An Optimal Rewiring Strategy for Cooperative Multiagent Social Learning

Extended Abstract

Hongyao Tang, Jianye Hao, Li Wang, Zan Wang
College of Intelligence and Computing, Tianjin University
Tianjin, China
{bluecontra,jianye.hao,wangli,wangzan}@tju.edu.cn

Tim Baarslag
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
T.Baarslag@cwi.nl

ABSTRACT

Multiagent coordination is a key problem in cooperative multiagent systems (MASs). It has been widely studied in both fixed-agent repeated interaction setting and static social learning framework. However, two aspects of dynamics in real-world MASs are currently neglected. First, the network topologies can change during the course of interaction dynamically. Second, the interaction utilities can be different among each pair of agents and usually unknown before interaction. Both issues mentioned above increase the difficulty of coordination. In this paper, we consider the multiagent social learning in a dynamic environment in which agents can alter their connections and interact with randomly chosen neighbors with unknown utilities beforehand. We propose an optimal rewiring strategy to select most beneficial peers to maximize the accumulated payoffs in long-run interactions. We empirically demonstrate the effects of our approach in a variety of large-scale MASs.

KEYWORDS

Learning agent-to-agent interactions (negotiation, trust, coordination); Multiagent learning; Social simulation

ACM Reference Format:
Hongyao Tang, Jianye Hao, Li Wang, Zan Wang and Tim Baarslag . 2019. An Optimal Rewiring Strategy for Cooperative Multiagent Social Learning. 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

Multiagent coordination in cooperative multiagent systems (MASs) is a significant and widely studied problem. It requires agents to have the capability of coordinating with others effectively towards desirable outcomes. A wide spectrum of works have studied the multiagent coordination problems in cooperative MASs [4, 6–9, 11]. One line of research is multiagent social learning that study the multiagent coordination problem among a population of cooperative agents with sparse and local interactions [1, 5, 10, 12, 13, 15].

However, most existing works under the social learning framework assume that agents are located in a static network. Thus, two aspects of dynamics in real-world MASs are currently neglected. First, the interaction utilities for agent pairs are not necessary to be identical due to the difference of agents' preferences and the contexts they are situated in [2, 16]. Second, the network topologies can be dynamic, i.e. agent changes their interacting partners autonomously. To this end, we study the multiagent coordination in cooperative MASs with taking above two aspects into consideration. We consider a dynamic environment where agents can alter their connections by rewiring, and propose an optimal rewiring approach to select most beneficial peers among all reachable peers to maximize the accumulative payoff during the long-run interactions.

2 PROBLEM DESCRIPTION

We consider a population of agents N, in which each agent i has a set of *reachable peers*, defined as $\{O_i \cup \bar{O}_i\}$. Agent i can only interact with its neighborhood O_i through the connections, and also has a probability φ to be able to establish a new connection to a potential agent $j \in \bar{O}_i$ with cost c_j^i through rewiring. For each rewiring, an old connection should be broken before establishing a new one to model agents' limited communication ability in practice [16].

We model the strategic interaction between each pair of agents as a cooperative game. A general form of two-action cooperative games between agent i and j is denoted as $G_i^j = [u_a, \alpha, \alpha, u_b]$, where u_a (or u_b) is the payoff when agent i and j both choose action a (or b) and α ($\leq u_a(u_b)$) is the outcome for mis-coordination. To model the uncertainty and diversity of agents' utilities, the coordination payoff u_a (or u_b) is sampled from a stochastic variable x_a (or x_b) following a cumulative probability distribution $F_a(x)$ (or $F_b(x)$). Moreover, $F_a(x)$ (or $F_b(x)$) is unique for each game. The value of u_a (or u_b) is unknown before interaction and is revealed when the corresponding outcome is reached once. Each agent can observe the actions of its interaction neighbor at the end of each interaction.

3 OPTIMAL REWIRING STRATEGY

The overall interaction protocol is shown in Algorithm 1, including rewiring phase (Line 2-4) and interaction phase (Line 5-7).

Algorithm 1 Overall interaction protocol for agent $i \in N$.

```
1: for a number of interaction rounds do
2:     if random variable p ≤ φ then
3:         Perform rewiring action (including NOOP).
4:     end if
5:     Play game G_i^j with randomly chosen player j ∈ O_i.
6:     Obtain payoff and update its policy.
7:     Update neighbor j's action model.
8: end for
```

* Corresponding author: Jianye Hao.

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.

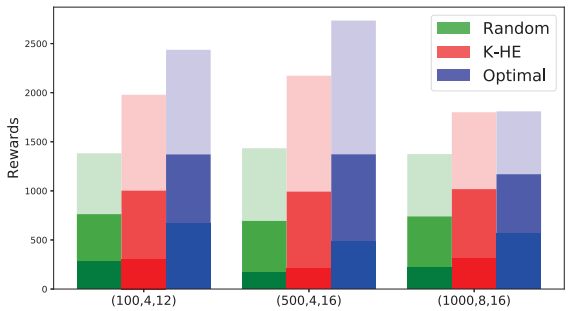


Figure 1: The rewards of rewiring strategies in different topologies (horizontal axis). The tuple denotes the initial size of agents, neighborhood and reachable peers. For each color, three degrees of opacity represent the value of minimum, average and maximum separately.

Estimation of Expected Interaction Payoff

The expected payoffs of agent i interacting with j according to an known or unknown payoff matrix G_i^j, i.e., v_i^j or x_i^j, are evaluated respectively as follows:

$$v_i^j = \max_{m \in A_i} p_i^j(m) u_m + \left(1 - p_i^j(m)\right) \alpha, \tag{1}$$

$$x_i^j = \max_{m \in A_i} p_i^j(m) x_m + \left(1 - p_i^j(m)\right) \alpha. \tag{2}$$

Agent j's policy p_i^j can be estimated from historical actions.

K-Sight Index and Rewiring Strategy

Each agent's situated environments are continuously changing due to rewiring and thus we model it as an Markov Decision Process. Each state s of agent i can be represented as a tuple $\langle \bar{O}_i, y_i \rangle$, with the set of potential peers $\bar{O}_i$ and the current baseline value $y_i = \min_{j \in O_i} v_i^j$. With a long sight, the K-step utility function $U_K(\pi_i^*, s)$ of an optimal strategy π_i^* is formulated as,

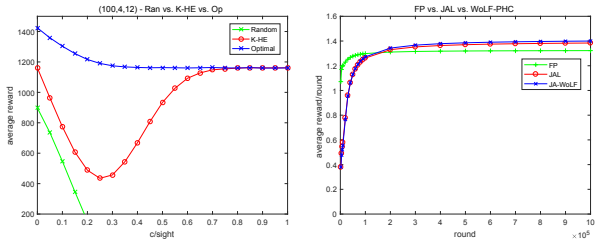
$$\begin{aligned} U_K(\pi_i^*, s) = \max \Big\{ & K y_i + U_K(\pi_i^*, \langle \bar{O}_i, y_i' \rangle), \\ & \max_{j \in \bar{O}_i} \left\{ -c_i^j + K y_i' + U_K(\pi_i^*, \langle \bar{O}_i \backslash \{j\}, y_i' \rangle) \right\} \Big\}. \end{aligned} \tag{3}$$

To compute the optimal policy in Equation 3, inspired from Pandora's Rule [14] and Negotiation Problem [2], we calculate the K-sight rewiring index Λ_i^j as follows,

$$\Lambda_i^j = \int_{-\infty}^{\infty} y_i' \cdot \mathrm{d}F_i^j(x) - y_i, \tag{4}$$

where y_i' is the new baseline value after rewiring and $F_i^j(x)$ is the distribution of x_i^j (Equation 2). The benefit index Λ_i^j captures the relevant information about agent j: it should be rewired when it has the highest positive value of the accumulated net benefit in the following K rounds interaction.

Thus, we propose K-sight rewiring strategy: at each rewiring phase, an agent i first calculates the interaction baseline value y_i. Second, for each potential peer $j \in \bar{O}_i$, the index Λ_i^j is computed by following Equation 4. Finally, agent i choose the agent t with highest benefit index as its rewiring target, i.e., $\Lambda_i^t = \Lambda_{max} = \max_{j \in \bar{O}_i}(K\Lambda_i^j - c_i^j)$. Then agent i makes the rewiring decision accordingly: to rewire target agent t if $K\Lambda_{max} - c_i^t \geq 0$, or not to rewire otherwise. For each rewiring, agent i breaks the worst connection, i.e., $\arg\min_{j \in O_i} v_i^j$.



(a) (b)

Figure 2: Performance comparison for (a) different rewiring strategies and (b) different interaction strategies.

Interaction Strategies

We consider three representative learning strategies in interaction phase, i.e., Fictitious play (FP), Joint-Action Learner (JAL) [4] and Joint-Action WoLF-PHC (JA-WoLF) [3].

4 EXPERIMENTAL EVALUATIONS

To evaluate our K-sight rewiring strategy (Optimal), we compare it with two benchmark strategies, i.e., Random and K-sight Highest Expect (K-HE) that rewires the agent with the highest positive value of K-round expected payoff minus the cost.

First, we conduct experiments under different topologies for each rewiring strategy. The average accumulated payoff over 1000 rounds of each agent are shown in Figure 1. We can observe that our optimal rewiring strategy outperforms benchmark strategies in terms of average, best and worst cases across all settings. Second, in Figure 2(a) we evaluate our approach under the settings with the rewiring cost c varying in the range of $[0.0, 200.0]$ and the fixed $K = 200$. The results show that our approach significantly outperforms others across almost all c/K settings. For both Figure 1 and Figure 2(a), we use FP as the interaction strategy.

Moreover, we analyze the performance of three interaction strategies with our rewiring strategy. Figure 2(b) shows the average single-round interaction payoffs of each agent during the long-term interaction. We can observe that FP strategy can fast reach a good payoff level while JAL and JA-WoLF outperform FP in the long term due to their better convergence on optimal Nash equilibrium.

5 ACKNOWLEDGEMENTS

The work is supported by the National Natural Science Foundation of China (Grant Nos.: 61702362, U1836214), Special Program of Artificial Intelligence, Tianjin Research Program of Application Foundation and Advanced Technology (No.: 16JCQNJC00100), Special Program of Artificial Intelligence of Tianjin Municipal Science and Technology Commission (No.: 569 17ZXRGGX00150), and the Netherlands Organisation for Scientific Research (No.: 639.021.751).

REFERENCES

[1] S. Airiau, S. Sen, and D. Villatoro. 2014. Emergence of conventions through social learning - Heterogeneous learners in complex networks. *AAMAS* 28, 5 (2014), 779–804.

[2] T. Baarslag and E. H. Gerding. 2015. Optimal Incremental Preference Elicitation during Negotiation. In *IJCAI*. 3–9.

[3] M. H. Bowling and M. M. Veloso. 2001. Rational and Convergent Learning in Stochastic Games. In *IJCAI*. 1021–1026.

[4] C. Claus and C. Boutilier. 1998. The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems. In *AAAI*. 746–752.

[5] J. Hao and H. Leung. 2013. The Dynamics of Reinforcement Social Learning in Cooperative Multiagent Systems. In *IJCAI*. 184–190.

[6] S. Kapetanakis and D. Kudenko. 2002. Reinforcement Learning of Coordination in Cooperative Multi-Agent Systems. In *AAAI*. 326–331.

[7] M. Lauer and M. A. Riedmiller. 2000. An Algorithm for Distributed Reinforcement Learning in Cooperative Multi-Agent Systems. In *ICML*. 535–542.

[8] L. Matignon, G. Laurent, and N. Le Fort-Piat. 2008. A study of FMQ heuristic in cooperative multi-agent games.. In *AAMAS Workshop on Multi-Agent Sequential Decision Making in Uncertain Multi-Agent Domains*, Vol. 1. 77–91.

[9] L. Matignon, G. J. Laurent, and Nadine Le Fort-Piat. 2012. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. *Knowledge Eng. Review* 27, 1 (2012), 1–31.

[10] M. Mihaylov, K. Tuyls, and A. Nowé. 2014. A decentralized approach for convention emergence in multi-agent systems. *AAMAS* 28, 5 (2014), 749–778.

[11] L. Panait and S. Luke. 2005. Cooperative Multi-Agent Learning: The State of the Art. *AAMAS* 11, 3 (2005), 387–434.

[12] S. Sen and S. Airiau. 2007. Emergence of Norms through Social Learning. In *IJCAI*. 1507–1512.

[13] D. Villatoro, J. Sabater-Mir, and S. Sen. 2011. Social Instruments for Robust Convention Emergence. In *IJCAI*. 420–425.

[14] M. L. Weitzman. 1979. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society* (1979), 641–654.

[15] C. Yu, M. Zhang, F. Ren, and X. Luo. 2013. Emergence of social norms through collective learning in networked agent societies. In *AAMAS*. 475–482.

[16] C. Zhang and V. R. Lesser. 2013. Coordinating multi-agent reinforcement learning with limited communication. In *AAMAS*. 1101–1108.