Tuning the Diversity of Open-Ended Responses From the Crowd

Walter S. Lasecki
University of Rochester
Computer Science
Rochester, NY 14627 USA
wlasecki@cs.rochester.edu

Christopher M. Homan
Rochester Institute of Technology
Computer Science
Rochester, NY 14623 USA
cmh@cs.rit.edu

Jeffrey P. Bigham
Carnegie Mellon University
HCI and LTI Institutes
Pittsburgh, PA 15213
jbigham@cs.cmu.edu

Introduction

Crowdsourcing can solve problems beyond the reach of state-of-the-art fully automated systems (Bigham et al. 2010; Lasecki et al. 2011; 2012; Bernstein et al. 2011; von Ahn and Dabbish 2004; Attenberg, Ipeirotis, and Provost 2011; Aral, Ipeirotis, and Taylor 2011). A common pattern found in many such systems is for the workers to discover, in parallel, a number of candidate solutions and then vote on the best one to pass forward, often within a fixed amount of time.

Given limited human resources, then, how much effort should be spent on discovering new solutions versus deliberating over those that have already been proposed? Too many proposals and it may be too hard for the remaining workers to discriminate among them and make a clear group decision. Too few and the best answer might not be found. Clearly, the optimal balance depends on many factors specific to the crowd and the problem itself, so a flexible approach is needed to make it easy for system designers to elicit responses appropriately.

We present the *propose-vote-abstain* mechanism for eliciting from crowd workers the proper balance between solution discovery and selection. Each crowd worker is given a choice among *proposing* an answer, *voting* among the answers proposed so far, or *abstaining*, i.e., doing nothing. When a stopping condition is reached, the mechanism returns the answer with the most votes. Workers are paid a base amount, with bonuses if they propose or vote for the winning answer.

This mechanism has several virtues, from a crowdsourcing perspective: (1) It is simple and easy to understand, which saves on worker time and cognitive load, leading to better use of valuable human resources. (2) This simplicity also makes it easy to parallelize, manipulate, an plug into other existing systems. (3) By providing each worker the option of abstaining, it provides a release valve for those workers who lack confidence in any particular answer, thus removing noise from this system.

We provide a game-theoretic analysis that shows, among other things, the baseline behavior of the system under different worker incentive structures. We also study experimentally a fundamental problem in this setting, namely, how to best distribute limited resources between answer discovery and voting. How many alternatives— absolutely or relative to the number of voters—is ideal may vary dynamically on the answers proposed so far or on the nature of the questions asked. We show that the propose-vote-abstain mechanism can help find the best balance for particular questions, and can even tune this balance in real time.

The Propose-Vote-Abstain Mechanism

This propose-vote-abstain mechanism is designed for settings in which the first input is a *request*— for information or an answer to a specific question. Workers are then recruited into the system. When a worker joins, he or she is presented with a request and the solutions proposed by all previous workers. The worker is then offered the following choices: *propose* an new answer with a reward of π if that answer eventually receives the most votes, *vote* for an existing candidate answer with a reward of ν if that candidate eventually receives the most votes, or abstain with reward α.

Game-Theoretic Analysis

We use game theory to get a sense of how we might expect the propose-vote-abstain mechanism to perform under ideal circumstances. In order to make the analysis tractable, we make several simplifying assumptions:

1. The game has a single turn with a indeterminate (and unknown to the workers) number of players.
2. The game does not terminate unless there is at least one answer proposed and one vote.
3. If at termination more than one candidate has the most votes, then one of these candidates is selected uniformly at random as the winner.
4. We assume that each player has equal confidence in alternatives winning, including ones proposed by the player.
5. The only information the workers know about the system are the candidate answers and the request.

Assumptions in 4–5 are, admittedly, quite strong. However, they mean that (for the purposes of analysis) the only state information we need to consider are the number of alternatives m_t proposed so far, where t is the current time.

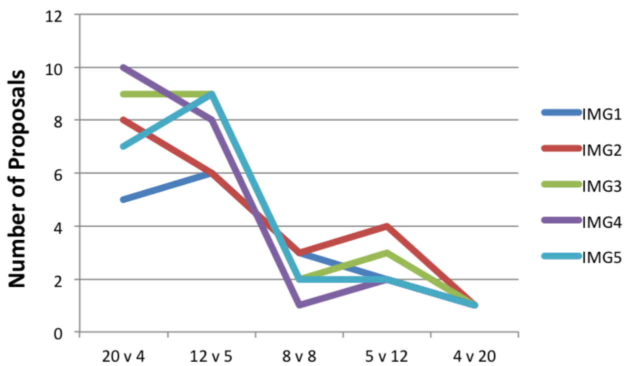

Figure 1: Results from varying relative payments for voting and proposing responses to an image description task.

This yields several results about the behavior of the crowd as a function of the payoffs for each response.

Proposition 1. *Under the assumptions above:*

1. *If* $\alpha \geq \min\{\pi, \nu\}$*, then abstaining is a dominant strategy for all players.*
2. *If* $\min\{\pi, \nu\} > \alpha$ *then the dominant strategy for the first* $\min\{\lfloor \nu/\pi \rfloor, \lfloor \pi/\alpha \rfloor, 1\}$ *workers is to propose and for the remaining workers it is to vote.*

Experimental Analysis

To test our model's ability to elicit different levels of response diversity from workers, we setup a simple image description task. We recruited 100 Mechanical Turk workers and asked them to view a set of 5 images (presented in random order) and either propose, vote for, or abstain from contributing to the image's description. Our relative pricing levels between the vote and propose actions were traded off to test a range of three options between 4 and 20 cents (per image). Rewards were given based on the result of the aggregate decision at the end of our experiments. The abstain payment was fixed at 2 cents for all of the questions in our tasks.

Figure 1 shows the results of our tests. As the payment for voting becomes large relative to the proposal payment, the number of total answers generated by the system significantly decreases from an average of 7.8 responses to 1 response for all 5 pictures we saw ($p < .0001$). The decreasing trend was linear with $R^2 = 0.802$. Note that there is a disproportional drop at the break-even point when payment is equal for both options. This is consistent with what we expect because voting requires less effort than generating a response, so there is a slight bias in its favor.

While all of the images eventually converged to a single response as the vote payment increased, the number of responses generated in the opposing case (where the proposal reward is high and workers are incentivized to generate several answers) varies from 5 to 10 responses each. This is likely dependent on how subjective the image is an how many answer could be considered plausible with high confidence. This trend is seen throughout the results as each response set trends towards a single response. This suggests that the content of the task does play a role in worker trends, but in the convergent limit this can be overridden by financial incentives.

Conclusion

In this paper, we presented the propose-vote-abstain mechanism for eliciting answers from crowd workers. Consistent with our theoretical results (which predict no abstentions) very few abstentions occurred. The theoretical results suggest that abstention plays an important role in regulating the number of proposals, even though few participants actually abstain. Further work is needed to better understand how exactly the theoretical results translate to practice.

Acknowledgments

Funding for this research was provided by NSF Awards #IIS-1149709 and #SES-1111016, and a Microsoft Research Graduate Fellowship.

References

Aral, S.; Ipeirotis, P.; and Taylor, S. 2011. Content and context: Identifying the impact of qualitative information on consumer choice. In Galletta, D. F., and Liang, T.-P., eds., *ICIS*. Association for Information Systems.

Attenberg, J.; Ipeirotis, P. G.; and Provost, F. J. 2011. Beat the machine: Challenging workers to find the unknown unknowns. In *Human Computation*, volume WS-11-11 of *AAAI Workshops*. AAAI.

Bernstein, M. S.; Brandt, J.; Miller, R. C.; and Karger, D. R. 2011. Crowds in Two Seconds: Enabling Realtime Crowd-powered Interfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2011)*, 33–42.

Bigham, J. P.; Jayant, C.; Ji, H.; Little, G.; Miller, A.; Miller, R. C.; Miller, R.; Tatarowicz, A.; White, B., W. S.; and Yeh, T. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2010)*, 333–342.

Lasecki, W. S.; Murray, K. I.; White, S.; Miller, R. C.; and Bigham, J. P. 2011. Real-time Crowd Control of Existing Interfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2011)*, 23–32.

Lasecki, W. S.; Miller, C. D.; Sadilek, A.; Abu-Moussa, A.; and Bigham, J. 2012. Real-time captioning by groups of non-experts. In *In Submission*. http://hci.cs.rochester.edu/pubs/pdfs/legion-scribe.pdf.

von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, 319–326. New York, NY, USA: ACM.