Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence

Threats and Trade-Offs in Resource Critical Crowdsourcing Tasks over Networks

Swaprava Nath[1], **Pankaj Dayama,**[2] **Dinesh Garg,**[3] **Y. Narahari,**[4] **James Zou**[5]
[1]PhD Candidate, Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560 012, India
swaprava@gmail.com, (091) 944484 19311, http://swaprava.byethost7.com
[2] Global General Motors R&D - India Science Lab [3]IBM India Research Lab
[4]Dept. of Computer Science and Automation, Indian Institute of Science
[5]School of Engineering and Applied Sciences, Harvard University

Abstract

In recent times, crowdsourcing over social networks has emerged as an active tool for complex task execution. In this paper, we address the problem faced by a planner to incentivize agents in the network to execute a task and also help in recruiting other agents for this purpose. We study this mechanism design problem under two natural resource optimization settings: (1) cost critical tasks, where the planner's goal is to minimize the total cost, and (2) time critical tasks, where the goal is to minimize the total time elapsed before the task is executed. We define a set of fairness properties that should be ideally satisfied by a crowdsourcing mechanism. We prove that no mechanism can satisfy all these properties simultaneously. We relax some of these properties and define their approximate counterparts. Under appropriate approximate fairness criteria, we obtain a non-trivial family of payment mechanisms. Moreover, we provide precise characterizations of cost critical and time critical mechanisms.

Introduction

Advances in the Internet and communication technologies have made information aggregation from a crowd easier. Examples are labeling millions of images, prediction of stock markets, seeking answers to specific queries, searching for objects across a wide geographical area, etc. This mode of information aggregation is popularly known as *crowdsourcing* (for details, see (Surowiecki 2005) and (Howe 2009)). Many emerging crowdsourcing applications operate through an underlying social network. The task owner initially recruits individuals from its immediate network to participate in executing the task. These individuals, apart from attempting to execute the task by themselves, recruit other individuals in their respective social networks to also attempt the task and further grow the network. The success of such crowdsourcing applications depends on providing appropriate incentives to individuals for both (1) executing the task by themselves and/or (2) recruiting other individuals. Designing a proper incentive scheme (crowdsourcing mechanism) is crucial to the success of any such crowdsourcing based application. *Amazon Mechanical Turk* is one of the early crowdsourcing mechanisms, and afterwards many popular examples emerged in the recent past. For example, oDesk, Netflix challenge, *Red Balloon Challenge* (DARPA 2010) and CLIQR quest (DARPA 2012), etc., all of which aim to leverage the collective intelligence of the crowd. In the red balloon challenge, the winning team from MIT demonstrated that a crowdsourcing mechanism can be employed to accomplish such a challenging task (Pickard et al. 2011). Other scenarios where such crowdsourcing based mechanisms can be very effective include *query incentive networks* (Kleinberg and Raghavan 2005), *multi-level marketing* (Emek et al. 2011), etc.

The relevant literature can be classified into two focal areas: one which aims at (a) **crowdsourcing** applications, such as geographical area scanning, viral marketing, multi-level marketing, query answering, etc. (Lakhani et al. 2006), (Ipeirotis 2010), and the other dealing with (b) **sybilproof** mechanisms, (Iwasaki et al. 2007), (Conitzer et al. 2010). There are few examples in literature that address sybilproofness in crowdsourcing applications. The red balloon challenge (DARPA 2010) and its novel solution method proposed in (Pickard et al. 2011) can be considered as an early work that motivated the study of strategic aspects in crowdsourcing applications. In this paper, we formally address the issue of sybil attacks in crowdsourcing settings and provide theoretical limits of achievability.

A major challenge in deploying such crowdsourcing mechanisms in real world problems is their vulnerability to different kinds of manipulations (e.g. false name attacks, also known as *sybil attacks* in literature (Conitzer et al. 2010)) that rational and intelligent participants may attempt. In this paper, we discuss the problem of designing optimal crowdsourcing mechanisms that are robust to such manipulations. The work which is most relevant to our paper is (Emek et al. 2011), though our problem setting is different from theirs in many ways. First, in our setting, the planner's utility lies solely in getting an time or cost critical atomic task completed. Second, the task accomplishment does not yield any monetary benefit to the planner, unlike in their model. Third, an agent may get incentives through multiple chains in their setting whereas that is not allowed in our model. Finally, the two different kinds of sybil attacks considered in their setting turn out to be the same for our setting due to the atomic nature of the task that we consider. Though the underlying research questions are related, our specific contribution lies in deriving non-trivial, custom results for a much different class of problems and in proposing and using novel notions of approximate fairness measures to obtain a

deeper understanding of incentive design in crowdsourcing.

Model and Main Results

In this paper, we consider a set of agents connected over a social graph. A distinguished node, called the planner, selects a subset of her neighbors and asks them to execute an atomic task and offers a monetary reward scheme. A task is *atomic*, if it is indivisible and executable by a single individual. These nodes can choose to execute the task themselves or forward it to their neighbors. We say that the latter nodes are *recruited* by the former nodes. We consider mechanisms where no node can be recruited by multiple nodes, which results in a tree recruitment structure. The successful executer of the atomic task is called the *winner* and the corresponding chain to the root is called the *winning chain*.

On this setting, we identify a set of desiderata that are reasonable for an ideal crowdsourcing mechanism. (1) *Downstream Sybilproofness (DSP)*: this property ensures that no agent can gain by creating fake identities below her in the recruitment tree. (2) *Collapse-proofness (CP)*, which ensures agents collectively do not gain by collapsing to a single node. (3) *Strict and Weak Contribution Rationality (SCR & WCR)*, which ensures positive or non-negative payment to the agents of the winning chain. (4) *Budget Balance (BB)*, and (5) *Security to Winner (SEC)*, so that the winner gets at least a constant fraction of the total reward. The precise definitions of these properties are given in (Nath et al. 2012). Properties (1) and (2) are responsible for a truthful revelation of the network. Property (3) encourages participation of the members of the winning chain to forward the task to a potential winner, while (5) is a motivation for a node to do the task himself. Property (4) is natural for any mechanism design problem so that cost does not spill over budget. We assume that any non-winning chain derive a zero payment. We have the following two results.

Theorem 1 *No reward mechanism can satisfy DSP, SCR, and CP together.*

Theorem 2 *Properties DSP, WCR, CP, and BB are equivalent to a* winner-takes-all *mechanism.*

These results lead us to the question: *if approximation is inevitable, which properties do we care about under certain circumstances?* We undertake two different streams of investigation: (a) **cost critical tasks**, where the goal is to minimize the total cost, (b) **time critical tasks**, where the goal is to minimize the total time for executing the task. Notice that *query incentive networks* (Kleinberg and Raghavan 2005) and *multi-level marketing* (Emek et al. 2011) fall under the category of cost critical (revenue maximizing) tasks, while search-and-rescue operations and the red balloon challenge (DARPA 2010) fall under that of time critical tasks. For case (a), we relax DSP to ϵ-DSP (one cannot gain by more than a factor of ϵ by creating sybils) and show that,

Theorem 3 *For all $\epsilon > 0$, the space of mechanisms satisfying ϵ-DSP, CP, BB, and SCR is nonempty.*

However, these mechanisms could be quite unfair to the participants. Hence we introduce concepts of δ-SCR (each node gets at least δ fraction of its child) and γ-SEC (winner gets at least γ fraction of the total reward), which leads to the following characterization theorem.

Theorem 4 *If $\delta \leq \min\{1-\gamma, \frac{\epsilon}{1+\epsilon}\}$, a mechanism is a cost minimizer over the class of mechanisms satisfying ϵ-DSP, δ-SCR, γ-SEC, and BB iff it is (γ, δ)-GEOM mechanism.*

Mechanism (γ, δ)-GEOM offers γ fraction of total reward to the winner and geometrically decreases the reward with factor δ towards the root.

For case (b), we consider mechanisms that maximize the payment to the leaf node (winner). It incentivizes nodes to solve the task themselves (rather than simply forwarding) and in turn minimizes the completion time. We call these class of mechanisms `MAXLEAF`. We show that,

Theorem 5 *If $\delta \leq \frac{\epsilon}{1+\epsilon}$, a mechanism is* `MAXLEAF` *over the class of mechanisms satisfying ϵ-DSP, δ-SCR, and BB iff it is δ-GEOM mechanism.*

Mechanism δ-GEOM offers $\frac{1-\delta}{1-\delta^t}$ fraction of the total reward to the winner and geometrically decreases the rewards towards root with the factor δ, where t is the length of the winning chain. Details of all proofs, definitions, and justifications of the choice of properties are in (Nath et al. 2012).

Conclusions and Future Work

In this paper we show certain impossibilities in crowdsourcing mechanisms. We provide approximation schemes that are meaningful under certain application domains. This opens up a bunch of interesting research questions which we plan to investigate as a future work.

References

Conitzer, V.; Immorlica, N.; Letchford, J.; Munagala, K.; and Wagman, L. 2010. False-name-proofness in social networks. In *Proceedings of the 6th international conference on Internet and network economics*, WINE'10, 209–221.

DARPA. 2010. DARPA Network Challenge Project Report. In *http://archive.darpa.mil/networkchallenge/*.

DARPA. 2012. DARPA CLIQR Quest Challenge.

Emek, Y.; Karidi, R.; Tennenholtz, M.; and Zohar, A. 2011. Mechanisms for Multi-Level Marketing. In *Proceedings of the 12th ACM Conference on Electronic Commerce (EC)*.

Howe, J. 2009. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. Crown Business.

Ipeirotis, P. G. 2010. Analyzing the Amazon Mechanical Turk Marketplace. *XRDS* 17:16–21.

Iwasaki, A.; Kempe, D.; Saito, Y.; Salek, M.; and Yokoo, M. 2007. False-name-proof mechanisms for hiring a team. In *Proceedings of the 3rd international conference on Internet and network economics*, WINE'07, 245–256.

Kleinberg, J., and Raghavan, P. 2005. Query Incentive Networks. In *Proceedings 46th IEEE Symposium on Foundations of Computer Science*, 132–141.

Lakhani, K. R.; Jeppesen, L. B.; Lohse, P. A.; and Panetta, J. A. 2006. The Value of Openness in Scientific Problem Solving. Working Paper.

Nath, S.; Dayama, P.; Garg, D.; Narahari, Y.; and Zou, J. 2012. Threats and Trade-offs in Resource Critical Crowdsourcing Tasks over Networks. Technical report, Indian Institute of Science.

Pickard, G.; Pan, W.; Rahwan, I.; Cebrian, M.; Crane, R.; Madan, A.; and Pentland, A. 2011. Time-critical social mobilization. *Science* 334:509–512.

Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor.