Analysing Incentive Mechanisms for Crowdsourcing

E. del Val[1], G. Martínez-Cánovas[2] V. Botti[1], and P. Hernández[2]

[1] Universitat Politècnica de València
[2] Universitat de València

Abstract. Crowdsourcing applications are truly dependent on the contribution of users with their data. In this paper, we propose two types of mechanisms (*Blind* and *Informed*) that promote the participation of users before a deadline through monetary rewards. *Blind* mechanisms do not have knowledge about the expected rewards of potential participants. *Informed* mechanism is based on the information that the potential participants provide. Mechanisms were evaluated in populations with rational and irrational dynamic behaviors.

1 Introduction

Currently, there are many organizations that move towards participatory models. In order to develop a crowdsourcing application, organizations should consider a set of characteristics to be successful: (i) the task should be modular; (ii) a community of interest must be engage; (iii) utilize the output from the crowd in a manner that creates value [8]. In this paper, we focus on the promotion of users' participation [7][1][10][5]. Individuals may provide their contribution free since they have intrinsic motivations or enjoyment [6] or, they may expect an economic reward in exchange of their contribution. Services based on crowdsourcing are usually related to real-time applications (i.e., citizens behavior monitoring [3], traffic monitoring [2], noise monitoring [4]) and periodically require a high number of contributions. For potential participants, each contribution may require resources. Therefore, it is important to ensure participation through the use of mechanisms. In this paper, we consider two type of mechanisms that take into account the number of required samples, time and budget constraints. Both mechanisms allow the adaptation of the reward per contribution to populations where individual behavior patterns evolve with time.

2 Incentive Mechanisms

We consider a system S that needs to obtain small contributions (samples or data) in order to properly offer its service ρ. There is a set of agents N that are potential participants. Each agent chooses to participate in exchange for a certain reward r_i, or to do nothing. The system S requires $X \leq N$ samples to properly offer its service. There is a time constraint of T rounds to obtain X and S has budget B to spend in rewards. Considering the previous constraints, S tries to collect X samples minimizing the cost of the rewards offered to participants. X, B, and T are private information of S.

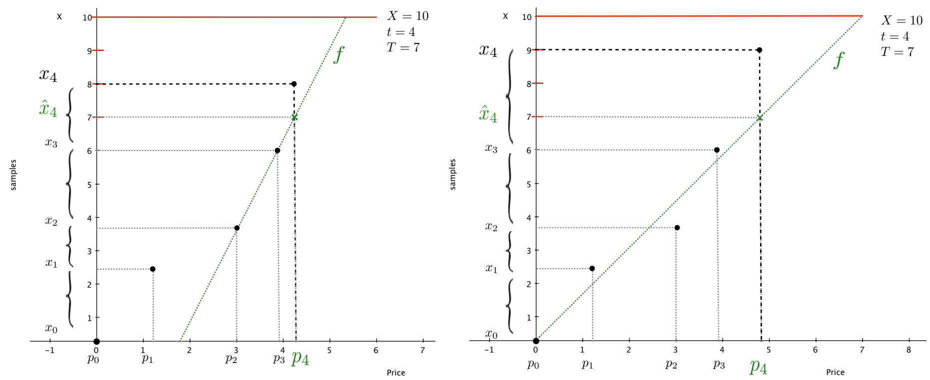

Fig. 1. Function f in round $t = 4$ using *l2p* and considering $X = 10$ and $T = 7$. X-axis represents the prices from previous rounds (i.e., $\mathbf{P} = \{p_0, p_1, p_2, p_3\}$). Y-axis represents the number of cumulative participations in previous rounds (i.e., $\sum_{j=0}^{j<t} x_j \in \mathbf{X}$). Left figure f is calculated with *l2p* and in right figure f is calculated with *lm*.

Blind mechanism starts establishing the reward that S will give for the participants in the first round (p_1) assuming a linear cumulative distribution function *cdf* (i.e., $p_1 = \frac{X}{T}$). Considering this reward (p_1), S starts a *call for participation* protocol (*cfp* hereafter). An agent i will participate providing its data to S, if the reward offered in the *cfp* is greater than its expected reward r_i ($r_i \leq p_1$). In that case, agent i will receive p_1. The number of agents that participate x_t and the reward p_t given by S in round $t = 1$ is stored in $\mathbf{X}$ and in $\mathbf{P}$ respectively.

A similar process is repeated in the following rounds ($1 < t < T$). While the number of participations until current round t is lower than X (i.e., $\sum_{j=0}^{j=t} x_j < X$), and there is enough budget ($expense < B$), S continues the process and estimates the price for the following round. Based on the information collected from previous rounds (i.e., $\mathbf{P}$ and $\mathbf{X}$), the mechanism creates a linear function f that establishes the relation between the expected number of contributions that S might obtain given a certain reward per participation (i.e., an estimation of the *cdf* of the market at round t). We establish two estimation methods to generate f: (i) Last two Points (*l2p*) that estimates *cdf* using a linear function that passes through the last two points in $\mathbf{P}$ and the last two cumulative samples $\sum_{j=0}^{j=t-2} x_j \in \mathbf{X}$ and $\sum_{j=0}^{j=t-1} x_j \in \mathbf{X}$ (see Figure 1); (ii) Linear Model (*lm*) that is more informed method than *l2p*. *lm* estimated *cdf* using the Least Squares method taking all the information in previous rounds (i.e., $\mathbf{P}$ and $\mathbf{X}$) into account.

Once f is estimated for the following round, S calculates the number of samples collected until now (*samples*) and the total number of samples that S expects to obtain in round $t+1$ ($\hat{x}_{(t+1)}$). Based on $\hat{x}_{(t+1)}$, function f, and the number of rounds until reach T, S estimates if there is enough budget to continue with round $t+1$. If this cost is lower than the current budget, the system continues offering prices to agents. The next price is calculated using f.

f receives as input parameter the cumulative number of samples that S expects to reach at the end of this time step ($samples+\hat{x_t}$). Then, S starts a *cfp* protocol and the agents whose expected reward (r_i) is under the price proposed by S, will provide their sample to S. After that, S adds the price established and the number of samples obtained in round t, and updates the expensed budget.

Informed mechanism starts calculating the total samples that S expects to collect in round t ($\hat{x_t}$) taking into account the number of samples that S already has (i.e., $\frac{N-samples}{T-t}$). Then, S asks for the expected rewards of agents (i.e., $\tilde{r_i}$). After that, assuming that each potential participant a_i may contribute with one sample, S creates a partial ordered subset that contains the agents with the lowest expected rewards $\mathbf{R} = \{(a_i, \tilde{r_i}), (a_j, \tilde{r_j}), \ldots, (a_n, \tilde{r_n})\} : r_i \leq r_n \leq r_j \wedge |\mathbf{R}| = \hat{x_t}$. From $\mathbf{R}$, S selects the highest expected reward ($\mathbf{R}[\hat{x_t}]$). Based on this reward, S starts a *cfp* protocol. In this protocol, S behaves differently depending on the agent it is interacting with. If an agent a_i was one of the agents with the lowest values of expected reward ($a_i \in \mathbf{R}$), S offers the expected reward $\tilde{r_i}$. Otherwise, S offers p_t to the rest of agents that expected a reward higher than p_t. Then, each agent decides to participate or not depending on their real expected reward (r_i). The algorithm returns the total number of samples (x_t), the rewards provided to agents that participated ($\mathbf{P}$), and calculates the expense in the current round t. This process is repeated until the number of required samples X is reached, the number of rounds do not exceed T, and there is enough budget to continue asking for samples.

3 Experiments

The following tests focus on how the previous mechanisms are able to adapt the prices that S offers to the potential participants N in order to reach the number of required samples X minimizing the cost of S. To simulate the dynamic economic behaviors of potential participants, we consider rational and irrational behavior patterns (pattern 1 and pattern 5) from [9]. We evaluated the following configurations: Population 1 with 25% rational 75% irrational users and Population 2 with 75% rational 25% irrational users.

In the experiments, we considered that the potential number of participants was $N = 1000$, the number of samples that S required was $X = 800$, and the number of rounds was $T = 10$. Table 1 shows the results obtained with *Blind2points* and *BlindLeastSquares* mechanisms considering populations 1 and 2. It was observed that in populations where more than the 50% of the population were irrational, in the last rounds, the mechanism *Blind2points* estimated the rewards that potential participants expected better than *BlindLeastSquares*. This fact can be observed in the percentage error, the total expense, and the final number of samples. The results obtained using the mechanism that interacts with potential participants to ask for their expected reward values $\tilde{r_i}$. The *informed* mechanism shows a similar performance independently of the behavior of the potential participants. The possibility of asking for the expected reward makes that the mechanism adjusts better the rewards and the final expense is lower than in *Blind2points* and *BlindLeastSquares*. However, with population 1 (i.e., when there are irrational potential participants), S does not always collect the expected number of samples X.

Table 1. Comparison between *Blind2points*, *BlindLeastSquares*, and *Informed*.

	participation			expense			error	
	b2p	bls	informed	b2p	bls	informed	b2p	bls
population 1	845	965	797	13173.3	16784.91	13161.6	23.7%	34.48%
population 2	820	876	802	12038.18	13842	12326.26	6.03%	10.7%

4 Conclusions

The mechanisms described in the paper minimize the cost of the potential participants contributions and adapt the reward in each round considering that the expected rewards may change with time. The experiments show that for populations where the majority of potential participants follow an irrational pattern and it is not possible to obtain information from them, the best mechanism is *Blind2points*. If there is information about the expected reward of the potential participants, the *Informed* mechanism offers more accurate rewards than the other mechanisms independently of the behavior of the population.

Acknowledgements

This work was partially supported by the following projects: TIN2014-55206-R, TIN2012-36586-C03-01, and PROMETEOII/2013/019.

References

1. Del Val, E., Rebollo, M., Botti, V.: Promoting cooperation in service-oriented mas through social plasticity and incentives. JSS 86(2), 520–537 (2013)
2. Koukoumidis, E., Peh, L.S., Martonosi, M.R.: Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory. In: MobiSys. pp. 127–140 (2011)
3. Liu, Y., Zhao, Y., Chen, L., Pei, J., Han, J.: Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays. TPDS 23(11), 2138–2149 (2012)
4. Maisonneuve, N., Stevens, M., Niessen, M.E., Steels, L.: Noisetube: Measuring and mapping noise pollution with mobile phones. In: ITEE, pp. 215–228 (2009)
5. Martínez-Cánovas, G., Del Val, E., Botti, V., Hernández, P., Rebollo, M.: A formal model based on game theory for the analysis of cooperation in distributed service discovery. INS 326, 59–70 (2016)
6. Nov, O., Naaman, M., Ye, C.: What drives content tagging: the case of photos on flickr. In: Proc. of the SIGCHI. pp. 1097–1100 (2008)
7. Reddy, S., Estrin, D., Srivastava, M.: Recruitment framework for participatory sensing data collections. In: Pervasive Computing, pp. 138–155 (2010)
8. Rowe, M., Poblet, M., Thomson, J.D.: Creating value through crowdsourcing: The antecedent conditions. In: Outlooks and Insights on Group Decision and Negotiation, pp. 345–355 (2015)
9. Spann, M., Tellis, G.J.: Does the internet promote better consumer decisions? the case of name-your-own-price auctions. Journal of Marketing 70(1), 65–78 (2006)
10. del Val, E., Rebollo, M., Botti, V.: Emergence of cooperation through structural changes and incentives in service-oriented mas. In: Proc. of the AAMAS. pp. 1355–1356 (2012)