Incentivized Information Fusion with Social Sensors

Sujay Bhatt[*]
Dept. of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
sh2376@cornell.edu

Vikram Krishnamurthy
Dept. of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
vikramk@cornell.edu

ABSTRACT

This paper deals with the problem of incentivized information fusion, where a controller seeks to infer an unknown parameter by incentivizing a network of social sensors to reveal the information. The social sensors gather information on the parameter after interacting with other social sensors, to optimize a local utility function.

We are interested in finding incentive rules that are easy to compute and implement. In particular, we give sufficient conditions on the model parameters under which the optimal rule for the controller is provably a threshold decision rule, i.e, don't incentivize when the estimate (of the parameter) is below a certain level and incentivize otherwise.

We will further provide a complete sample path characterization of the optimal incentive rule, i.e, the nature (average trend) of the optimal incentive sequence resulting from the controller employing the optimal threshold rule. We show that the optimal incentive sequence is a sub-martingale, i.e, the optimal incentives increase on average over time.

1. INTRODUCTION

Autonomous decision making involves information acquisition and identifying a course of action that optimizes an objective function. To gather the required information, an autonomous system (controller) can avail the services of an information network or a multi-agent sensor network. For example, e-commerce companies like Amazon gather ratings and reviews from users (sensors) regarding the services and products listed on their websites; where the social network serves as the information network. The sensors that are a part of the information network exhaust their own resources in acquiring the information, for example, the customers who share the experiences on Amazon are those who have purchased the product. Thus, there is a need to design autonomous systems that suitably compensate the sensors for information, to motivate their participation by considering that the sensors in the information network may learn and influence each other. We term this problem as *incentivized information fusion*, where a controller seeks to infer an unknown parameter by incentivizing a network of social sensors to reveal the information they gather about the underlying parameter.

Sensors having the following attributes will be called *Social Sensors*:

i.) They affect the behaviour of other sensors.
ii.) They share quantized information (decisions/actions) and have their own dynamics.
iii.) They have limited processing capabilities - boundedness.
iv.) They are rational - they fuse all available information using Bayes' rule to take an action that maximizes the reward.

A social sensor (human) provides information about its state (sentiment, social situation, quality of product, label) to a social network after performing social learning. Social learning is the process by which the sensors learn and influence each other. The availability of review platforms like Yelp, Expedia, Amazon etc, facilitates social learning; see [8] for an empirical study of peer effects on consumption, and [10] for a more formal treatment.

In this paper, we present a model of Bayesian social learning with focus on the interaction between the controller and the multi-agent network of social sensors.

The contribution of this paper is two fold:

1.) **Threshold Incentive Rule**: We study the interaction of a controller and a network of social sensors, where the controller can modify the utility function of the sensors which in turn affects their decisions. We give sufficient conditions on the model parameters - that are intuitive - under which the optimal rule for the controller is provably a threshold decision rule, i.e, don't incentivize when the estimate (of the parameter) is below a certain level and incentivize otherwise.
2.) **Sub-martingale Property**: We provide a complete sample path characterization of the optimal incentive rule, i.e, the nature (trend) of the optimal incentive sequence over time. We show that the optimal incentive sequence for the controller is a sub-martingale, i.e, the optimal incentives increase on average over time.

Related Literature

[5] considers the interaction of a global controller and a network of social sensors, where the sensors perform social learning to estimate an underlying parameter and to optimize a local utility function. The objective of the controller is to detect a change in the parameter as soon as possible by observing the actions of the sensors. In this paper, unlike [5],

[*]This material is based upon work supported, in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under grant 12346080, National Science Foundation under grant 1714180, and Schmidt Sciences.

IFIP WG 7.3 Performance 2017. Nov. 14-16, 2017, New York, NY USA

the controller can influence the behaviour of the sensors by modifying their utility function.

Crowdsourcing large scale problems like image classification [11], annotation [9], recommendation, peer prediction for eliciting truthful and informative feedback; *where the sensors (Turkers) are allowed to interact and influence each other*, are some of the applications that can be analyzed using framework in this paper. [4] considers a general model of crowdsourcing and the problem of minimizing the total incentive that must be paid to achieve a target overall reliability. The problem considered in this paper can be seen as an extension of [4], in that we also allow the possibility of social learning between the workers (Turkers). Our work is similar in flavor to *Peer Prediction* [7,12], where reward based scoring schemes are devised to elicit informative feedback; but a key difference is that there is Bayesian social learning.

Organization

Sec. 2 details the Bayesian social learning model for the process of information fusion by the social sensors and the incentivization protocol employed by the controller. Sec. 3 formulates the stochastic control problem faced by the controller as a POMDP and is solved using dynamic programming. The structure of the value function and optimal decision rule is completely characterized for the controller objectives.
Sec. 4 describes the nature of the incentive sequence that is input to the multi-agent network. It is shown that the controller offers smaller incentives initially and subsequently increases it to learn the true parameter.

2. SOCIAL LEARNING MODEL AND CONTROLLER OBJECTIVE

2.1 Social Learning Model and Incentive Protocol

The following illustrates an example application of the problem considered in this paper. Travel website companies like TripAdvisor and e-commerce companies like Amazon, offer monetary compensation for people (social sensors) to reveal truthful or honest information (in the form of reviews) regarding the services and products listed on their websites. The social sensors want to maximize their utility - a function of the compensation, experience with the product or services, and the experience of other sensors. The controller (TripAdvisor or Amazon) wants to minimize the expected payout for gathering truthful information from the social sensors.

Let $k = 1, 2, \cdots$ denote the discrete time instant when sensor k acts. Let $x \in \mathcal{X} = \{1, 2\}$ denote the underlying parameter, which is assumed to be a random variable[1]. The network of social sensors and the controller, seek to estimate the realization of the random variable. Let

$$\pi_{k-1}(i) = \mathbb{P}(x = i | a_1, \ldots, a_{k-1})$$

denote the estimate of the parameter (termed as public belief) at time $k-1$, and let the initial estimate be denoted as $\pi_0 = (\pi_0(i), i \in \mathcal{X})$, where $\pi_0(i) = \mathbb{P}(x_0 = i)$. Let the belief space, i.e, the set of distributions π over the parameter be denoted as

$$\Pi(2) \triangleq \{\pi \in \mathbb{R}^2 : \pi(1) + \pi(2) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2\}\}.$$

2.2 Controlled Information Fusion Objective

The controller wants to estimate the underlying parameter x by providing incentives to the social sensors in exchange for information. The controller considers the following objective function:

$$J_\mu(\pi) = \mathbb{E}_\mu\{\sum_{k=1}^{\infty} \rho^k c(p_k) | \pi_0 = \pi\}. \quad (1)$$

Here p_k denotes the incentive, $\rho \in [0, 1)$ denotes the economic discount factor indicating the degree of impatience of the controller, π denotes the estimate of the parameter conditioned on the decisions of the social sensors - termed as *public belief*, $c(p_k)$ denotes the cost of information fusion incurred by the controller, and μ denotes the decision rule for the controller that maps π to an incentive $p \in [0, 1]$.
The controller seeks to find the optimal incentive rule μ^* such that

$$J_{\mu^*}(\pi_0) = \inf_{\mu \in \boldsymbol{\mu}} J_\mu(\pi_0). \quad (2)$$

2.3 Social Learning Model and Incentive Protocol

We will model the dynamics of the social sensors (observation model and decision rule optimization), the information fusion cost that models the cost of information acquisition for the controller, the dynamics of the parameter estimate computed from the decisions of the social sensors (public belief), and incentive rules that associate parameter estimates with incentives to be provided.
We consider the classical sequential social learning framework [2, 5]. The decision of social sensor $a_k \in \mathcal{A} = \{1, 2\}$ depends on the decisions of the first $k - 1$ sensors and its own estimate of the parameter (termed private belief) after receiving noisy private valuations, $y_k \in \mathcal{Y} = \{1, 2\}$, of the parameter x. The controller estimates the parameter by providing incentives $p_k \in [0, 1]$ at each time k (or at each sensor k). It is assumed that each sensor decides once[2] in a predetermined sequential order indexed by k.

Social Sensor Dynamics

1. **Social Sensor's Private Observation**: Sensor k's private observation denoted by $y_k \in \mathcal{Y} = \{1, 2\}$ is a noisy measurement of the true parameter. It is obtained from the observation likelihood distribution as,

$$B_{ij} = \mathbb{P}(y_k = j | x = i). \quad (3)$$

   The discreteness of the observation distribution captures the *boundedness* or the limited processing capabilities of the sensor.

[1] For example x could denote {Bad quality, Good quality} for a product or service, {Object absent, Object Present} in case of a simple annotation problem, etc.

[2] The model and the results presented in the paper can easily be interpreted in the case where more than one sensor acts every period. Consider the situation where multiple homogenous sensors act at time k. A naive approach to map to the model considered in this paper is to use Majority voting and label the majority vote as a. Majority voting simply chooses what the majority of sensors agree on. At time $k+1$, a group of sensors decide by considering the Majority decisions at times $k < k + 1$. The controller updates the knowledge about the underlying parameter based on the Majority decisions.

2. **Social Learning and Private Belief update**: Information on the parameter conditioned on the private valuation is the private belief (η^y), and is computed using Bayes' rule. Sensor k updates its private belief by fusion of the observation y_k and the prior public belief π_{k-1} as the following Bayesian update

$$\eta^{y_k} = \frac{B_{y_k}\pi_{k-1}}{\mathbf{1}' B_{y_k}\pi_{k-1}} \tag{4}$$

where B_{y_k} denotes the diagonal matrix

$$\begin{bmatrix} \mathbb{P}(y_k|x = \ 1) & 0 \\ 0 & \mathbb{P}(y_k|x = 2)) \end{bmatrix}$$

and $\mathbf{1}$ denotes the 2-dimensional vector of ones.

3. **Social Sensor's Action**: Sensor k executes an action $a_k \in \mathcal{A} = \{1, 2\}$ myopically to maximize the reward. Let $r(x, y, a_k)$ denote the reward accrued if the sensor takes action a_k when the underlying parameter is x and the observation is y.

In this paper, we assume for simplicity that all social sensors have the same reward function $r(x, y, a)$ and we choose this as

$$r(x, y, a) = \delta_a p - \alpha_a \mathcal{I}(a \neq x) - \beta_a \mathcal{I}(a \neq y) - \gamma_a \tag{5}$$

where $\delta_a, \alpha_a, \beta_a, \gamma_a \in [0, 1]$. The form is inspired by the quasi-linear utility in [1]. Here, δ_a is interpreted as the fraction of the monetary compensation p received, α_a and β_a are the losses incurred for not taking appropriate actions, and γ_a is the cost incurred in obtaining observation.

The sensor chooses an action a_k to maximize the reward as

$$a_k = \arg\max_{a \in \mathcal{A}} r_a' \eta^{y_k} \tag{6}$$

where

$$r_a = [r(1, a) \ r(2, a)] \text{ and } r(x, a) = \sum_{j=1}^{2} r(x, y = j, a) B_{xj}. \tag{7}$$

Remark. For example, consider the situation where an e-commerce website like Amazon is soliciting honest customer reviews from those who have purchased a particular product.

Here $a_k \in \mathcal{A} = \{1(\text{Bad Review}), 2(\text{Good Review})\}$, $\delta_a p$ could indicate the compensation in exchange for the review, α_a denotes the cost incurred for making a decision not appropriate for the product quality, β_a denotes the cost incurred for making a decision not appropriate for the information gathered on the product quality, and γ_a denotes the cost of acquiring information regarding the product quality before the purchase.

Remark. It is assumed that $\delta_2 > \delta_1$, $\alpha_1 > \alpha_2$, $\beta_1 > \beta_2$, $\gamma_2 > \gamma_1$. These assumptions are intuitive, for example if Amazon is soliciting honest reviews, $\delta_2 > \delta_1$ models higher compensation offered to the sensor to write a high quality review; $\gamma_2 > \gamma_1$ models higher cost invested by the sensor in information acquisition to write a high quality review; $\alpha_1 > \alpha_2$ and $\beta_1 > \beta_2$ model higher cost for writing a bad review when the quality/ information indicates otherwise.

Information Fusion cost

The controller wants to maximize the number of sensors that act according to their evaluations/observations[3] while minimizing the following cost:

Expenditure: The controller offers a compensation to receive truthful accounts of the information gathered by the social sensors, i.e,

$$c(p_k) = p_k - \phi_r \mathcal{I}(a_k = y_k|\pi_{k-1}) \tag{8}$$

where $\mathcal{I}$ denotes the indicator function, $\phi_r \in (0, 1)$ denotes the benefit from truthful information gathering and p denotes the expenditure.

Public Belief Dynamics

Information on the parameter conditioned on the new action is the public belief (π), and is computed using Bayes' rule. Sensor k's action is shared by the controller with the multi-agent network and the public belief on the quality is updated according to the social learning Bayesian filter (see [5]) as follows

$$\pi_k = T^{\pi}(\pi_{k-1}, a_k) = \frac{R_{a_k}^{\pi_{k-1}} \pi_{k-1}}{\mathbf{1}' R_{a_k}^{\pi_{k-1}} \pi_{k-1}}. \tag{9}$$

Here, $R_{a_k}^{\pi_{k-1}} = \text{diag}(\mathbb{P}(a_k|x = i, \pi_{k-1}), i \in \mathcal{X})$ is the decision or action likelihood matrix, where $\mathbb{P}(a_k|x = i, \pi_{k-1}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(a_k|y, \pi_{k-1})\mathbb{P}(y|x = i)$ and

$$\mathbb{P}(a_k|y, \pi_{k-1}) = \begin{cases} 1 & \text{if } a_k = \arg\max_{a \in \mathcal{A}} r_a' \eta^{y_k}; \\ 0 & \text{otherwise.} \end{cases}$$

Note that π_k belongs to the unit simplex $\Pi(2) \triangleq \{\pi \in \mathbb{R}^2 : \pi(1) + \pi(2) = 1, 0 \leq \pi(i) \leq 1 \text{ for } i \in \{1, 2\}\}$.

Information Fusion Incentive

The controller fuses (aggregates) the information by providing incentives to the social sensors. The history of past incentives and decisions $\mathcal{H}_k = \{\pi_0, p_1, \cdots, p_k, a_k\}$ is recorded by the controller and the multi-agent network. The controller chooses

$$p_{k+1} = \ \mu_{k+1}(\mathcal{H}_k) \in [0, 1] \tag{10}$$

for the sensor $k + 1$. Here μ_{k+1} denotes the decision rule at time $k + 1$ that associates the history $\mathcal{H}_k$ with an incentive p_{k+1}. Since $\mathcal{H}_k$ is increasing with time k, to implement a controller, it is useful to obtain a sufficient statistic that does not grow in dimension. The public belief π_k computed via the social learning filter (9) forms a sufficient statistic for $\mathcal{H}_k$ and the incentive in (10) is given as

$$p_{k+1} = \mu_{k+1}(\pi_k). \tag{11}$$

3. STRUCTURE OF OPTIMAL INCENTIVE RULE

The stochastic control problem faced by the controller is equivalent to a partially observed Markov decision process (POMDP) with dynamics (9) and objective (2), and is

[3] This is consistent with the objective of truthful information reporting in Peer Prediction literature; see [7]. Acting according to self-valuations or observations also improves the quality of social learning as the decisions are informative (in the Blackwell sense, see [6]); see also Footnote 4.

solved using dynamic programming. In this section, we give sufficient conditions under which the optimal incentive rule for the controller is provably a threshold rule.

Assumptions

(A1) The observation distribution $B_{xy} = \mathbb{P}(y|x)$ is TP2, i.e, the determinant of the matrix B is non-negative.

(A2) The reward vector r_a is supermodular, i.e, $r(1,1) > r(2,1)$ and $r(2,2) > r(1,2)$ for every $p \in [0,1]$.

Discussion of the Assumptions

1. Assumption (A1) is on the underlying stochastic model, and enables the posteriors to be compared. The observation distribution being TP2 (total positive of order 2) implies that for higher parameter values, the probability of receiving higher valuations is higher than for lower values.
2. (A2) is required for the problem to be non-trivial. If it doesn't hold and $r(i,1) > r(i,2)$ for $i = 1,2$, then $a = 1$ always dominates $a = 2$; the sensors provide no useful information.

The optimal incentive rule μ^* and the optimal cost (value function) $V(\pi)$ for the POMDP satisfy the Bellman's dynamic programming equation

$$Q(\pi,p) = c(p) + \rho \sum_{a \in \mathcal{A}} V(T^{\pi}(\pi,a))\sigma(\pi,a), \qquad (12)$$

$$\mu^*(\pi) = \operatorname*{argmin}_{p \in [0,1]} Q(\pi,p),$$

$$V(\pi) = \min_{p \in [0,1]} Q(\pi,p), \quad J_{\mu^*}(\pi_0) = V(\pi_0).$$

Main Result: 1 (Threshold Incentive Rule)

The theorem shows that the optimal incentive rule is threshold; i.e, belief space $\Pi(2)$ is divided to two connected regions. In one of the regions, it is optimal to not provide any incentives; and in the other, it is optimal to incentivize using a well defined deterministic function of the parameter estimate. Define the following function:

$$\Delta(\eta^y) = [l_1 \ \ - l_2] \frac{B_y \pi}{\mathbf{1}' B_y \pi} + l_3 \qquad (13)$$

where η^y is the private belief update (4), $\pi = \begin{bmatrix} 1 - \pi(2) \\ \pi(2) \end{bmatrix}$,

$$l_1 = \frac{\alpha_2 + \beta_2 B_{11} - \beta_1 B_{12}}{\delta_2 - \delta_1}, \ l_2 = \frac{\alpha_1 - \beta_2 B_{21} + \beta_1 B_{22}}{\delta_2 - \delta_1},$$

$l_3 = \frac{\gamma_2 - \gamma_1}{\delta_2 - \delta_1}$, and $\alpha, \beta, \delta, \gamma$ are as in (7). The function $\Delta(\eta^y)$ will henceforth be referred to as the *incentive function.*

Theorem 1. *Let $\phi_r \in (0,1)$. Under (A1) and (A2), the optimal incentive rule for the controller $\mu^*(\pi) = \operatorname{argmin}_p Q(\pi,p)$ is*

$$\mu^*(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi_r^*(2)); \\ \Delta(\eta^{y=2}) & \text{if } \pi(2) \in [\pi_r^*(2), 1]. \end{cases} \qquad (14)$$

where $\pi_r^*(2) \in (0,1)$.

Discussion of Main Result

The optimal incentive rule $p = \mu^*(\pi)$ seeks to maximize the number of sensors that act according to their observations - to improve the quality of social learning[4] - while minimizing the payout. According to Theorem 1, the regions in the belief space $\Pi(2)$ where it is optimal choose $p = \Delta(\eta^{y=2})$ and $p = 0$ are connected and convex. Therefore, computing the optimal decision rule amounts to finding the belief π^*, below[5] which it is optimal not to provide any compensation $p = 0$; and above which it is optimal to compensate $\Delta(\eta^{y=2})$ at every belief, to minimize the cost.

The practical usefulness of the threshold incentive rule in the theorem stems from the following: (i) the search space of incentive rules μ reduces from an infinite class of functions (over $\Pi(2)$) to those that have a simple threshold structure as in Fig.1; (ii) to compute the optimal incentive rule μ^*, one can compute the belief π^* offline. This is advantageous as the control problem (POMDP) has a continuous state space $\Pi(2)$, and computing optimal policies involves a PSPACE hard dynamic programming recursion offline; (iii) at each instant (or belief) the controller only needs to decide between $\Delta(\eta^{y=2})$ and $p = 0$ depending on the threshold rule (14); (iv) having connected regions reflects the confidence of the controller in implementing the incentive schemes– if its is optimal to not incentivize at a particular belief, it is optimal to discontinue incentivization when the belief is lower.

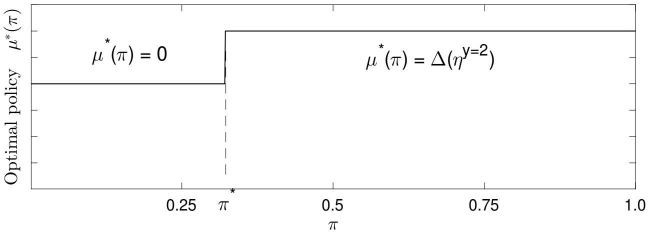

Figure 1: Expenditure Minimization: Optimal fusion rule for the controller. When $\pi < \pi^*$, it is optimal not to incentivize $\mu^*(\pi) = 0$, and when $\pi > \pi^*$, the controller incentivizes using the incentive function $\mu^*(\pi) = \Delta(\eta^y)$. The following parameters were chosen: $\phi_r = 0.3$ and discount factor $\rho = 0.7$. Let e_1 and e_2 denote indicator vectors. For $\delta_1 = 0.3$, $\delta_2 = 0.95$, and $B = \left(\begin{smallmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{smallmatrix}\right)$, the following parameters were obtained as a solution of $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$ for the reward vector: $\alpha_1 = 0.288, \alpha_2 = 0.278, \beta_1 = 0.11, \beta_2 = 0.1, \gamma_1 = 0.1, \gamma_2 = 0.414$.

4. STOCHASTIC PROPERTIES OF THE OPTIMAL INCENTIVE SEQUENCE

In this section, we will describe the relation between the optimal decision rule in Theorem 1, and the incentive sequence $p_k = \mu^*(\pi_{k-1})$, for $k = 1,2,\ldots$ In particular, we provide a complete sample path characterization of the incentive sequence (over) that results from the controller applying the optimal incentive rule.

Main Result: 2 (Sub-martingale Property)

The theorem gives a complete sample path characterization

[4] Acting according to self valuations or private observations is equivalent to truthful information reporting, and is informative. Since the sensors take into account the actions or decisions of the preceding sensors, fusion of informative decisions lead to improved estimate of the parameter and hence improves social learning.

[5] $\pi_2 \geq \pi_1$ if the determinant

$$\begin{vmatrix} \pi_1(1) & \pi_1(2) \\ \pi_2(1) & \pi_2(2) \end{vmatrix} \geq 0$$

of optimal decision rule implemented by the controller. It is shown that when the controller wants to minimize the expected payout for gathering truthful information, the incentive sequence is a sub-martingale[6]; i.e, it increases on average[7] over time.

Theorem 2. *Let π_0 denote the initial belief. Under (A1), the optimal incentive sequence $p_k = \mu^*(\pi_{k-1})$ for the information fusion problem is a sub-martingale.*

Discussion of Main Result

According to Theorem 2, the optimal decision rule of the controller is such that the sample path of the incentive sequence displays an increasing trend, i.e, the incentives increase on average over time.

The usefulness of the theorem stems from the following: (i) it gives a complete sample path characterization of the optimal incentive rule implemented by the controller; (ii) the sub-martingale property assures that that larger beliefs and higher compensations are concomitants of high parameter values.

The controller starts compensating with a low incentive initially to learn about the quality of the product or service. If the quality looks promising, it gradually increases the compensation to encourage the sensors to act according to their assessments. For example, when Amazon is soliciting honest reviews, gradual increase in the compensation when the quality looks promising will lead to good quality reviews and this in turn will increase the sales of the product or services in the future.

5. CONCLUSION

This paper considered the problem of incentivized information fusion, where a controller compensates a network of social sensors in exchange for information on an underlying parameter. It was shown that the optimal fusion/decision rule for the controller is a threshold rule; and a sample path characterization of the optimal incentive rule employed by the controller was also provided. In particular, it was shown the when the controller wants to maximize the number of social sensors that report truthfully, the sequence of incentives should display an increasing trend on average; i.e, it's a sub-martingale.

Crowdsourcing large scale problems like image annotation, data labeling, recommendation, peer prediction for eliciting truthful and informative feedback; where the sensors (Turkers) are allowed to interact and influence each other, are some of the applications that can be analyzed using framework in this paper.

APPENDIX

Preliminaries and Definitions:

Definition 1 (Submodular function [6])**.** $\nu : \Pi(2) \times \mathbb{A} \to \mathbb{R}$ is submodular in (π, a) if

$$\nu(\pi, a) - \nu(\pi, \bar{a}) \geq \nu(\bar{\pi}, a) - \nu(\bar{\pi}, \bar{a}) \qquad (15)$$

for $a > \bar{a}$ and[8] $\bar{\pi} \geq \pi$.

Define

$$\mathcal{F}_k : \sigma - \text{algebra generated by } (\pi_0, p_1, a_1, \ldots, p_k, a_k).$$

Definition 2 (Martingale)**.** ([3]) Let $\mathcal{F}_k$ denote the sigma algebra. A sequence X_k such that $\mathbb{E}[|X_k|] < \infty$ is a martingale (with respect to $\mathcal{F}_k$) if

$$\mathbb{E}[X_{k+1}|\mathcal{F}_k] = X_k, \text{ for all } k.$$

If $\mathbb{E}[X_{k+1}|\mathcal{F}_k] \geq X_k,$ for all $k.$, the sequence X_k is a *sub-martingale.*

Definition 3. ([3]) A sequence H_k is said to be a predictable sequence if $H_k \in \mathcal{F}_{k-1}$.

Theorem 3 ([6])**.** *Let the parameter be a random variable and (A1) and (A3) hold. For $p \in [0, 1]$, the belief space $\Pi(2)$ can be partitioned into at most 3 non-empty regions such that, the sensor decision likelihood matrix R^π in (9) is a constant with respect to the belief parameter π and is given as*

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \textit{ and } \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Theorem 4 ([6], Chapter 8)**.** *Consider a POMDP with possibly continuous-valued observations. Assume that for each action p, the instantaneous costs $c(\pi, p)$ are decreasing with respect to $\pi \in \Pi(2)$. Under (A1), the value function $V(\pi)$ is decreasing in π.*

Theorem 5 ([6], Chapter 12)**.** *Consider a POMDP with possibly continuous-valued observations. Assume that for each action p, the instantaneous costs $c(\pi, p)$ are decreasing with respect to $\pi \in \Pi(2)$, and $c(\pi, p)$ is submodular. Under (A1), there is a unique threshold π^* such that*

$$\mu^*(\pi) = \begin{cases} p_1 & \textit{if } \pi \leq \pi^*; \\ p_2 & \textit{otherwise.} \end{cases}$$

where p_1 and p_2 are two possible actions.

Theorem 6 ([3])**.** *Let W_k be a sub-martingale. If $H_k \geq 0$ is predictable and each H_k is bounded, then $(H.W)_k$ is a sub-martingale.*

Theorem 6 corresponds to Theorem 5.2.5 in [3].

From Theorem 3, we have the following results. The proofs are omitted due to lack of space.

Theorem 7. *Let $\Delta(\eta^{y=1})$ and $\Delta(\eta^{y=2})$ be two possible incentives at belief π. Under (A1) and (A2), the Q function in (12) can be simplified as:*

$$Q(\pi, p) = \begin{cases} p + \rho V(\pi) & \textit{if } p \in [0, \Delta(\eta^{y=2})); \\ p - \phi_r + \rho \mathbb{E} V(\pi) & \textit{if } p \in [\Delta(\eta^{y=2}), \Delta(\eta^{y=1})); \\ p + \rho V(\pi) & \textit{if } p \in [\Delta(\eta^{y=1}), 1]. \end{cases} \qquad (16)$$

and $V(\pi) = \min Q(\pi, p)$. Here,

$$\mathbb{E} V(\pi) = \boldsymbol{1}' B^\pi_{y=1} \pi \times V(\eta^{y=1}) + \boldsymbol{1}' B^\pi_{y=2} \pi \times V(\eta^{y=2}).$$

Theorem 7 represents the Q function (12) over the range $[0, 1]$ into *three* regions. The following corollary highlights why such a partition is useful.

Corollary 8. *At every public belief $\pi \in \Pi(2)$, it is sufficient to choose one of the three incentives $\{0, \Delta(\eta^{y=2}), \Delta(\eta^{y=1})\}$.*

[6] See Appendix for definition.

[7] Here average is over different iterations of the estimation process. For example, each round of labelling/classification in Crowdsourcing can be seen as one iteration.

[8] See Footnote 5

Discussion: A consequence of Corollary 8 is that the value function (12) computation reduces to:

$$\begin{aligned} V(\pi) &= \min\{\rho V(\pi), \Delta(\eta^{y=2}) - \phi_r + \rho \mathbb{E} V(\pi), \\ &\quad \Delta(\eta^{y=1}) + \rho V(\pi)\}. \\ \Rightarrow V(\pi) &= \min\{0, \Delta(\eta^{y=2}) - \phi_r + \rho \mathbb{E} V(\pi)\}. \qquad (17) \end{aligned}$$

In words, it is sufficient for the controller to choose either $p = \Delta(\eta^{y=2})$ or $p = 0$ at every belief π, as opposed to $p \in [0, 1]$ at every belief.

Lemma 9. *Under (A1), $\Delta(\eta^{y=1})$ is concave in π, and $\Delta(\eta^{y=2})$ is convex in π.* □

Proofs:

Proof of Theorem 1:
By definition, we know that $\Delta(\eta^y) \in [0, 1]$, $\Delta(e_1) = 1$ and $\Delta(e_2) = 0$. It can be seen by substitution that $\mathbb{E}V(\pi) = V(\pi)$ when $\pi = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$, and let $V(0)$ and $V(1)$ denote the corresponding values. So the incentive function $\Delta(\eta^{y=2})$ is decreasing with π. The value function computation is given by (17). Let

$$\begin{aligned} \nu(\pi, p_1) &= 0 \\ \nu(\pi, p_2) &= \Delta(\eta^{y=2}) - \phi_r \end{aligned}$$

where $p_1 = 0$ and $p_2 = \Delta(\eta^{y=2}) > 0$. The function $\nu(\pi, p)$ is submodular with $p_1 = 0$ and $p_2 = \Delta(\eta^{y=2}) > 0$. From Theorem 4 and Theorem 5, it is easily seen that there exists a unique π_r^* such that

$$\min Q(\pi, p) = \begin{cases} 0 & \text{if } \pi \leq \pi_r^*; \\ \Delta(\eta^{y=2}) - \phi_r + \rho \mathbb{E} V(\pi) & \text{otherwise.} \end{cases}$$

where $\pi_r^*(2)$ is given by

$$\pi_r^*(2) = \{\pi(2) | 0 = \Delta(\eta^{y=2}) - \phi_r + \rho \mathbb{E} V(\pi)\}.$$

Therefore the optimal incentive rule is given as:

$$\mu^*(\pi) = \begin{cases} 0 & \text{if } \pi \leq \pi_r^*; \\ \Delta(\eta^{y=2}) & \text{otherwise.} \end{cases}$$

□

Proof of Theorem 2:
Consider the sub-optimal policy $\hat{\mu}(\pi)$ given as

$$\hat{\mu}(\pi) = \begin{cases} \Delta(\eta^{y=2}) - \epsilon & \text{if } \pi(2) \in [0, \pi^*(2)); \\ \Delta(\eta^{y=2}) & \text{if } \pi(2) \in [\pi^*(2), 1]. \end{cases}$$

Here $\epsilon > 0$ and $\pi^*(2) \in [0, 1]$. Let $W_k = \hat{\mu}(\pi_{k-1})$.
From Lemma 9, $\Delta(\eta^{y=2})$ is convex in π. Let $u^S(\pi_{k+1}) = \Delta(\eta_{k+1}^{y=2})$ denote the incentive at time $k+1$. So $u^S(\pi)$ is convex in π.
We know that the public belief π_k is a martingale [2], i.e, $\mathbb{E}[\pi_{k+1} | \mathcal{G}_k] = \pi_k$. For $\epsilon \to 0$,

$$\mathbb{E}[W_{k+1} | \mathcal{G}_k] = \mathbb{E}[u^S(\pi_{k+1}) | \mathcal{G}_k] \geq u^S(\mathbb{E}[\pi_{k+1} | \mathcal{G}_k]) \geq u^S(\pi_k) \geq W_k$$

by Jensen's inequality and martingale property of the public belief. Therefore $W_k (= \hat{\mu}(\pi_k))$ is a sub-martingale.
Consider a function $\bar{\mu}(\pi)$ given by

$$\bar{\mu}(\pi) = \begin{cases} 0 & \text{if } \pi(2) \in [0, \pi^*(2)); \\ 1 & \text{if } \pi(2) \in [\pi^*(2), 1]. \end{cases}$$

Let $H_k = \bar{\mu}(\pi_{k-1})$. From Theorem 6, $(H.W)_k$ is a sub-martingale. But $(H.W)_k = p_k$. Therefore, the optimal incentive sequence $p_k = \mu^*(\pi_{k-1})$ is a sub-martingale,

$$\mathbb{E}[p_{k+1} | \mathcal{F}_k] \geq p_k$$

i.e, it increases on average over time. □

A. REFERENCES

[1] Barrera, J., and Garcia, A. Dynamic incentives for congestion control. *IEEE Transactions on Automatic Control 60*, 2 (2015), 299–310.

[2] Chamley, C. *Rational herds: Economic models of social learning.* Cambridge University Press, 2004.

[3] Durrett, R. *Probability: theory and examples.* Cambridge university press, 2010.

[4] Karger, D. R., Oh, S., and Shah, D. Iterative learning for reliable crowdsourcing systems. In *Advances in neural information processing systems* (2011), pp. 1953–1961.

[5] Krishnamurthy, V. Quickest detection POMDPs with social learning: Interaction of local and global decision makers. *IEEE Transactions on Information Theory 58*, 8 (2012), 5563–5587.

[6] Krishnamurthy, V. *Partially Observed Markov Decision Processes.* Cambridge University Press, 2016.

[7] Miller, N., Resnick, P., and Zeckhauser, R. Eliciting informative feedback: The peer-prediction method. *Management Science 51*, 9 (2005), 1359–1373.

[8] Moretti, E. Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies 78*, 1 (2011), 356–393.

[9] Nowak, S., and Rüger, S. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval* (2010), ACM, pp. 557–566.

[10] Papanastasiou, Y., Bakshi, N., and Savva, N. Social learning from early buyer reviews: Implications for new product launch. *History* (2013).

[11] Tran-Thanh, L., Venanzi, M., Rogers, A., and Jennings, N. R. Efficient budget allocation with accuracy guarantees for crowdsourcing classification tasks. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems* (Richland, SC, 2013), AAMAS '13, International Foundation for Autonomous Agents and Multiagent Systems, pp. 901–908.

[12] Witkowski, J., Bachrach, Y., Key, P., and Parkes, D. C. Dwelling on the negative: Incentivizing effort in peer prediction. In *First AAAI Conference on Human Computation and Crowdsourcing* (2013).