A Quality Assuring Multi-armed Bandit Crowdsourcing Mechanism with Incentive Compatible Learning

(Extended Abstract)

Shweta Jain
IISc Bangalore, India
jainshweta@csa.iisc.ernet.in

Sujit Gujar
EPFL Lausanne, Switzerland
sujit.gujar@epfl.ch

Onno Zoeter
XRCE Meylan, France
onno.zoeter@xrce.xerox.com

Y. Narahari
IISc Bangalore, India
hari@csa.iisc.ernet.in

ABSTRACT

We develop a novel multi-armed bandit (MAB) mechanism for the problem of selecting a subset of crowd workers to achieve an assured accuracy for each binary labelling task in a cost optimal way. This problem is challenging because workers have unknown qualities and strategic costs.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Miscellaneous; I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents

Keywords

Mechanism Design, Multi-armed Bandit, Crowdsourcing

1. INTRODUCTION

Consider a company (requester) that provides financial advice to its clients on whether or not to invest in a particular security. In order to advice the clients, the company has a pool of financial consultants (crowd workers). Gathering the opinion of as many consultants as possible and aggregating their opinions increases the probability of providing a high quality advice, however, it also entails increased costs. The company has two conflicting business requirements, firstly to keep the costs low, and secondly, to provide a quality of advice that meets a minimum threshold. The financial consultants have heterogeneous and unknown skill sets and their costs are typically private information. Given noisy labels from selected financial consultants, the company seeks to aggregate the labels to achieve a certain target accuracy, at the same time giving the right incentives to the workers so that they report their costs truthfully. The problem addressed in this paper is motivated by such real world problems.

We propose a novel framework, Assured Accuracy Bandit (AAB), in which we formulate an optimization problem for minimizing total cost subject to the constraint that the probability of occurrence of the most likely outcome that leads to an error is below a certain threshold level when majority voting is used for aggregation. We then provide an exploration separated algorithm which we call a *Constrained Confidence Bound* (CCB) algorithm that ensures that the constraint is satisfied for each task with high probability. We provide an upper bound on the number of exploration steps and show that the CCB algorithm produces an ex-post monotone allocation rule which can be transformed into an ex-post incentive compatible and ex-post individually rational mechanism.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

2. THE MODEL

Let $\mathcal{N}$ be the set of n crowdsourcing workers, who are available for completing T similar crowdsourcing tasks. Quality q_i of any worker i represents the probability of the agent providing the correct answer and is assumed to be independent of the qualities of other workers. A cost c_i is associated with each worker that can be reported strategically by the workers. Let $1-e$ be the target accuracy provided by the requester that determines the trade-off between cost and accuracy to be achieved. We assume that the workers are not spammers and the quality of their work is at least better than just random selection of a label ($q_i \geq 0.5\ \forall i$).

2.1 Assured Accuracy Bandit (AAB) Framework

Let $\tilde{y}_i$ denote the noisy label from worker i and $\tilde{y}(S)$ be the vector of noisy labels if set $S = \{1, 2, \ldots, s\}$ is selected with quality vector $q = \{q_1, q_2, \ldots, q_s\}$ such that $q_1 \leq q_2 \leq \ldots \leq q_s$. Let $\hat{y}$ and y denote the predicted label and true label respectively. We bound the probability of occurrence of the most likely outcome that leads to an error $P(E_{S(q)})$. If the aggregation rule is majority voting then:

$$\begin{aligned} P(E_{S(q)}) &= \max_{\tilde{y}(S)\in\{0,1\}^S} P(\tilde{y}(S), \hat{y} \neq y | y) \\ &= (1-q_1)(1-q_2)\ldots(1-q_{s'})q_{s'+1}\ldots q_s, \\ &\text{where } s' = \lfloor((s+1)/2)\rfloor \end{aligned}$$

Bounding the probability of the most likely outcome is the first step as it makes our analysis easier, though a natural approach would be to bound the probability of error. The optimization problem is then given as:

$$\min_{S\subseteq\mathcal{N}} \sum_{i\in S} c_i \text{ s.t. } \tilde{P}(E_{S(q)}) < e \tag{1}$$

$$\text{where, } \tilde{P}(E_{S(q)}) = (1-q_1)(1-q_2)\dots(1-q_{s'}) \qquad (2)$$

We further bound $P(E_{S(q)})$ by $\tilde{P}(E_{S(q)})$ due to the following reasons: 1)$P(E_{S(q)}) \leq \tilde{P}(E_{S(q)})$ $\forall q$, $\forall S$ and 2) $\tilde{P}(E_{S(q)})$ is monotonic i.e. with two quality profiles, q and q' such that $q_i \leq q'_i$ $\forall i \in S$, then $\tilde{P}(E_{S(q)}) < e$ implies $\tilde{P}(E_{S(q')}) < e$. We assume that $\Delta = e - \tilde{P}(E_{S(q)}) > 0$, which plays a crucial role in bounding the number of exploration steps.

3. THE CCB ALGORITHM

ALGORITHM 1: CCB Algorithm

Input: Parameter e, tasks T, workers $\mathcal{N}$, confidence μ
Output: Labeler selection set S^t, Label $\hat{y}_t$ for task t

1 $\mu' = (1-(1-\mu)^{2/n})$, $\forall i$, $\hat{q}_i^+ = 1$, $\hat{q}_i^- = 0.5$, $c_{i,1} = 0$
2 $S^1 = \mathcal{N}$, collect $\tilde{y}(S^1)$ and $\hat{y}_1 = \text{MAJORITY}(S^1)$
3 Observe true label y_1
4 $\forall i \in \mathcal{N}$, $n_{i,1} = 1$, $c_{i,1} = 1$ if $\tilde{y}_i = y_1$ and $\hat{q}_i = c_{i,1}/n_{i,1}$
5 **for** $t = 2$ *to* T **do**
6   Let $S^t_{opt} = \arg\min_{S \subseteq \mathcal{N}} \sum_{i \in S} c_i$ s.t. $\tilde{P}(E_{S(\hat{q}^+)}) < e$
7   **if** $\tilde{P}(E_{S^t_{opt}(\hat{q}^-)}) > e$ **then**
8         $S^t = \mathcal{N}$, $\hat{y}_t = \text{MAJORITY}(S^t)$ % Explore
9   Observe true label y_t
10         $\forall i \in S^t$, $n_{i,t} = n_{i,t} + 1$, $c_{i,t} = c_{i,t} + 1$ if $\tilde{y}_i = y_t$, $\hat{q}_i = c_{i,t}/n_{i,t}$, $\hat{q}_i^+ = \hat{q}_i + \sqrt{\frac{1}{2n_{i,t}} ln(\frac{1}{\mu'})}$, $\hat{q}_i^- = \hat{q}_i - \sqrt{\frac{1}{2n_{i,t}} ln(\frac{1}{\mu'})}$
11   **else**
12         $t^* = t$
13         $S^{t^*} = S^t_{opt}$, $\hat{y}_t = \text{MAJORITY}(S^t)$
14   Break % Goto Step 16
15
16 **for** $t = t^* + 1$ *to* T **do**
17     $S^t = S^{t^*}$, $\hat{y}_t = \text{MAJORITY}(S^t)$ % Exploit

The CCB algorithm is presented in Algorithm 1. We assume that the true label is observed once the task is completed and there are enough number of workers such that the constraint is satisfied with respect to the true qualities by selecting all the workers. The algorithm works on the principle of the UCB algorithm [1] and ensures that the constraint in (1) is satisfied with high confidence μ. Upper confidence ($\hat{q}^+$) and lower confidence ($\hat{q}^-$) bounds on qualities are maintained based on Hoeffding's inequality such that true qualities lie between these bounds with probability $1-\mu'$. The idea is to first solve the optimization problem with respect to the upper confidence bound. Select all the workers (Exploration round) if selected set S^t does not satisfy the constraint with respect to lower confidence bound else S^t is the optimal set with probability $1-\mu$ (Lemma 3.1) and hence select the set S^t for all the remaining rounds (Exploitation rounds).

Lemma 3.1 *The set S^{t^*} returned by Step 13 is the optimal set with probability $1-\mu$*

4. KEY RESULTS

The CCB algorithm is adaptive exploration separated where the number of exploration rounds in which all the workers are selected depends on the way learning progresses. Theorem 1 provides bounds on the number of exploration steps.

Theorem 1 *The number of exploration rounds is bounded by $\frac{8}{\Delta^2} n^2 ln(\frac{1}{\mu'})$ with probability $1-\mu$ where $(1-\mu) = (1-\mu')^{n/2}$*

Let the optimal set be denoted by S^*. We prove the above theorem by showing that with probability $(1-\mu)$, after $t^* = \frac{8}{\Delta^2} n^2 ln(\frac{1}{\mu'})$ rounds of uniform exploration no other set $S \neq S^*$ with lower cost satisfies the constraint with respect to the upper confidence bound using Hoeffding's inequality. Moreover, $\hat{P}(E_{S^*(\hat{q}^-)}) < e$ with probability $(1-\mu)$ after t^* uniform exploration rounds. Thus, the expected number of exploration rounds is given by $(1-\mu)\frac{8}{\Delta^2} n^2 ln(\frac{1}{\mu'}) + T\mu$

Theorem 2 *The CCB algorithm gives an ex-post monotone allocation rule i.e. for every random realization, $\forall i$, $\forall \hat{c}_{-i}$, $\hat{c}_i \leq \hat{c}_i^+ \Rightarrow \mathcal{A}_i(\hat{c}_i, \hat{c}_{-i}) \geq \mathcal{A}_i(\hat{c}_i^+, \hat{c}_{-i})$ where $\mathcal{A}_i(\hat{c}_i, \hat{c}_{-i})$ is the number of tasks given to the i^{th} worker with bids $\hat{c}_i$ and $\hat{c}_{-i}$*

For a fixed random realization, quality updates remain the same for the same number of exploration steps. Since the optimization problem involves cost minimization, if quality updates are the same then, if the optimal set contains i with bid $\hat{c}_i^+$ then i belongs to the optimal set with bid $\hat{c}_i$. Let $t^*(\hat{c}_i)$ and $t^*(\hat{c}_i^+)$ represent the number of exploration steps with bid $\hat{c}_i$ and $\hat{c}_i^+$ respectively. We show that if i belongs to the optimal set with bid $\hat{c}_i$ but not with bid $\hat{c}_i^+$ then $t^*(\hat{c}_i^+) \leq t^*(\hat{c}_i)$ and hence, monotonicity follows. In the other cases when $i \in S^*$ with bids $\hat{c}_i$ and $\hat{c}_i^+$ or when $i \notin S^*$ with bids $\hat{c}_i$ and $\hat{c}_i^+$, then we show that $t^*(\hat{c}_i) = t^*(\hat{c}_i^+)$.

Now the transformation presented in [2] can be used to produce a randomized mechanism which is ex-post incentive compatible and universally ex-post individually rational.

5. SUMMARY AND FUTURE WORK

To the best of our knowledge this is the first contribution that combines learning of qualities and eliciting of true costs in a crowdsourcing environment. The algorithm works well when the number of workers is manageable and qualities are high enough. However, the scalability of the algorithm needs to be investigated and worked upon further. The algorithm is currently exploration separated in order to achieve ex-post monotonicity. One can investigate existence of other algorithms satisfying desirable mechanism properties with lower regret. Currently, the proposed monotonic constraint is stiffer than the desirable constraint. It would be nice to see if the constraint can be stronger compared to the actual one. One can also use different rules to aggregate crowd answers like weighted majority voting.

Acknowledgments

We thank Satyanath Bhat for providing valuable suggestions. This work is part of a collaborative project between Xerox Research and Indian Institute of Science.

6. REFERENCES

[1] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, May 2002.

[2] Moshe Babaioff, Robert D. Kleinberg, and Aleksandrs Slivkins. Truthful mechanisms with implicit payment computation. In *EC-2010*, pages 43–52. ACM, 2010.