Efficient and Robust Emergence of Conventions through Learning and Staying

Wei Liu
Department of Computer
and Information Science
Southwest University
Chongqing, China
lwmiracle123@email.swu.edu.cn

Shuyue Hu
Department of Computer Science
and Engineering
The Chinese University of Hong Kong
Hong Kong, China
syhu@cse.cuhk.edu.hk

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

Wu Chen (corresponding author)
Institute of Logic and Intelligence
Southwest University
Chongqing, China
chenwu@swu.edu.cn

Siyuan Chen
Department of Computer
and Information Science
Southwest University
Chongqing, China
sychen@email.swu.edu.cn

Yong Yu
Department of Computer
and Information Science
Southwest University
Chongqing, China
iyuyong@email.swu.edu.cn

***Abstract*—In a multi-agent system (MAS), conventions serve as an effective mechanism to reduce frictions among agents and hence solve coordination problems. Convention emergence studies how agents' behavior patterns give rise to conventions and how efficiently a convention forms. In a networked MAS, the question focuses on how conventions can arise when the agents' positions are constrained. In this paper, we investigate convention emergence under the multi-player synchronous interaction model in networked MASs. In particular, we focus on the scenario that the agents is not informed the actions played by other agents, and the only information agents can perceive is whether an interaction is success or not. To facilitate the emergence of conventions, we propose a novel approach, namely Win-Stay-Lose-Learn (WSLL), to solve the problem of no observation and shorten the action transformation time when convention seeds conflict among agents. We conduct experiments to verify the robustness and effectiveness of our proposed method, experimental results show that our method outperforms other baseline approaches in terms of convergence speed under various circumstances.**

***Index Terms*—convention emergence, coordination, networked MASs, reinforcement learning**

I. Introduction

Social conventions, such as driving on a particular side of the road and using the same channel for message dissemination in wireless sensor networks, is an effective mechanism to achieve coordination in both human society and multi-agent systems (MASs). In MAS research, a convention is usually defined to be "a social law that restricts the agents behavior to one particular strategy [1]". To introduce conventions into MASs, there are two lines of approaches: the prescriptive approach [2]–[4] and the emergence approach [5]–[7]. While the former one assumes that a priori existence of conventions, the latter one addresses conventions as the natural result of local interactions among agents and is thus more desirable for distributed MASs [8].

Since the early works [1], [9], one main research question of the emergence approach has been: what leads to the efficient emergence of conventions? To date, a number of mechanisms that can be further categorized into two classes: the spreading-based mechanism and the learning-based mechanism, are proposed. The spreading-based mechanisms exquisitely specify how individual agents should behave to spread the convention seeds, and usually equip agents with the capability of observation and imitation from the local neighborhoods [7], [9], [10]. Thus, this type of mechanisms tend to have strength in the particular type of scenarios which they are designed for. On the other hand, the learning-based (in particular, reinforcement learning-based) mechanisms model the local interactions among agents to be coordination games, and assume agents to independently learn from trial-and-error [11]–[13]. Therefore, the learning-based mechanisms in nature should be applicable to a much wider domain scenario, particularly in which observation and imitation are not feasible.

Researches have studied convention emergence under various interaction models. The interaction model can be categorized as the asynchronous interaction model and the synchronous interaction model based on the number of agents participated in interactions during each iteration, the interactions happened between only one pair or group agents during each iteration under the asynchronous interaction model [1], [7], [10], where happened among all agents under the synchronous interaction model [10]–[12]. The interaction model can also be categorized into the 2-player interaction model [6], [12], [14] and the multi-player interaction model [10], [11], [15], [16] based on the number of agents involved in an interaction.

In MASs, agents are not always able to observe or imitate

This work was supported by Project of Key Liberal Science Research Base in Chongqing (18SKB047).

the actions of others due to constraints and costs on the interaction channel. It is therefore important to take into considerations of such constraints when modeling interactions. Mihaylov, Tuyls and Nowé proposed Win-Stay Lose-probabilistic-Shift(WSLpS) and applied it to various interaction models and scenarios [10]. However, the WSLpS plays a poor performance when agents can not observe the actions of their neighbors. Moreover, when the number of available actions become large, it always fails to establish conventions for agents blindly search for new actions. Learning-based methods do not rely on the agents' ability of observation, agents interact with others and learn the utility of each action based on the feedback reward, however, agents may stay at one particular action for a long time due to adhering to their learning experiences when convention seeds conflict, such phenomenon also called sub-conventions which has been studied by some researches [17], [18]. In this paper, we integrate the advantages of 'Win-Stay' and reinforcement learning to tackle the problems discussed above. More specifically, we harness the power of reinforcement learning to solve the problem of no observation and shorten the action transformation time by resetting the agents' learning experiences and combining the idea of 'Win-Stay'.

The remainder of the paper is organized as follows. We review the related work in the next section. Section 3 explains some basic concepts about convention emergence problem and introduces the scenario we studied. The proposed method will be interpreted in Section 4. In Section 5, we present the experimental results. Finally, we make our conclusion in Section 6.

II. Related Work

Conventions, as an effective mechanism to regulate agents' behaviors, have attached a wide range of attention in MASs. Methods for convention emergence through agents' local interactions have been studied for many years. One line of method for convention emergence is the spreading-based method, Shoham and Tennenholtz introduced the convention emergence problem into MASs [9], they proposed four basic types of strategy update rules and showed that the external majority (EM) strategy performs best among all these strategies, later, they proposed Highest Cumulative Reward (HCR) strategy and modeled the agents' interactions as coordination games. Delgado proposed generalized simple majority (GSM) rule, agents adopt an action with a probability based on the action distribution of their neighbors [7]. More recently, Mihaylov, Tuyls and Nowé proposed WSLpS for various interaction models, agents maintain their current strategies when they win, and with a probability shift their strategies when they lose [10]. Another line for convention emergence is the learning-based method, Sen and Airiau proposed social learning framework and equipped agents with reinforcement learning algorithm, they showed that conventions can successfully emergence through agents' synchronous pairwise learning [12], later, this work is extended by lots of researches by taking consideration of network topologies [6], [17], [19] or other interaction models [11], [13], [15].

There are some researches that combined the learning-based method with other mechanisms for convention emergence. Villatoro, Sabater-Mir and Sen introduced two kinds of social instruments: rewiring and observation, to tackle the effect of subconvention emergence, and hence facilitate the emergence of global conventions [18]. Yu et al. proposed a hierarchical learning framework, subordinate agents report their interaction information to their corresponding supervisors, supervisors gather these information and interchange it with other supervisors to generate guide policies [20]. More recently, Wang et al. deployed teacher-student mechanism on top of the learning method to accelerate the emergence of language conventions [8]. Our research goal in this paper is to combine the idea of 'Win-Stay' and reinforcement learning to facilitate the emergence of conventions under the multi-player synchronous interaction model.

III. Preliminaries

In this section, we explicate some basic concepts and formalize the scenario we studied in this paper. The notation we used in this paper has been summarized in Table I

TABLE I
Summary of notation

Notation	Description
A	available action set
$\|A\|$	number of available actions
α	learning rate
ε	exploration rate
Q_i	the Q table of agent i
$N(i)$	agent set of i's neighbors
R	the sum of rewards in an iteration
γ	the threshold value of win
β	the ratio of successful interactions in an iteration

Definition 1 (Convention)**.** *A social law that restricts the agents behavior to one particular strategy is called a* (social) convention. *[1]*

A typical example used by existing literatures is the scenario of 'rules of the road', when we drive a car on a road, there are need a rule specifies which side to drive to avoid collision with other cars, quite evidently, no matter which side is specified, the traffic order can be ensured, therefore, the rule of 'drive on the left side' or 'drive on the right side' can both be regarded as a convention.

Convention Emergence Framework: The scenario we studied in this paper given rise to conventions under the *multi-player synchronous interaction model* in *networked MASs*. The interaction process is given by Algorithm 1. The interaction between agents is modeled as a 2-player-m-action pure coordination game, agents get a $+1$ payoff when they play a same action (*successful interaction*), otherwise, they are punished with a -1 payoff. There are m Nash equilibria in a 2-player-m-action pure coordination game. In this paper, particularly, we focus on the scenario that agents can not inform the actions played by their neighbors, whereas the only information agents can perceive is whether an interaction is successful. A typical scenario under this assumption is [10]: In a sensor network, nodes coordinate on a same channel for messages dissemination; By using protocols such as Time Division Multiple Access, all nodes interact and update their actions at every time steps, nodes can determine whether an interaction is successful or not according to the messages received.

Algorithm 1 Multi-player synchronous interaction model

```
INPUT Maximum iterations $T$
  for each agent $i$ do
      initializes a random action $a_i$
  end for
  for each iteration $t \in \{1, 2 \cdots T\}$ do
      for each agent $i$ do
          for each agent $j \in N(i)$ do
              agent $i$ plays coordination game with $j$ and receives reward $r_j$
          end for
      end for
      for each agent $i$ do
          updates its action $a_i$ using a certain method
      end for
  end for
```

Reinforcement Learning and Rewards: Reinforcement Learning (RL) has shown to be suitable for convention emergence. Q-learning, a standard RL algorithm, is well adopted [21]. The decision update rule of Q-learning used in a convention emergence-context (under the pure coordination game) is shown in (1). Agents learn the utility of each action based on the reward from each iteration. Under the multi-player synchronous interaction model, agents interact with all their neighbors in an iteration. Therefore, the reward is defined as the sum of payoffs from each bilateral interaction.

$$Q^t(a) = (1 - \alpha) \times Q^{t-1}(a) + \alpha \times R \tag{1}$$

IV. Win-Stay-Lose-Learn

We present the action update method Win-Stay-Lose-Learn. This method integrates the idea of 'Win-Stay' and RL. A sketch of WSLL is given by Algorithm 2. There are two parts in WSLL, the first part is that agents maintain their current strategies and reset their learning experiences when they win; the second is that agents learn the utilities of their current actions when they lose.

Intuitively, an agent should maintain its current strategy for coordination when its action is consistent with most of its neighbors. As the scenario formalized above, agents can not observe the actions of their neighbors. Instead, the agents may only know how popular their current actions are among their neighbors based on the number of successful interactions in each iteration. In this case, the meaning of 'win' in WSLL is the ratio of successful interactions in one iteration greater than a threshold value γ. We use β to denote the ratio of successful interactions and we will study how γ influences the emergence of conventions in the experimental section.

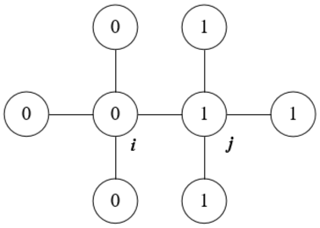


Fig. 1. Subconvention

Resetting the learning experience is another main operation that agents should perform when they 'win'. The main advantage of this setting is that agents can quickly shift their strategies when they see conflict. Without resetting the learning experience, the agents may take a long time to fix their learning experiences by receiving negative rewards toward their current actions or sometimes even fail to build conventions. Agents also take a small probability to select a new action (also called *exploration*) when $\gamma \leq \beta < 1$ to prevent the emergence of subconventions. A typical case is shown in Fig. 1, if the threshold value γ is less than 0.75, all the eight agents are in the state of 'win'. However, the global convention fails to emergence as that state is stable if agents do not explore. In this case, agents choose a new action with a small probability when $\gamma \leq \beta < 1$ will help to escape from that state. Take the scenario of Fig. 1 as an example and assume $\alpha = 0.5$, if the agent i choose a new action 1 in a certain iteration, the state of i and $N(i) - \{j\}$ all changed from 'win' to 'lose' in that iteration, their corresponding Q tables are updated as $Q_i(1) = -1$ and $Q_{N(i)-\{j\}}(0) = -0.5$. Therefore, i and $N(i) - \{j\}$ will select action 0 and 1 respectively in the next iteration. These agents will still 'lose' in the second iteration. However, the Q-tables of these agents will update as $Q_i(1) = -1, Q_i(0) = -2$ and $Q_{N(i)-\{j\}}(0) = Q_{N(i)-\{j\}}(1) = -0.5$, the agents of $N(i) - \{j\}$ will choose a random action in the third iteration. The only stable state of the system is that all the agents play a same action. Therefore, it is necessary for agents to select a new action with a small probability even their actions are consistent with most of their neighbors. This case also shows the advantage of resetting the learning experience when agents are in the state of 'win', in the above simulation process, we assume the initialized Q values of all agents toward each action are equal to 0, however, the Q values of the agents' current actions are always largest among all available actions if not reset the learning experience, therefore, it will take a long time for agents to fix their learning experiences. It is easy to know with the number of available actions increased, the probability of forming the phenomenon like Fig. 1 is decreased, so we set the explore probability decreased along with the number of available actions increased.

Algorithm 2 WSLL

```
INPUT action $a_i$ of agent $i$ of current iteration, vector $\vec{R_i}$
    containing the rewards from each neighbors.
OUTPUT the new action $a_i$ of agent $i$ for next iteration.
  $R = \sum_{j \in N(i)} r_j$
  if $\beta \geq \gamma$ then
      $Reset\ Q_i$
      $rnd \leftarrow generate\ Random\ Number \in (0, 1)$
      if $rnd < \varepsilon / |A|$ and $\beta < 1$ then
          $a_i \leftarrow choose\ Random\ a \in A - \{a_i\}$
      else
          $a_i \leftarrow a_i$
      end if
  else
      $Q_i^t(a_i) = (1 - \alpha) \times Q_i^{t-1}(a_i) + \alpha \times R$
      $a_i \leftarrow argmax Q_i^t(a_i)$
  end if
```

V. Experimental Study

We firstly investigate the influence of γ on convention emergence, then we compare the performance of our proposed method with other baseline approaches under various settings. Unless otherwise specified, we take the following settings by default: 90% as the criterion of convention emergence, 100 agents as population size, Scale-free network with 3 as power law exponent, the number of available actions is set as 5, maximum iterations is set as 5000 in each simulation, all results are averaged over 500 simulation runs, the learning rate α is set as 0.3, the exploration rate ε is set as 0.1, the baseline methods and their settings are as below:

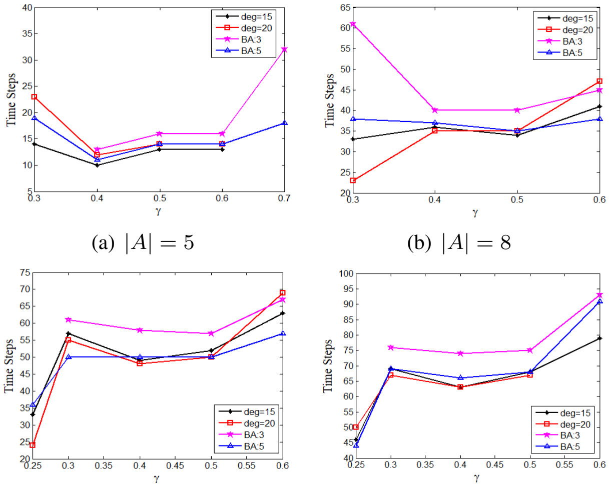
Win-Stay Lose-probabilistic-Shift (WSLpS): The shift probability is set as 0.8 which is optimal for multi-player synchronous interaction model as suggested by the authors.

Q-learning (Q): The learning rate is set as 0.3 and the exploration rate is set as 0.1.

Collective learning(CL): Q-learning as the learning strategy, the learning rate and exploration rate is same as Q-learning, global exploration as the exploration mode, majority voting as the ensemble method. We mention that the authors assumed the common observation in their first work [11]; however, they eliminate this assumption in their follow-up research [15].

A. The influence of γ on convention emergence

In this subsection, we investigate the influence of γ on convention emergence, the questions we main concerned are what is the range of γ helps to establish convention and what factors will influence it. For we take the ratio of successful interactions as the criterion of 'win', one major factor that will directly affect the expectation of successful interactions is the number of available actions, the expectation is $\frac{1}{|A|}$ in the totally stochastic condition, there should be less agents play the same action with the focal agent when the number of available actions increased, therefore, we investigate the influence of γ on convention emergence by varying the $|A|$ as $5, 8, 10, 12$. Fig. 2 represents the experimental results, the $x-$axis is value of γ and the $y-$axis indicates the corresponding convergence time steps, the outlier value are not displayed for legibility. Each experiment is conducted under two regular networks which have 15 and 20 fixed degree separately, this setting makes each agent have a same circumstance, we also conduct each experiment under two scale-free network which the degree of nodes follow a pow-law distribution, this setting helps to vary the circumstance of each agent, the power law exponent we choose are 3 and 5. Firstly, we can observe that our approach have a robust performance under various circumstances, convention can successfully emergence under a wide range value of γ in all experiments, more precisely, the performance of WSLL under each $|A|$ is very robust when $0.3 \leq \gamma \leq 0.6$. Secondly, the valid value of γ left shift along with $|A|$ increased, we hypothesise that this phenomenon may due to the expectation of successful interactions in random case decreased along with the $|A|$ increased, therefore, a small ratio can be seen as 'win' compared to a small expectation of successful interactions. Thirdly, the optimal value of γ for convention emergence is always $2 \sim 3$ times $\frac{1}{|A|}$, we conjecture it may due to the agents can explicitly distinguish the majority and minority of their current actions and the processes of 'Stay' and 'Learn' are balanced under such value. In the next subsections, we compare WSLL with other baseline methods under various settings and choose 0.5 as the threshold value of win.



(a) $|A| = 5$

(b) $|A| = 8$

(c) $|A| = 10$

(d) $|A| = 12$

Fig. 2. The influence of γ on convention emergence under different available actions

B. Convention emergence under different available actions

We test the effectiveness of WSLL under the settings of different available actions by varing the $|A|$ as $2, 5, 8, 10$ and compare it with other three methods. Fig. 3 presents the results of different approaches establish conventions, each bar represents the averaged convergence time of each approach and the black mark of each bar indicates the corresponding standard error, besides, we randomly sample 50 times simulation results of each method under the setting of $|A| = 5$ to vividly show

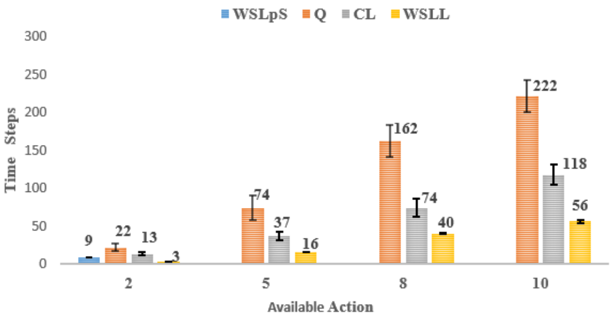
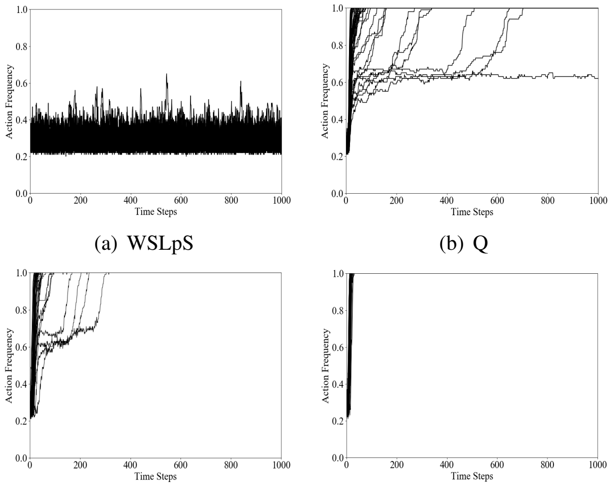


Fig. 3. Speed of convention emergence with respect to the available actions



(a) WSLpS (b) Q

(c) CL (d) WSLL

Fig. 4. The dynamics of convention emergence with 5 available actions

the dynamics of convention emergence, which is shown in Fig. 4. Firstly, we can observe that not only WSLL establish convention faster than other three methods, but also WSLL has the most stable performance. Secondly, the WSLpS performs better than other two learning-based approaches under the setting of 2 available actions, the rationale behind the effectiveness of WSLpS is that the agents stay with their current strategies with a big probability when their current actions are consistent with most of their neighbors, and agents can shift to another 'right' action once they are consistent with little of their neighbors because the actions played by agents are mutually exclusive when there are only two available actions, besides, our method even outperforms WSLpS, because the strategy of 'shift' in WSLpS is with a probability, however, our method is absolute, and with the help of learning, agents can select the 'right' actions as well. Thirdly, when $|A| >= 5$, the WSLpS fails to establish conventions, we can observe the dynamics of WSLpS in Fig. 4 (a), agents can not choose a appropriate action due to blindly searching for new actions when they can not observe the actions played by their neighbors. The other two learning-based baseline methods are more robust to the circumstances of no observation because of the ability of recording the utility of each action, it should note that the main difference between WSLL and them is that agents reset their learning experiences when their current actions are consistent with most of their neighbors, as discussed in section IV,the advantage of resetting learning experiences is to help agents quickly adjust their strategies when convention seeds conflict, the agents with learning-based methods may adhere to their current strategies for a long time, as we can see in Fig. 4 (b) and (c), there are some lines become horizontal during some time steps which show the processes of agents coordinate their strategies when convention seeds conflict, it should mention that there are 1% simulations fail to establish convention in Q-learning and Collective learning. For the poor performance of WSLpS under the setting of large action space and the default $|A|$ is set as 5 in our experimental settings, therefore, we will not conduct experiments with WSLpS in the next subsections.

C. Convention emergence under different network topologies

One main factor that will influence the emergence of conventions is the topologies of MASs, the reality MASs often form as complex networks, for instance, the web sites usually form as Scale-free networks. To this end, we compare WSLL with other baseline methods under three types of complex network to verify the effectiveness of it, the network types we choose are a Random network with 10 averaged degree, a Small-world network with 12 averaged neighbors and 0.1 as rewiring probability, a Scale-free network with 5 as power law exponent.

As the results shown in Fig. 5, WSLL establishes conventions faster than other methods under all the three types of network, besides, we can observe that the convergence time in Random and Scale-free network are almost same, the similar phenomenon also can be seen in Fig. 2, all lines in each subfigure always close to each other, these phenomenons demonstrate that WSLL has a strong stable performance under various network topologies. Another significant phenomenon is that Q-learning and Collective learning fail to establish conventions under the Small-world network which further verifies the necessity of resetting the learning experience for agents coordinate their strategies when convention seeds conflict, agents tend to cluster in Small-world networks and hence are more likely to form different convention seeds in each cluster, agents may take a long time or even fail to adjust their strategies when their action seeds conflict with the agents in other clusters, in this case, agents with WSLL can quickly shift their strategies.

D. The influence of population size on Convention emergence

Another issue we concerned is that how the MASs scale influences the performance of WSLL. To test the adaptability of WSLL on MASs scale, we vary the MASs size with 200,500,1000,2000 agents. As the results shown in Fig. 6, WSLL takes fewer time to establish conventions than other two baseline approaches under all settings, besides, WSLL has the smallest growth rate among all approaches, as the population size grows from 200 to 2000, the averaged convergence time

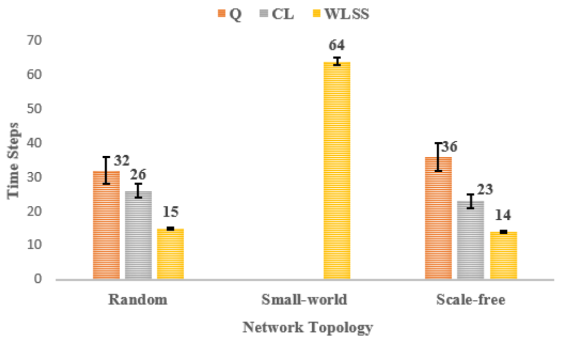


Fig. 5. Speed of convention emergence with respect to network topology

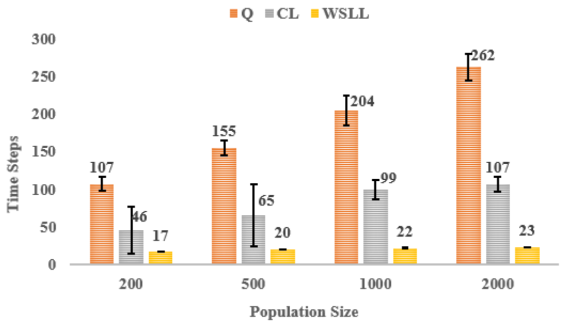


Fig. 6. Speed of convention emergence with respect to population size.

only increases 35%, more importantly, the growth rate reduced along with the population size increased, when the population size grows from 1000 to 2000, the growth rate only increases 0.05%, these results indicate that WSLL can be applied to large scale MASs.

VI. Conclusions and Future Work

In this paper, we study the problem of convention emergence under multi-player synchronous interaction model in networked MASs, more especially, we focus on the scenario that agents can not observe the actions played by their neighbors, the only information agents can perceive is whether an interaction is success or not. A novel action update strategy, Win-Stay-Lose-Learn(WSLL), is proposed. Experimental results show that our proposed method outperforms other baseline methods in terms of convergence speed, besides, the performance of WSLL is very robust under various circumstances. In the future, we will try to apply our method to other interaction models to capture more scenarios.

References

[1] Y. Shoham and M. Tennenholtz, "On the emergence of social conventions: modeling, analysis, and simulations," *Artificial Intelligence*, vol. 94, no. 1-2, pp. 139–166, 1997.

[2] C. D. Hollander and A. S. Wu, "The current state of normative agent-based systems," *Journal of Artificial Societies and Social Simulation*, vol. 14, no. 2, p. 6, 2011.

[3] J. Morales, M. López-Sánchez, J. A. Rodríguez-Aguilar, M. Wooldridge, and W. Vasconcelos, "Synthesising liberal normative systems," in *Proc. of AAMAS*, 2015, pp. 433–441.

[4] G. Boella and L. van der Torre, "An architecture of a normative system: counts-as conditionals, obligations and permissions," in *Proc. of AAMAS*, 2006, pp. 229–231.

[5] N. Salazar, J. A. Rodriguez-Aguilar, and J. L. Arcos, "Robust coordination in large convention spaces," *Ai Communications*, vol. 23, no. 4, pp. 357–372, 2010.

[6] S. Hu and H.-f. Leung, "Achieving coordination in multi-agent systems by stable local conventions under community networks." in *IJCAI*, 2017, pp. 4731–4737.

[7] J. Delgado, "Emergence of social conventions in complex networks," *Artificial intelligence*, vol. 141, no. 1-2, pp. 171–185, 2002.

[8] Y. Wang, W. Lu, J. Hao, J. Wei, and H.-F. Leung, "Efficient convention emergence through decoupled reinforcement social learning with teacher-student mechanism," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 795–803.

[9] Y. Shoham and M. Tennenholtz, "Emergent conventions in multi-agent systems: Initial experimental results and observations," in *Proc. of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, 1992, pp. 225–231.

[10] M. Mihaylov, K. Tuyls, and A. Nowé, "A decentralized approach for convention emergence in multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 28, no. 5, pp. 749–778, 2014.

[11] C. Yu, M. Zhang, F. Ren, and X. Luo, "Emergence of social norms through collective learning in networked agent societies," in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 475–482.

[12] S. Sen and S. Airiau, "Emergence of norms through social learning." in *IJCAI*, vol. 1507, 2007, p. 1512.

[13] S. Hu, C.-w. Leung, H.-f. Leung, and J. Liu, "To be big picture thinker or detail-oriented?: Utilizing perceived gist information to achieve efficient convention emergence with bilateralism and multilateralism," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 2021–2023.

[14] D. Villatoro, S. Sen, and J. Sabater-Mir, "Exploring the dimensions of convention emergence in multiagent systems," *Advances in Complex Systems*, vol. 14, no. 02, pp. 201–227, 2011.

[15] J. Hao, J. Sun, G. Chen, Z. Wang, C. Yu, and Z. Ming, "Efficient and robust emergence of norms through heuristic collective learning," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 12, no. 4, p. 23, 2018.

[16] Q. Liu, H. Zheng, W. Li, J. Liu, B. Yan, and H. Su, "A model of minority influence in preferential norm formation," in *International Symposium on Knowledge and Systems Sciences*. Springer, 2019.

[17] D. Villatoro, S. Sen, and J. Sabater-Mir, "Topology and memory effect on convention emergence," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 2. IEEE, 2009, pp. 233–240.

[18] D. Villatoro, J. Sabater-Mir, and S. Sen, "Social instruments for robust convention emergence," in *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[19] O. Sen and S. Sen, "Effects of social network topology and options on norm emergence," in *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*. Springer, 2009, pp. 211–222.

[20] C. Yu, H. Lv, F. Ren, H. Bao, and J. Hao, "Hierarchical learning for emergence of social norms in networked multiagent systems," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2015, pp. 630–643.

[21] H. Zhao, H. Su, Y. Chen, J. Liu, H. Zheng, and B. Yan, "A reinforcement learning approach to gaining social capital with partial observation," in *Pacific Rim International Conference on Artificial Intelligence*. Springer, 2019, pp. 113–117.