for $\hat{A}_{-i}$, we simulate consensus system for $L_\emptyset$, the set of all possible sequences of worker reports that reach a consensus about the correct answer.

$$Pr_f(\hat{A}_{-i} = a | C_i = r_i) = \sum_{r' \in L_\emptyset} Pr_f(r'|r_i)\, \mathbf{1}_{\{a\}}(M_\pi(r', f))$$

$$Pr_f(r'|r_i) \propto \sum_{a^* \in A} M_A(a^*, f)\, M_R(r_i, a^*, f_i) \prod_{r_l \in r'} M_R(r_l, a^*, f_l)$$

The second step of τ_i^c calculation is predicting the realized value for $\hat{A}_{-i}$ based on r_{-i}, the actual set of reports collected from workers excluding worker i. Doing so requires simulating $L_{r_{-i}}$, the set of all report sequences that start with r_{-i} and reach a consensus on the correct answer as follows:

$$\hat{a}_{-i} = \underset{a \in A}{\operatorname{argmax}} \sum_{r' \in L_{r_{-i}}} Pr_f(r'|r_{-i})\, \mathbf{1}_{\{a\}}(M_\pi(r', f))$$

4. FUTURE WORK AND CONCLUSIONS

We presented an approach to developing truthful and fair incentive mechanisms for crowdsourcing. Future work includes exploring new approaches for relaxing assumptions of common knowledge and designing truthful incentive mechanisms for a larger variety of tasks. We believe that the use of truthful and fair mechanisms promises to enhance the operation of crowdsourcing for both task authors and contributors, and can promote the wider use of such systems as a trusted methodology for problem solving.

5. REFERENCES

[1] E. Kamar, S. Hacker, and E. Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *AAMAS*, 2012.

[2] E. Kamar and E. Horvitz. Incentives and truthful reporting in consensus-centric crowdsourcing. Technical report, MSR-TR-2012-16, Microsoft Research, 2012.

[3] A. Kittur, E. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *SIGCHI*, 2008.

[4] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, pages 1359–1373, 2005.