Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Efficient Collaborative Crowdsourcing

Zhengxiang Pan[1], Han Yu[1], Chunyan Miao[1], Cyril Leung[1,2]
, [1]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly, Nanyang Tech. Univ., Singapore
[2]Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, Canada
{panz0012, han.yu, ascymiao}@ntu.edu.sg, cleung@ece.ubc.ca

Abstract

We consider the problem of making efficient quality-time-cost trade-offs in collaborative crowdsourcing systems in which different skills from multiple workers need to be combined to complete a task. We propose *CrowdAsm* - an approach which helps collaborative crowdsourcing systems determine how to combine the expertise of available workers to maximize the expected quality of results while minimizing the expected delays. Analysis proves that CrowdAsm can achieve close to optimal profit for workers in a given crowdsourcing system if they follow the recommendations.

Introduction

In recent years, intelligent task allocation has been recognized as a useful approach to make efficient quality-time-cost trade-offs in crowdsourcing. Existing approaches assume that a task can be effectively completed by an individual worker (Yu et al. 2013c; 2013a; 2014; 2015). As populations age, inter-generational crowdsourcing platforms start to emerge. Workers with diverse skills need to form teams to collaboratively complete crowdsourcing tasks requiring heterogeneous skills. An example is to proof-read traditional Japanese scripts where elderly workers contribute linguistic expertise while the young assist them with technical knowhow (Kobayashi et al. 2013).

To make efficient quality-time-cost trade-offs in collaborative crowdsourcing systems through intelligent task allocation, we propose the crowd assemble (CrowdAsm) approachIt dynamically assembles teams of workers considering the budgets, the availability of workers with the required skills and their track records in order to complete collaborative crowdsourcing tasks. It maximizes the expected rate of success for the tasks, minimizes the expected time elapse, and stays within the given budget. Through rigorous analysis, we prove that CrowdAsm can achieve near optimal profit for workers in a given collaborative crowdsourcing system if they follow the recommendations.

The CrowdAsm Approach

For a given crowdsourcing system, the availability of workers with a given skill m can be modeled as a queueing system $q_m(t+1) = q_m(t) - \mu_m(t) + a_m(t)$. $\mu_m(t) = \sum_{k=1}^{K} n_{m,k}\tilde{T}_k(t) - \tilde{a}_m(t)$ where $n_{m,k}$ is the number of workers with skill m required to complete a task of type k, and $\tilde{T}_k(t)$ is the actual number of tasks of type k completed during time step t. The range of values for $\tilde{T}_k(t)$ is $0 \leq \tilde{T}_k(t) \leq d_k(t)T_k(t)$. $T_k(t)$ is the number of tasks of type k demanded by the crowdsourcers at time step t. $d_k(t)$ is an indicator function reflecting the ability of the crowdsourcing system to satisfy demands for tasks of type k. $d_k(t) = 1$ if $q_m(t)$ can support $n_{m,k} > 0$ for all m required by a task; otherwise, $d_k(t) = 0$.

Due to the physical limitations of a crowdsourcing system, we assume that there exists an upper bound to $\mu_m(t)$ which can be expressed $\mu_m^{max}(t) = \sum_{k=1}^{K} n_{m,k}T_k^{max}(t)$. $\tilde{a}_m(t)$ is the number of workers with skill m who have become available at time step t (e.g., those who have completed previous tasks allocated to them or new workers joining the system). $a_m(t) \in \mathbf{Z}^+ \cup \{0\}$, is the number of workers to be mobilized by the CrowdAsm approach at t. For a given crowdsourcing system, the number for registered workers is finite. Thus, we assume that there is an upper bound a_m^{max} for $a_m(t)$ such that $0 \leq a_m(t) \leq a_m^{max}$.

Let $c_m(t)$ be the cost required to attract workers with skill m to become available under a given worker supply condition. The quality of the available workers, as measured by an average reliability value $0 < \tilde{r}_m(t) < 1$, is also taken into account. Such a measure $r_i(t)$ for a worker i can be computed based on his past performance using reputation-based approaches (Yu et al. 2013b). Thus, $\tilde{r}_m(t) = \frac{1}{N_m^{out}(t)} \sum_{i=1}^{N_m^{out}(t)} r_i(t)$, where $N_m^{out}(t)$ is the total number of workers with characteristic m who are not currently logged in to the crowdsourcing system and whose $r_i(t) \geq \varepsilon$. ε is a reliability threshold set by the crowdsourcing system or a crowdsourcer (e.g., the minimum "approval rate" in mTurk).

The expected profit for the crowdsourcing system at time step t can be expressed as $\delta(t) = \sum_{k=1}^{K} d_k(t) f_k(p_k(t), \tilde{r}_k(t)) p_k(t) - c(\mathbf{a}(t), \mathbf{x}(t))$, where $\mathbf{a}(t)$ is the set of workers to be mobilized by CrowdAsm and $\mathbf{x}(t)$ is the set of possible supply conditions of workers with various skills. Using the *Lyapunov drift* $\Delta(\mathbf{q}(t))$ (Neely 2010) as a metric to the overall level of congestion of demand on workers with various skills in the crowdsourcing system (i.e., expected delay), we formulate the *delay-minus-profit* objective function as $\Delta(\mathbf{q}(t)) - \rho\mathbb{E}\{\delta(t)|\mathbf{q}(t)\}$ which is to be minimized. $\rho > 0$ is a control variable determining the relative importance of delay and profit which can be set by a crowdsourcing system. $\Delta(\mathbf{q}(t)) \leq \xi + \sum_{m=1}^{M} q_m(t)(a_m(t) - \mu_m(t))$, where $\xi = \frac{1}{2}\sum_{m=1}^{M}[(a_m^{max}(t))^2 + (\mu_m^{max}(t))^2]$. Thus, the objective function can be re-expressed as:

$$\begin{aligned}
&\Delta(\mathbf{q}(t)) - \rho\mathbb{E}\{\delta(t)|\mathbf{q}(t)\} \leq \xi + \sum_{m=1}^{M} q_m(t)[\mathbb{E}\{(a_m(t)|\mathbf{q}(t)\} \\
&- \sum_{k=1}^{K} n_{m,k}\mathbb{E}\{\beta_k \tilde{r}_k(t)p_k(t)|\mathbf{q}(t)\} + \mathbb{E}\{\tilde{a}_m(t)|\mathbf{q}(t)\}] \\
&+ \rho\sum_{k=1}^{K}\mathbb{E}\{d_k(t)[\beta_k \tilde{r}_k(t)p_k(t)]p_k(t)|\mathbf{q}(t)\} \\
&- \rho\sum_{m=1}^{M}\mathbb{E}\{\frac{c_m(t)a_m(t)}{\tilde{r}_m(t)}|\mathbf{q}(t)\}
\end{aligned} \tag{1}$$

By choosing only terms containing $a_m(t)$ from Eq. (1), CrowdAsm needs to minimize the expression $\sum_{m=1}^{M} a_m(t)[q_m(t) - \rho\frac{c_m(t)}{\tilde{r}_m(t)}]$, subject to the constraints: 1) $a_m(t) \leq a_m^{max}$; 2) $r_i(t) \geq \varepsilon, \forall i \in a_m(t)$; and 3) $\sum_{m=1}^{M} c_m(t)a_m(t) \leq B(t)$, where $B(t)$ is the budget available for a given task.

To minimize Eq. (1), compute the values of the expression $[q_m(t) - \rho\frac{c_m(t)}{\tilde{r}_m(t)}]$ for all m. Sort all task requests in ascending order of their $[q_m(t) - \rho\frac{c_m(t)}{\tilde{r}_m(t)}]$ values. For each group of workers with skill m, send task requests to as many workers $a_m(t)$ as allowed by Constraints 1 to 3.

Analysis

In this section, we analyze the performance of the CrowdAsm approach. Specifically, we are interested in the proximity of the time averaged profit for the crowdsourcing system to the optimal time averaged profit.

CrowdAsm observes the queues of available workers with different skills $\mathbf{q}(t)$ at any time step t and helps the crowdsourcing system determine the values of $\mathbf{a}(t)$ to minimize Eq. (1). There exist at least a combination of $a_m^*(t)$ and $\mu_m^*(t)$ values which satisfy all constraints and produce the optimal time averaged profit $\mathbb{E}\{\delta^*(t)|\mathbf{q}(t)\} = \delta^{opt}$ for a given crowdsourcing system such that $\Delta(\mathbf{q}(t)) - \rho\mathbb{E}\{\delta(t)|\mathbf{q}(t)\} \leq \xi - \rho\mathbb{E}\{\delta^*(t)|\mathbf{q}(t)\} + \sum_{m=1}^{M} q_m(t)\mathbb{E}\{a_m^*(t) - \mu_m^*(t)|\mathbf{q}(t)\}$. As the optimal policy results in $\mathbb{E}\{a_m^*(t) - \mu_m^*(t)|\mathbf{q}(t)\} = 0$ for all m, we have $\Delta(\mathbf{q}(t)) - \rho\mathbb{E}\{\delta(t)|\mathbf{q}(t)\} \leq \xi - \rho\delta^{opt}$. Following the definition of $\Delta(\mathbf{q}(t))$ and taking expectations on both sides of the above expression, $\mathbb{E}\{L(\mathbf{q}(t+1))\} - \mathbb{E}\{L(\mathbf{q}(t))\} - \rho\mathbb{E}\{\delta(t)\} \leq \xi - \rho\delta^{opt}$. Summing the above expression over all time steps $t \in \{0, 1, ..., \tau - 1\}$ yields:

$$\begin{aligned}
&\sum_{t=0}^{\tau-1}[\mathbb{E}\{L(\mathbf{q}(t+1))\} - \mathbb{E}\{L(\mathbf{q}(t))\}] - \rho\sum_{t=0}^{\tau-1}\mathbb{E}\{\delta(t)\} \\
&\quad \leq \tau\xi - \tau\rho\delta^{opt}.
\end{aligned} \tag{2}$$

As $\rho > 0$, dividing both sides of the above expression by $\tau\rho$ results in $\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbb{E}\{\delta(t)\} \geq \delta^{opt} - \frac{\xi}{\rho} + \frac{1}{\tau}\mathbb{E}\{L(\mathbf{q}(\tau))\} - \frac{1}{\tau}\mathbb{E}\{L(\mathbf{q}(0))\}$. Since $L(\mathbf{q}(t)) \geq 0$ and $L(\mathbf{q}(0)) = 0$, we have:

$$\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbb{E}\{\delta(t)\} \geq \delta^{opt} - \frac{\xi}{\rho}. \tag{3}$$

Therefore, we have proven that the time averaged profit achievable for a given crowdsourcing system following CrowdAsm is within $O(\frac{1}{\rho})$ of the optimal profit, subject to the physical limitations of the crowdsourcing system.

Conclusions

In this paper, we proposed the CrowdAsm approach to dynamically assemble teams of workers considering the budgets, the availability of workers with the required skills and their track records to complete crowdsourcing tasks requiring collaboration among workers with heterogeneous skills. Theoretical analysis has shown that CrowdAsm can achieve close to optimal profit for workers in a collaborative crowdsourcing system if they follow the recommendations.

Acknowledgements

This research is supported, in part, by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office; and the Lee Kuan Yew Post-Doctoral Fellowship Grant.

References

Kobayashi, M.; Ishihara, T.; Itoko, T.; Takagi, H.; and Asakawa, C. 2013. Age-based task specialization for crowdsourced proofreading. In *LNCS*, volume 8010, 104–112.

Neely, M. J. 2010. *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan and Claypool Publishers.

Yu, H.; Miao, C.; An, B.; Leung, C.; and Lesser, V. R. 2013a. A reputation management model for resource constrained trustee agents. In *IJCAI'13*, 418–424.

Yu, H.; Shen, Z.; Leung, C.; Miao, C.; and Lesser, V. R. 2013b. A survey of multi-agent trust management systems. *IEEE Access* 1(1):35–50.

Yu, H.; Shen, Z.; Miao, C.; and An, B. 2013c. A reputation-aware decision-making approach for improving the efficiency of crowdsourcing systems. In *AAMAS'13*, 1315–1316.

Yu, H.; Miao, C.; An, B.; Shen, Z.; and Leung, C. 2014. Reputation-aware task allocation for human trustees. In *AAMAS'14*, 357–364.

Yu, H.; Miao, C.; Shen, Z.; Leung, C.; Chen, Y.; and Yang, Q. 2015. Efficient task sub-delegation for crowdsourcing. In *AAAI-15*, 1305–1311.