Ethically Aligned Multi-agent Coordination to Enhance Social Welfare

Extended Abstract

Han Yu
School of Computer Science and Engineering, Nanyang Technological University
Singapore
han.yu@ntu.edu.sg

Zhiqi Shen
School of Computer Science and Engineering, Nanyang Technological University
Singapore
zqshen@ntu.edu.sg

Lizhen Cui
School of Software Engineering, Shandong University
Jinan, China
clz@sdu.edu.cn

Yongqing Zheng
School of Software Engineering, Shandong University
Jinan, China
zhengyongqing@dareway.com.cn

Victor R. Lesser
School of Computer Science, University of Massachusetts Amherst
Amherst, MA, USA

ABSTRACT

In multi-agent systems (MASs), the complex interactions among self-interested agents can be modelled as stochastic games. Existing decision support approaches dealing with such situations focus on minimizing individual agent's regret through outperforming other agents in the competitive aspect of the game. Such an approach often results in social welfare not being maximized in the process. In this paper, we propose the regret-minimization-social-welfare-maximization (RMSM) approach. It contains a novel method to quantify how an agent's sacrifice increases and decreases over time based on queueing system dynamics. In this way, ensuring fairness of distribution of sacrifice among agents and compensating for their previous sacrifices can be translated into maintaining the stability of a queueing system.

ACM Reference Format:
Han Yu, Zhiqi Shen, Lizhen Cui, Yongqing Zheng, and Victor R. Lesser. 2019. Ethically Aligned Multi-agent Coordination to Enhance Social Welfare. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

Multi-agent systems (MASs) are often characterized by complex interactions among self-interested agents [7]. In most MASs, the overall social welfare depend on the choices made by all agents. Agents are generally assumed to be strategic and aware of other agents also reasoning strategically. Therefore, MAS research often studies what the most rational decisions by an agent shall be. One of the long-standing goals of collaborative artificial intelligence (AI) is how to enable agents to coordinate with other agents in order to agree on a shared strategy of problem-solving.

Over the years, game theoretic research in MAS has produced many decision support approaches to help individual agents minimize their regret [2–4, 9]. These approaches focus on helping agents select actions with minimal regret without sharing useful private observations of the environment to others. The sharing of such information can benefit the MAS as a whole. Without a mechanism to ensure that individual altruism is reciprocated, rational self-interested agents will not share such information with others.

In order to address this problem, we propose the *regret-minimization-social-welfare-maximization (RMSM)* approach. When an agent shares observations about actions with high payoffs, it is considered to have made a personal sacrifice to enhance the social good of an MAS as other agents may also join in to select these actions (thereby splitting the payoffs). Extensive simulation-based experiments demonstrate that self-interested rational agents learn policies which eventually converge on following RMSM.

2 THE PROPOSED APPROACH

In order to account for the information sharing regret, we introduce the concept of the *default action* by an agent i in round t. It represents an action a self-interested rational agent would select without information sharing from other agents or guidance from a third-party. Methods such as the greedy approach [10] can be used to derive the default actions. We denote the situation in which an agent i selects the default action $a_j(t) \in \mathbf{A}(t)$ in round t of the game without external advice as $d_j^{(i)}(t)$. Thus, the utility produced by $d_j^{(i)}(t)$ is $\hat{u}_i(t) = u(d_j^{(i)}(t), r_j(t))$.

To track the changes in an agent i's information sharing regret over different rounds of the game, we construct a sacrifice measure, $Y_i(t)$, based on queueing theory. The queueing

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

dynamics of $Y_i(t)$ is expressed as:

$$Y_i(t+1) = \max[Y_i(t) + \hat{u}_i(t) - u_i(t), 0]. \tag{1}$$

Here, $Y_i(t)$ is not allowed to drop below 0 as it only tracks the sacrifice made by agent i and the compensations as a result of coordinated action selection it receives. Once $Y_i(t) = 0$, the coordinator agent will not compensate agent i by offering it advantageous action choices. Payoffs received by i under such circumstances will not further reduce $Y_i(t)$.

In addition to measuring individual sacrifice over time, we also want to distribute the sacrifice among agents in as fairly a manner as allowed by the circumstances. In this way, the approach follows the Ethically Aligned Design (EAD) guidelines [12, 19] specified by the IEEE. For this purpose, we adopt the quadratic Lyapunov function [5, 6, 13–18], $L(t)$, to measure the distribution of sacrifice among agents at t as $L(t) = \frac{1}{2}\sum_{i=1}^{N} Y_i^2(t)$. A large $L(t)$ indicates that either many agents have made large sacrifices, or a small number of agents have made very large sacrifices or a combination of such situations, all of which are to be minimized. From the perspective of a given MAS, the overall objective is to maximize social welfare while minimizing agents' sacrifice:

max:

$$\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{N} u(a_j^{(i)}(t), r_j(t))[\omega + Y_i(t)] \tag{2}$$

s. t.:

$$a_j^{(i)}(t) \in \mathbf{A}(t), \forall i, t \tag{3}$$

where $\omega > 0$ is a control variable to determine the relative weight assigned to maximizing social welfare vs. minimizing sacrifice regret. The proposed RMSM approach to maximize equation (2) is given in Algorithm 1.

Algorithm 1 RMSM

Require: ω; $Y_i(t)$, $d_j^{(i)}(t)$ and $\mathbf{S}_i(t)$ from all i and t.
Aggregate $\mathbf{S}_i(t)$ from all agents to estimate $\mathbf{R}(t)$;
Sort all N agents in descending order of $\Upsilon_i(t)$;
for $i = 1$ to N **do**
  $a_j^{(i)}(t) = \underset{a_j(t)\in\mathbf{A}(t)}{\operatorname{argmax}}\ u(a_j(t), r_j(t))$ given the prescribed actions of previous agents;
  if multiple eligible $a_j(t)$ values are found **then**
    Choose the $a_j(t)$ selected by the fewest other agents;
  end if
end for
for $i = 1$ to N **do**
  $Y_i(t+1) = \max[Y_i(t) + \hat{u}_i(t) - u_i(t), 0]$;
end for
return $\left\{a_j^{(1)}(t), a_j^{(2)}(t), ..., a_j^{(N)}(t)\right\}$;

3 EXPERIMENTAL EVALUATION

To study whether a self-interested rational agent has incentive to follow an approach other than RMSM, we designed a simulation-based experiment. The agents in this experiment can choose to follow any of the five comparison approaches (i.e. arg max, softmax, softmax (S) which is softmax with information sharing, RLARM [8], and RMSM) at any point during the simulation. In the beginning of the experiment, each agent follows each of the comparison approaches with equal probability (i.e. 20%). An agent keep track of the average reward it has derived so far following each of the comparison approaches. Each comparison approach is regarded as an arm in an n-Armed Bandit model [1] ($n = 5$). Each agent learns an approach selection policy over time following the decreasing ϵ-greedy strategy [11].

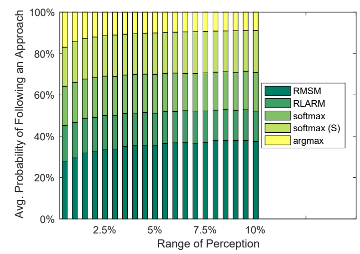

Figure 1: Agents' learnt behaviours

Figure 1 shows the probability of adopting each of the five comparison approaches at any given time step averaged over 10,000 agents after 1,000 rounds of simulation. Agents' range of perception variables are increased from 0.5% to 10% of the entire action space in steps of 0.5%, thereby increasing the degree of overlap of observations among agents. It can be observed that, under low overlaps in agents' perception, the advantage of RMSM compared to other approaches is less pronounced than under higher overlaps in agent perceptions. Nevertheless, under the different range of perception settings, self-interested agents always learns to follow RMSM with the highest probability. Agents who deviate from RMSM to follow other approaches receive worse rewards than those following RMSM. Over time, as ϵ becomes close to 0, all self-interested rational agents converge to adopting the RMSM approach.

4 CONCLUSIONS

In this paper, we propose the RMSM approach to help multi-agent games maximize social welfare while minimizing information sharing regret. Under the conditions facing MASs investigated by this paper, seemingly self-interest preserving local decisions made by agents may actually hurt their self-interest if done in an uncoordinated fashion. To promote information sharing and support coordinated actions among agents, a novel method to quantify the increase and decrease of an agent's sacrifice as a result of sharing private observations of action rewards over time following queueing system dynamics has been developed. It provides a framework to coordinate multi-agent sacrifice to enhance social welfare.

ACKNOWLEDGEMENTS

This research is supported, in part, by Nanyang Technological University, Nanyang Assistant Professorship (NAP).

REFERENCES

[1] M. N. Katehakis and A. F. Veinott. 1987. The Multi-Armed Bandit Problem: Decomposition and Computation. *Mathematics of Operations Research* 12, 2 (1987), 262–268.

[2] Marc Lanctot, Neil Burch, Martin Zinkevich, Michael Bowling, and Richard G. Gibson. 2012. No-Regret Learning in Extensive-Form Games with Imperfect Recall. In *Proceedings of the 29th International Conference on Machine Learning (ICML'12)*. 65–72.

[3] Boyang Li, Han Yu, Zhiqi Shen, Lizhen Cui, and Victor R. Lesser. 2015. An Evolutionary Framework for Multi-agent Organizations. In *Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'15)*. 35–38.

[4] Chunyan Miao, Han Yu, Zhiqi Shen, and Cyril Leung. 2016. Balancing quality and budget considerations in mobile crowdsourcing. *Decision Support Systems* 90 (2016), 56–64.

[5] MIchael J. Neely. 2010. *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan and Claypool Publishers. 211 pages.

[6] Zhengxiang Pan, Han Yu, Chunyan Miao, and Cyril Leung. 2016. Efficient Collaborative Crowdsourcing. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*. 4248–4249.

[7] Simon Parsons and Michael Wooldridge. 2002. Game Theory and Decision Theory in Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems* 5, 3 (2002), 243–254.

[8] Gabriel de O. Ramos, Bruno C. da Silva, and Ana L.C. Bazzan. 2017. Learning to Minimise Regret in Route Choice. In *Proceedings of the 16th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'17)*. 846–855.

[9] Zhiqi Shen, Han Yu, Chunyan Miao, and Jianshu Weng. 2011. Trust-based web service selection in virtual communities. *Web Intelligence and Agent Systems* 9, 3 (2011), 227–238.

[10] Eilon Solan and Nicolas Vieille. 2015. Stochastic games. *Prceedings of the National Academy of Sciences USA (PNAS)* 112, 45 (2015), 13743–13746.

[11] Richard S. Sutton and Andrew G. Barto. 1998. *Introduction to Reinforcement Learning*. The MIT Press.

[12] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2018. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2*. Technical Report. IEEE.

[13] Han Yu, Chunyan Miao, Bo An, Cyril Leung, and Victor R. Lesser. 2013. A Reputation Management Model for Resource Constrained Trustee Agents. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*. 418–424.

[14] Han Yu, Chunyan Miao, Bo An, Zhiqi Shen, and Cyril Leung. 2014. Reputation-aware Task Allocation for Human Trustees. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'14)*. 357–364.

[15] Han Yu, Chunyan Miao, Yiqiang Chen, Simon Fauvel, Xiaoming Li, and Victor R. Lesser. 2017. Algorithmic management for improving collective productivity in crowdsourcing. *Scientific Reports* 7, 12541 (2017), doi:10.1038/s41598–017–12757–x.

[16] Han Yu, Chunyan Miao, Cyril Leung, Yiqiang Chen, Simon Fauvel, Victor R. Lesser, and Qiang Yang. 2016. Mitigating Herding in Hierarchical Crowdsourcing Networks. *Scientific Reports* 6, 4 (2016), doi:10.1038/s41598–016–0011–6.

[17] Han Yu, Chunyan Miao, Zhiqi Shen, Cyril Leung, Yiqiang Chen, and Qiang Yang. 2015. Efficient Task Sub-delegation for Crowdsourcing. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15)*. 1305–1311.

[18] Han Yu, Zhiqi Shen, Chunyan Miao, and Bo An. 2013. A Reputation-aware Decision-making Approach for Improving the Efficiency of Crowdsourcing Systems. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'13)*. 1315–1316.

[19] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, Victor R. Lesser, and Qiang Yang. 2018. Building Ethics into Artificial Intelligence. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)*. 5527–5533.