To be Big Picture Thinker or Detail-Oriented? Utilizing Perceived Gist Information to Achieve Efficient Convention Emergence with Bilateralism and Multilateralism

Extended Abstract

Shuyue Hu, Chin-wing Leung, Ho-fung Leung
The Chinese University of Hong Kong
Hong Kong, China
{syhu,cwleung,lhf}@cse.cuhk.edu.hk

Jiamou Liu
The University of Auckland
Auckland, New Zealand
jiamou.liu@auckland.ac.nz

ABSTRACT

Recently, the study of social conventions (or norms) has attracted much attention. In this paper, we study the emergence of conventions from agents' repeated coordination games via bilateralism and multilateralism. We assume that agents can perceive the gist information, i.e., a big picture of how popular each action is in their neighbourhood. A novel reinforcement learning approach which utilizes the gist information is proposed. Experiment verifies that the proposed approach significantly outperforms the baseline and the state-of-the-art approaches, in terms of the speed of convention emergence.

KEYWORDS

Norm; Convention Emergence; Fuzzy Trace Theory

ACM Reference Format:
Shuyue Hu, Chin-wing Leung, Ho-fung Leung and Jiamou Liu. 2019. To be Big Picture Thinker or Detail-Oriented? Utilizing Perceived Gist Information to Achieve Efficient Convention Emergence with Bilateralism and Multilateralism. 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

Social conventions (or norms), such as driving on a particular side of the roads, have shown to be an efficient mechanism to solve coordination problems in human society [8]. Recently, the concept of social conventions has attracted much attention in multi-agent system research, since they may achieve systemwide coordination among agents [2–5]. By the *90% convergence* metric [7], a convention emerges in a system, if at least 90% of the agents coordinate by playing the same action. To introduce conventions into multi-agent systems, early research works [7, 13, 20] reveal that conventions can naturally emerge from agents' learning to play repeated coordination games. Subsequently, how to speed up convention emergence has become one of the main concerns in the literature [6, 15, 19].

In this paper, we accelerate convention emergence from networked multi-agent systems in which agents from time to time require coordination with all of their neighbours. We distinguish two manners to achieve coordination among agents: bilateralism and multilateralism. By bilateralism, agents "divide and conquer" and attempt to coordinate with each of their neighbours one by one. By multilateralism, agents attempt to achieve coordination with the entire neighbourhood all at once. We formalize these two manners as 2-player and n-player coordination games respectively. Although multilateral coordination has wide application [9–11], it is inadequately studied in the research of convention emergence.

According to the *fuzzy trace theory* [1], people in parallel form two types of mental representation for an event: *verbatim trace* and *gist trace* (or *gist*). While verbatim trace is the fine-grained representation which preserves abundant details, gist trace is the more generalized representation which only preserves the information of essence. Consider an example in [1], suppose that people, presented with two objects A and B, are informed of their lengths. People's verbatim trace about these two objects is the exact length of each object. On the contrary, the gist trace is simply which object is longer (or shorter).

We identify that most recent works integrate the verbatim information about each agent's neighbours into the agent's learning [12, 16, 17, 21], so as to speed up convention emergence. For example, an agent learns its strategy based on which action each of its neighbours uses [12, 17, 21], or more specifically, how each of its neighbours evaluates actions [16]. However, the fuzzy trace theory reveals that people actually prefer to base their decisions in daily life on the more generalized gist traces. Mimicking such preference, we define the gist trace for individual agent as how prevalent it perceives each action should be, and propose gist-based Q-learning. We experimentally compare gist-based Q-learning with the state-of-the-art approaches which are based on verbatim information. The results clearly indicate the significant outperformance of gist-based Q-learning, in terms of accelerating convention emergence.

2 CONVENTION EMERGENCE FRAMEWORK

We consider that at each time step, a proportion σ of agents are randomly selected to initiate coordination games with all of their neighbours. A higher value of σ reflects the more intensive need for coordination. By bilateralism, each initiating agent plays a 2-player coordination game with each of its neighbours. For each game, the two players are rewarded with payoff $+1$ if they coordinate by playing the same action, but are punished with payoff -1 if they play different actions. By multilateralism, each initiating agent plays an n-player coordination game with all of its neighbours, where n is the number of its neighbours plus itself. For each game, if all of the n players play the same action, they are fully rewarded with payoff $+1$. However, practically, if at least 75% of the n players

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.

coordinate to play the same action, the coordinating players are moderately rewarded with payoff $+0.5$. Otherwise, the players are punished with payoff -1 for discoordination. Note that by either bilateralism or multilateralism, agents may play more than one game at one time step. We here assume that the immediate payoff that an agent receives is averaged over all of the games it plays at each time step. Based on the received payoff, agents, aiming to maximize their own future payoff, reevaluate their strategy and choose the action for next time step.

3 GIST-BASED Q-LEARNING

Under the framework described above, a rational strategy for individual agent is to choose the most popular action among its neighbours. Therefore, the gist, the information that is essential for decision-making, should be which action is most popular among the neighbours. Intuitively, such information can be reflected by the proportion of neighbours using each action. For a certain action, a higher proportion of usage indicates that the action is more widely used, and thus is more popular in the neighbourhood. Consider a particular agent i with a set $A = \{a_1, \ldots, a_k\}$ of k available actions. Let $p_j^{i,t}$ denote the proportion of agent i's neighbours using action a_j at time t. We define the gist $\mathbf{p}^{i,t}$ for agent i at time t to be a vector of the proportion of usage of each action in its neighbourhood, i.e., $\mathbf{p}^{i,t} = [p_1^{i,t} \ \ldots \ p_k^{i,t}]^\intercal$.

We modify Q-learning [18] so that agents can learn the expected payoff of performing each action from their gist traces. We assume that, at time t, agent i's Q-value $Q_j^{i,t}$ for any action $a_j \in A$ largely depends on the agent's current gist trace $\mathbf{p}^{i,t}$, and takes the following form:

$$Q_j^{i,t} = b_j^i + \mathbf{c}_{-j}^{i}{}^\intercal \mathbf{p}_{-j}^{i,t} \tag{1}$$

where b_j^i denotes inherently how good of using action a_j is regardless of what others play; $\mathbf{p}_{-j}^{i,t} = [p_1^{i,t} \ \ldots \ p_{j-1}^{i,t} \ p_{j+1}^{i,t} \ \ldots \ p_k^{i,t}]^\intercal$ is deduced from agent i's gist trace $\mathbf{p}^{i,t}$ at time t, and $\mathbf{c}_{-j}^i = [c_{j,1}^i \ \ldots \ c_{j,j-1}^i \ c_{j,j+1}^i \ \ldots \ c_{j,k}^i]^\intercal$ is a vector of $k-1$ parameters, each of which, say, $c_{j,k}^i$, reflects the correlation between agent i's Q-value of action a_j and the proportion of usage of action a_k in the neighbourhood. During interactions, each agent i learns the values of b_j^i and each element in $\mathbf{c}_{-j}^i$ based on stochastic gradient descent method [14]. At the end of each time step, each agent selects the action with the highest Q-value with ϵ-greedy exploration for the next time step.

4 EXPERIMENTS

In our experiments, we consider a system of 500 agents, each of which has 2 available actions, and place them on random network with average degree 20. We compare gist-based Q-learning with three other learning approaches in convention emergence that can be directly applied to the framework described in Section 2, namely, Q-learning [18], collective learning [21] and multiple Q-learning [16]. While Q-learning is the commonly adopted baseline, collective learning and multiple Q-learning are the state-of-the-art approaches which utilize verbatim information.

Table 1: The average number of time steps required to achieve convention emergence under different settings. Q: Q learning, CL: collective learning, MQ: multiple Q-learning, and GQ: gist-based Q-learning.

	Bilateralism				Multilateralism			
	Q	CL	MQ	GQ	Q	CL	MQ	GQ
σ=0.05	13	15	**12**	14	N/A	N/A	1,898	**190**
σ=0.2	11	10	7	**5**	3,430	2,946	923	**127**
σ=0.5	6	6	5	**4**	3,090	2,737	634	**125**
σ=1.0	**4**	**4**	**4**	**4**	1,909	1,619	523	**124**

We vary the proportion σ of initiating agents in 0.05, 0.2, 0.5 and 1, so that its effect on convention emergence can be clearly identified. When σ is 0.05, at each time step, there are 25 agents initiating coordination games, and thus each agent practically participates in one game. As σ becomes larger, there will be more initiating agents, and hence agents participate in increasingly more games. For statistical purpose, we conduct 100 simulations for each of the settings. Each simulation consists of 5,000 time steps.

In Table 1, we present the average number of time steps required to achieve convention emergence, measured by 90% convergence metric, under different approaches. It is shown that when agents attempt to achieve coordination with bilateralism, all of the approaches are comparably fast and establish conventions within more or less 10 time steps. However, when agents attempt to achieve coordination with multilateralism, gist-based Q-learning achieve convention emergence much faster than the other approaches. In particular, when the proportion σ of initiating agents is as low as 0.05, while Q-learning and collective learning generally fail to establish conventions within 5,000 time steps, and multiple Q-learning spends around 1,900 time steps, gist-based Q-learning spends around 190 time steps to establish conventions.

We observe that under gist-based Q-learning, for any action $a_j \in A$, agents develop negative values for elements in c_{-j}^i. Intuitively, this means agents establish a negative correlation between the Q-value of an action and the proportion of usage of the other actions. That is, the less widely used the other actions are, the higher Q-value this action should be and vice versa. We conjecture that this directly contributes to the good performance of gist-based Q-learning. With such negative correlation, agents may choose the most popular action throughout the system, which should have the highest Q-value. This reinforces the popularity of the most popular action, which could be just slightly more popular than the others, and thus leads to fast convention emergence.

5 CONCLUSIONS

We propose a novel learning approach, gist-based Q-learning, for convention emergence through agents' bilateral and multilateral coordination. Experimental results show that when agents achieve coordination with bilateralism, gist-based Q-learning is comparable to the state-of-the-art approaches which are based on verbatim information. When agents achieve coordination with multilateralism, gist-based Q-learning establishes conventions significantly faster than the others.

REFERENCES

[1] Charles J Brainerd and Valerie F Reyna. 1990. Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review* 10, 1 (1990), 3–47.

[2] Henry Franks, Nathan Griffiths, and Arshad Jhumka. 2013. Manipulating convention emergence using influencer agents. *Autonomous Agents and Multi-Agent Systems* 26, 3 (2013), 315–353.

[3] Jianye Hao and Ho-fung Leung. 2013. The Dynamics of Reinforcement Social Learning in Cooperative Multiagent Systems.. In *Proceedings of 23rd International Joint Conference on Artificial Intelligence*, Vol. 13. 184–190.

[4] Mohammad Rashedul Hasan, Anita Raja, and Ana Bazzan. 2015. Fast Convention Formation in Dynamic Networks Using Topological Knowledge.. In *Proceedings of 29th AAAI Conference on Artificial Intelligence*. 2067–2073.

[5] Shuyue Hu and Ho-fung Leung. 2017. Achieving coordination in multi-agent systems by stable local conventions under community networks. In *Proceedings of 26th International Joint Conference on Artificial Intelligence*. 4731–4737.

[6] Shuyue Hu and Ho-fung Leung. 2018. Compromise as a Way to Promote Convention Emergence and to Reduce Social Unfairness in Multi-Agent Systems. In *Proceedings of Australasian Joint Conference on Artificial Intelligence*. Springer, 3–15.

[7] J. E. Kittock. 1993. Emergent conventions and the structure of multi-agent systems. In *Proceedings of the 1993 Santa Fe Institute Complex Systems Summer School*, Vol. 6.

[8] D. Lewis. 1969. Convention: a philosophical study. (1969).

[9] Walter Mattli. 2001. The politics and economics of international institutional standards setting: an introduction. *Journal of European Public Policy* 8, 3 (2001), 328–344.

[10] Mihail Mihaylov, Karl Tuyls, and Ann Nowé. 2014. A decentralized approach for convention emergence in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 28, 5 (2014), 749–778.

[11] Rajesh Raman, Miron Livny, and Marvin Solomon. 2000. Resource management through multilateral matchmaking. In *Proceedings of 9th International Symposium on High-Performance Distributed Computing*. 290–291.

[12] Ryosuke Shibusawa and Toshiharu Sugawara. 2014. Norm emergence via influential weight propagation in complex networks. In *Proceedings of the European Network Intelligence Conference*. 30–37.

[13] Yoav Shoham and Moshe Tennenholtz. 1997. On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence* 94, 1 (1997), 139–166.

[14] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[15] Daniel Villatoro, Jordi Sabater-Mir, and Sandip Sen. 2011. Social instruments for robust convention emergence. In *Proceedings of 22nd International Joint Conference on Artificial Intelligence*, Vol. 11. 420–425.

[16] Yixi Wang, Wenhuan Lu, Jianye Hao, Jianguo Wei, and Ho-fung Leung. 2018. Efficient Convention Emergence through Decoupled Reinforcement Social Learning with Teacher-Student Mechanism. In *Proceedings of 17th International Conference on Autonomous Agents and MultiAgent Systems*. 795–803.

[17] Yuchen Wang, Fenghui Ren, and Minjie Zhang. 2018. A Cyclical Social Learning Strategy for Robust Convention Emergence. In *Proceedings of International Conference on Agents*. 120–125.

[18] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine Learning* 8, 3-4 (1992).

[19] Tianpei Yang, Zhaopeng Meng, Jianye Hao, Sandip Sen, and Chao Yu. 2016. Accelerating Norm Emergence Through Hierarchical Heuristic Learning.. In *Proceedings of European Conference on Artificial Intelligence*. 1344–1352.

[20] H Peyton Young. 1996. The economics of convention. *Journal of Economic Perspectives* 10, 2 (1996), 105–122.

[21] Chao Yu, Minjie Zhang, Fenghui Ren, and Xudong Luo. 2013. Emergence of social norms through collective learning in networked agent societies. In *Proceedings of 12th International Conference on Autonomous Agents and Multiagent Systems*. 475–482.