Computational Results For Crowdsourced Judgement Elicitation

Virgile Audi (vaudi@g.harvard.edu)
Charles Liu (cliu02@g.harvard.edu)

March 5, 2016

Abstract

Our project will focus primarily on the "Crowdsourced Judgement Elicitation with Endogenous Proficiency" paper by Dasgupta and Ghosh [1]. Although the paper provided very strong results, giving a maximal Nash Equilibrium of truth-reporting under a variety of scenarios, they required many assumptions to prove their findings. For our final project we'd like to relax some of those assumptions and see if similar results are met. We will develop a Python package that can simulate over the number of tasks/participants and estimate expected rewards for participants.

1 Effort/Proficiency Distribution

Two central assumptions in the paper are:

- Effort is a binary decision - either 0 effort where you guess the correct result based off a coin flip or you give full effort and guess the correct result based off your proficiency

- Your proficiency for obtaining the correct result is greater than .5

The second assumption is central to proving that giving full effort is an equilibrium (Lemma 4). By relaxing that assumption, this should no longer be optimal for each participant. There are three scenarios to consider here - participant's own proficiency $< .5$, participant's expected reference rater's proficiency $< .5$, and both.

One of the stronger statements we found the paper to make was Lemma 8:

> "Suppose the probability of agent i using strategy (1, X) is δ and strategy $(0, r_i)$ is $1 - \delta$ for each task $j \in J(i)$. Suppose i's potential reference raters $r_j(i)$ use strategies (1, X) and $(0, r_{r_j}(i))$ with probabilities $\epsilon_{r_j}(i)$ and $1 - \epsilon_{r_j}(i)$ respectively, for each task $j \in J(i)$. If $\epsilon_{r_j}(i) > 0$ for any reference rater with proficiency $p_{r_j}(i) > \frac{1}{2}$, then agent i has a (strict) profitable deviation to $\delta' = 1$, i.e., to always using strategy (1, X), for all values of $r_i \in [0, 1]$"

If the rater has any chance of giving full effort, then the optimal strategy is to give full effort. Aside from the proficiency being at least .5, this is reliant on the idea that effort is a binary measure. Another way of looking at this is that the proficiency of a participant is some function of effort, where effort is over a distribution. In our implementation, we will specify some prior on effort that would be used as a parameter in our proficiency distribution.

2 Uncommon Prior

3 Summary

The paper notes:

> "We assume that the maximum proficiency $p_i \geq \frac{1}{2}$ for all i this minimum requirement on agent ability can be ensured in online crowdsourcing settings by prescreening workers on a representative set of tasks (Amazon Mechanical Turk, for instance, offers the ability to prescreen workers [13, 4], whereas in peergrading applications such as on Coursera, students are given a set of pre-graded assignments to measure their grading abilities prior to grading their peers, the results of which can be used as a prescreen.)"

We believe this is a very large assumption - in many cases, like the product review setting that was used in the peer prediction paper, the quality of reviewers may not be known. This leads to relaxing different assumptions - on the one hand clearly proficiency, but also the prior on receiving some signal.

4 References

1. http://www.arpitaghosh.com/papers/elicit_arxiv_2.pdf

2. http://www.eecs.harvard.edu/econcs/pubs/witkowski_ec12.pdf