

Hierarchical Learning for Emergence of Social Norms in Networked Multiagent Systems

Chao Yu¹(✉), Hongtao Lv², Fenghui Ren³, Honglin Bao¹, and Jianye Hao⁴

¹ School of Computer Science and Technology,
Dalian University of Technology, Dalian 116024, China
cy496@dlut.edu.cn, BHL19931025@mail.dlut.edu.cn

² School of Innovation Experiment,
Dalian University of Technology, Dalian 116024, China
lvhongtao@mail.dlut.edu.cn

³ School of Computer Science and Software Engineering,
University of Wollongong, Wollongong 2500, Australia
fren@uow.edu.au

⁴ School of Software, Tianjin University, Tianjin 300072, China
jianye.hao@tju.edu.cn

Abstract. In this paper, a hierarchical learning framework is proposed for emergence of social norms in networked multiagent systems. This framework features a bottom level of agents and several levels of supervisors. Agents in the bottom level interact with each other using reinforcement learning methods, and report their information to their supervisors after each interaction. Supervisors then aggregate the reported information and produce guide policies by exchanging information with other supervisors. The guide policies are then passed down to the subordinate agents in order to adjust their learning behaviors heuristically. Experiments are carried out to explore the efficiency of norm emergence under the proposed framework, and results verify that learning from local interactions integrating hierarchical supervision can be an effective mechanism for emergence of social norms.

Keywords: Norm emergence · Learning · Multiagent systems

1 Introduction

Social norms have been considered to be an effective mechanism to facilitate coordination and cooperation in Multi-Agent Systems (MASs). How social norms can establish themselves automatically as an emerging process is a key problem in the research of social norms. It has been well recognized that learning from individual experience is a robust mechanism to facilitate emergence of social norms [1]. A great deal of previous work has thus studied emergence of social norms either through agent learning from random interactions in a well-mixed population of agents (e.g., [1, 2]), or through agent learning from local interactions with neighbors in networked agent societies (e.g., [3–7]).

Although these studies provide us with some valuable insights into general mechanisms behind emergence of norms achieved through agent learning behaviors, there exist several limitations in these studies inevitably. **First**, rich organizational characteristics such as hierarchical structure of the agent systems are not considered in these studies. For self-regulated MASs, however, these organizational characteristics can have significant impacts on opinion evolution and exchange of information among agents, thereby substantially influence the emergence of norms in the entire system. Thus, it is necessary to reflect these organizational characteristics in the local learning process of agents during the process of norm emergence. **Second**, most current studies investigate norm emergence in a small population and small action space. When the action space and population become large, subnorms may exist in the system and traditional models cannot form norms efficiently. In real-life human societies, however, social norms can always emerge efficiently in complicated organizations, ranging from enterprises to communities and nations. This is because in these organizations, there is some kind of centralized control governing the distributed interactions among agents. Therefore, it is necessary to integrate this kind of centralized control into distributed interactions among agents in order to facilitate the process of norm emergence, especially when the system is becoming more complex.

Based on the above consideration, this paper studies the influence of organizational structure on norm emergence, by proposing a hierarchical learning framework in networked MASs. In the framework, agents are separated into different clusters. Each agent in a cluster interacts with others using reinforcement learning methods, and reports its learning information to an upper layer supervisor in the cluster. After synthesizing all the interaction information in the cluster, the supervisor then generates a guide policy through learning from neighboring supervisors, and passes down this policy to subordinate agents in order to adapt their learning strategies heuristically. The main feature of the proposed framework is that through hierarchical supervision between subordinate agents and supervisors, a compromising solution can be made to elegantly balance distributed interactions and centralized control for norm emergence. Experiments show that learning from local interactions integrating hierarchical supervision can be an effective mechanism for emergence of social norms.

The remainder of the paper is organized as follows. Section 2 introduces social norms and reinforcement learning. Section 3 introduces the proposed learning framework. Section 4 presents experimental studies. Section 5 discusses related work. Finally, Sect. 6 concludes the paper with directions for future research.

2 Social Norms and Reinforcement Learning

In this study, we use the scenario of “rules of the road” [8,9] as a metaphor to study norm emergence. This rule means that when we drive on the road, we would choose driving on the left side (L) or on the right side (R), and we aim to establish a norm that one side (L or R) will be chosen by all agents in the society. This rule can be viewed as a coordination game [9] in Table 1. There are

Table 1. Payoff matrix of the coordination game

	Left (L)	Right (R)
Left (L)	1, 1	-1, -1
Right (R)	-1, -1	1, 1

two pure Nash-equilibria: both agents choose left or both agents choose right, and there is no preference on L or R in the coordination game.

In the real applications of MASs, the number of actions available to the agents is an important factor in the emergence of social norms. Most current studies on norm emergence, however, only deal with a relatively simple norm space in which a norm is chosen from two possible action alternatives. In order to study norm emergence in a larger action space, we extend the coordination game into a more general form by considering N_a actions in the action space. In every time step, agent i and agent j choose action a_i and action a_j , respectively. If $a_i = a_j$, they can get immediate payoff of 1, and -1 otherwise.

Reinforcement learning algorithms [12] have been widely used for agent interactions in previous research of norm emergence. We focus on Q-learning algorithm in this paper, in which an agent makes a decision through the estimation of a set of Q-values. Its one-step updating rule is given by Eq. 1,

$$Q(s, a) = Q(s, a) + \alpha_i [r(s, a) + \lambda \max_{a'} Q(s', a') - Q(s, a)] \quad (1)$$

where $\alpha_i \in (0, 1]$ is a learning rate of agent i , $\lambda \in [0, 1)$ is a discount factor, $r(s, a)$ and $Q(s, a)$ are the immediate and expected reward of choosing action a in state s at time step t , respectively, and $Q(s', a')$ is the expected discounted reward of choosing action a' in state s' at time step $t + 1$.

Q-values of each state-action pair are stored in a table for a discrete state-action space. At each time step, agent i chooses the best-response action with the highest Q-value based on the corresponding Q-value table with a probability of $1 - \epsilon$ (i.e., exploitation), or chooses other actions randomly with a trial-and-error probability of ϵ (i.e., exploration).

3 The Proposed Learning Framework

3.1 Overview of the Hierarchical Learning Framework

Consider the society structure in a country. There is a government to manage citizens, and a senior government to manage lower level governments, and so on. In this scenario, the government is playing a role of supervisor. It supervises the state of subordinate citizens and educates them to act properly. Inspired from the structure of society, we propose a hierarchical network as shown in Fig. 1. The agent network is separated into a series of clusters $C_x, x \in (1, 2, \dots, X)$, where X is the number of clusters, and x denotes a supervisor for each cluster. Supervisors can be one of the subordinate agents or a dedicated agent. They are

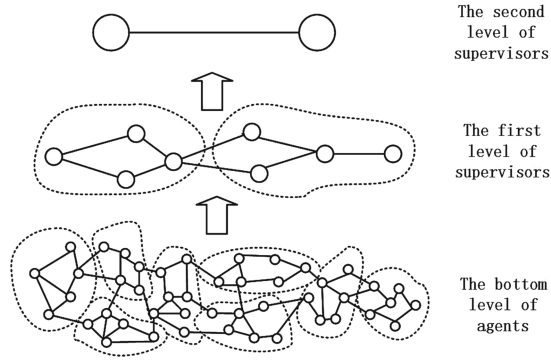


Fig. 1. An example of a hierarchical network with three levels

Algorithm 1. The interaction protocol of the proposed learning framework

```

1 for each step  $t(t = 1, \dots, T)$  do
2   for each agent  $i(i = 1, \dots, n)$  do
3     Chooses an action  $a_i$  based on its Q-values;
4     Interacts with  $j, \forall j \in N(i)$ , and gets a reward  $r_i$ ;
5     Reports  $a_i$  and  $r_i$  to its supervisor;
6   end
7   for each supervisor  $x(x = 1, \dots, m)$  do
8     Combines actions of subordinate agents into  $a_x$ ;
9     Communicates with  $y, \forall y \in N(x)$ , and converts  $a_x$  to  $a_y$  with a
    probability  $p$ ;
10    Passes down  $a_x$  to its subordinate agents;
11  end
12  for each agent  $i(i = 1, \dots, n)$  do
13    Updates  $\alpha_i$  or  $\epsilon_i$  based on the  $a_x$  of its supervisor;
14    Updates Q-values;
15  end
16 end

```

also connected to each other based on the lower network structure. Supervisors can collect learning information of their subordinate agents, and they can also interact with their neighbors. Similarly, a higher level of supervisors can be established based on the network of the first level of supervisors. In this way, we can establish several levels of supervisors until the population of supervisors in the top level is small enough to make a final decision efficiently. In this paper, we use two-level hierarchy networks for simplification.

The interaction protocol of our hierarchical learning framework is given in Algorithm 1. In each time step, each agent i chooses an action a_i with highest Q-value or randomly chooses an action with an exploration probability of ϵ (Line 3). Then, agent i interacts with a neighbor j randomly selected from the set of

neighbors $N(i)$ and gets a payoff of r_i (Line 4). The action a_i and payoff r_i are reported to agent i 's supervisor x (Line 5), and supervisor x combines all actions of its subordinate agents into an action a_x (Line 8). Supervisor x then interacts with one of its neighbors and updates a_x based on the performance difference between the actions of supervision x and its neighbor (Line 9). This process will be presented in detail in the following part. When supervisor x has taken a final decision a_x , it issues a_x to its subordinate agents (Line 10). Based on this information from supervisor x , agent i adjusts its learning rate α_i or the exploration probability ϵ_i (Line 13). Finally, each agent updates its Q-values using new learning rate α_i (Line 14). This process is iterated for T time steps.

The main feature of the hierarchical learning framework is that there is hierarchical supervision governing over the distributed interactions among agents. As supervisors have a wider view than subordinate agents, they can direct their subordinate agents by learning from other supervisors. Through the feedback from supervisors, subordinate agents can understand whether their actions comply with the norms in the society, and thereby adjust their learning behaviors accordingly. By integrating supervisors' central control into agents' local learning processes, norms can be established more effectively and efficiently in the hierarchical learning framework, compared with the traditional social learning framework based on pure distributed interactions.

3.2 The Hierarchical Learning Mechanism

The Report Mechanism and Assemble Method. Each agent reports its action a_i and payoff r_i to its supervisor x . The supervisor aggregates the information into two tables QS_x and RS_x . $QS_x(a)$ means the overall acceptance of the action a in the cluster and $RS_x(a)$ means the overall reward of action a in the cluster. $QS_x(a)$ can be calculated by Eq. 2,

$$QS_x(a) = \sum_{i \in C_x} s_i \quad (2)$$

where s_i is an indicator function given as follows:

$$s_i = \begin{cases} 1 & \text{if } a_i = a, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

$RS_x(a)$ can be calculated by Eq. 4,

$$RS_x(a) = \frac{1}{QS_x(a)} \sum_{i \in C_x, a_i = a} r_i \quad (4)$$

especially, if $QS_x(a) = 0$, $RS_x(a)$ is set to 0.

Each supervisor then gets its aggregated action a_x . We use democratic voting mechanism in this paper, which means that a_x is simply the action most accepted by the cluster (randomly choose an action if there exists a tie). The aggregated action a_x can be given by Eq. 5,

$$a_x = \arg \max QS_x(a) \quad (5)$$

The Interaction Mechanism of Supervisors. After combining actions of the cluster, each supervisor generates an assemble action a_x , which indicates the norm (i.e., action) accepted by the society. In order to generate a better supervision policy, each supervisor then interacts with a randomly chosen neighboring supervisor and learns from the neighbor by comparing the performance of their assemble actions. The motivation of this comparison comes from evolutionary game theory [13], which provides a powerful methodology to model how strategies evolve over time based on their performance. In this theory, strategies compete with each other and evolve while agents learn (mainly through imitation) from others with higher fitness. One of the widely used imitation rules is the proportional imitation, which can be given by Eq. 6 [14],

$$p = \frac{1}{1 + e^{-\beta(u_y - u_x)}} \quad (6)$$

where p is a probability for supervisor x to imitate the neighbor y , $u_x = RS_x(a_x)$ means the fitness (average payoff of the assemble action in this paper) of supervisor x , $u_y = RS_y(a_y)$ means the fitness of y , and β is a parameter to control the density of selection.

In Eq. 6, probability p can be large when u_x is smaller than u_y . Every supervisor x then converts a_x into a_y as the final assemble decision with probability p , or remains a_x with probability $1 - p$. In this way, supervisors imitate better actions taken by their neighbors to improve the behaviors of their clusters.

Strategy Adaption of Subordinate Agents. How to adjust learning strategies of subordinate agents using the guide information from supervisors is the key problem in the hierarchical learning framework. Based on the guide information from supervisors, subordinate agents can realize whether they have complied with the norm in the society and adjust their learning behaviors accordingly. To explicitly evaluate the agents' performance in terms of complying with the norm, we borrow the concept of WoLF (Win-or-Learn-Fast) principle [15], which is a well-known mechanism in multiagent learning, to the hierarchical learning framework. Each subordinate agent i compares a_x with its action a_i . If $a_i = a_x$, we say that agent i is "winning", and "losing" otherwise. Based on the "winning" or "losing" results, two fundamental learning parameters (i.e., learning rate α and exploration probability ϵ) can be adapted as follows.

Algorithm 2 shows the mechanism of adapting α , which is denoted as *hierarchical learning- α* . Notation s is the changing direction of learning rate. If $a_i = a_x$, the action chosen by agent i is approved by its supervisor. Agent i then decreases its learning rate α_i to maintain the "winning" state. However, if $a_i \neq a_x$, agent i increases the learning rate α_i to learn more from experience. Being more adaptive may bring about a better choice for the next round play.

Similarly, Algorithm 3 shows the mechanism of adapting ϵ , which is denoted as *hierarchical learning- ϵ* . When $a_i = a_x$, agent i decreases its exploration probability ϵ_i to maintain the "winning" state and reduces the mistake probability

because of exploration. However, if $a_i \neq a_x$, agent i increases the exploration probability ϵ_i to explore more actions to get rid of the “losing” state.

The third approach is to combine the above two approaches (i.e., adapting α and ϵ at the same time), which is denoted as *hierarchical learning-both*.

Algorithm 2. hierarchical learning- α	Algorithm 3. hierarchical learning- ϵ
Input: a_x, a_i, α_i ; Output: α_i ; 1 for each agent i do 2 if $a_i = a_x$ then 3 $s = -1$; 4 $\Delta\alpha_i = \gamma\alpha_i$; 5 end 6 else 7 $s = 1$; 8 $\Delta\alpha_i = \gamma(1 - \alpha_i)$; 9 end 10 $\alpha_i = \alpha_i + s\Delta\alpha_i$ 11 end	Input: a_x, a_i, ϵ_i ; Output: ϵ_i ; 1 for each agent i do 2 if $a_i = a_x$ then 3 $s = -1$; 4 $\Delta\epsilon_i = \gamma\epsilon_i$; 5 end 6 else 7 $s = 1$; 8 $\Delta\epsilon_i = \gamma(1 - \epsilon_i)$; 9 end 10 $\epsilon_i = \epsilon_i + s\Delta\epsilon_i$ 11 end

4 Experiment and Results

In this section, experiments are carried out to demonstrate the performance of our framework. We compare the hierarchical learning method with the conventional social learning method in [10], and explore the influence of some parameters on norm emergence under the proposed framework.

4.1 Experimental Setting

In this paper, we test the hierarchical learning framework on a $10 * 10$ grid network. This framework consists of two levels: subordinate agents and supervisors. The network is separated into some $4 * 4$ clusters unless otherwise specified. In Q-learning algorithm, we consider each state as the same, which means that there is no state transition. Each agent can choose from 4 actions as default. Parameters α and ϵ of each agent are initially set as 0.1 and 0.01, respectively. Moreover, parameter β to control the tensity of selection in Eq. 6 and parameter γ in Algorithms 2 and 3 are both 0.1. Unless specified otherwise, the final results are averaged over 100 Monte-Carlo runs.

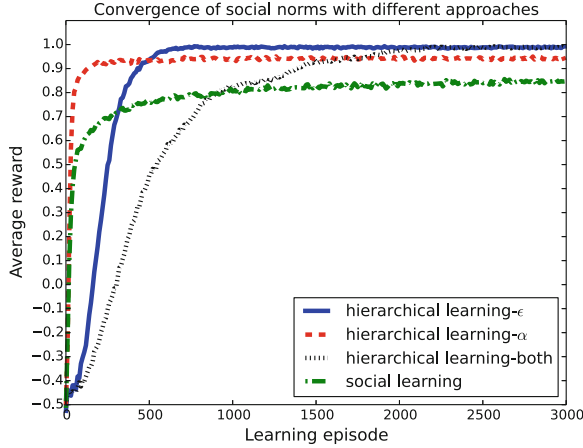


Fig. 2. The comparison of different learning approaches.

4.2 Results and Analysis

Convergence of Social Norms. Previous work has shown that norms can always emerge if agents interact with a randomly chosen neighbor. We compared the performance between the approach based on traditional social learning framework [10] and the three approaches based on our hierarchical learning framework. Figure 2 gives the learning dynamics of the four different learning approaches. From the results, we can see that, with both cases of social learning and hierarchical learning approaches, the average reward of the system increases and then keeps stable, which means the social norms emerge with all these four approaches. The convergence level and efficiency, however, are distinguishable among these approaches. The hierarchical learning approaches (especially hierarchical learning- ϵ and hierarchical learning-both) are slower at first but finally reach a higher convergence level than the social learning approach. Moreover, hierarchical learning- α emerges norms much faster than the other two hierarchical learning approaches. These results indicate that with the guide information of supervisors, agents can know more about the society, and the hierarchical mechanism is an effective way to remove subnorms in MASs to reach a higher level of norm emergence. Hierarchical learning- α is more efficient than the other two approaches because when agents adapt learning strategy with hierarchical learning- ϵ or hierarchical learning-both, the agents have to spend more time on exploring actions at first. But finally, when ϵ reduces to nearly 0, the agents can choose a correct action more firmly than other approaches.

Table 2 gives the detailed performance of the different approaches in 1000 runs. In order to better demonstrate the different performance of these approaches, we also include the results when 100% agents have chosen the same action as the norm. Achieving 100% norm emergence is an extremely challenging issue due to the widely recognized existence of subnorms. To calculate the

Table 2. Performance of different learning approaches (averaged over 1000 runs)

	Reward	90 % convergence		100 % convergence	
		Success rate (%)	Speed (steps)	Success rate (%)	Speed (steps)
Hierarchical learning- α	0.941	82.3	1838	81.8	1926
Hierarchical learning- ϵ	0.988	100	329	100	425
Hierarchical learning-both	0.989	99.9	590	99.9	710
Social learning	0.848	47.4	5660	45.4	5898

average steps needed for norm emergence, we take 10000 episodes as the upper limit in this run when a norm fails to emerge in 10000 episodes. The average reward, success rate (i.e., how many runs a norm can emerge successfully in the 1000 runs) and the speed (i.e., time steps needed to emerge a norm) are compared. We can see that hierarchical learning approaches are obviously more efficient and effective than traditional social learning in all these three criteria. Although hierarchical learning- α can emerge a norm more quickly than the other two approaches (refer to Fig. 2), the average steps to emerge a norm in hierarchical learning- α are larger because of a small probability of failing to emerge a norm (i.e., 10000 steps are used for calculation when a norm fails to emerge).

Dynamics of Learning Rate α and Exploration Probability ϵ . We studied the dynamics of the exploration probability ϵ and the learning rate α with different action sizes in Figs. 3 and 4 (dynamics of α and ϵ with hierarchical learning-both overlap with each other because of the identical update method). We can see that in both cases, exploration probability ϵ and learning rate α rise at first but then decline to 0 afterwards. When there are only 2 actions, both α and ϵ rise to a small extent, and then decline to nearly 0 quickly. When the action number increases to 4, it still takes a short period of time to reduce α to 0 using hierarchical learning- α . However, it takes longer time for the values to decline to 0 using hierarchical learning- ϵ and hierarchical learning-both approaches.

The dynamics of α and ϵ indicate that the agents do not know what they ought to do and therefore choose actions randomly at first. As the norm is emerging, the agents can realize which action should be the norm and thus reduce

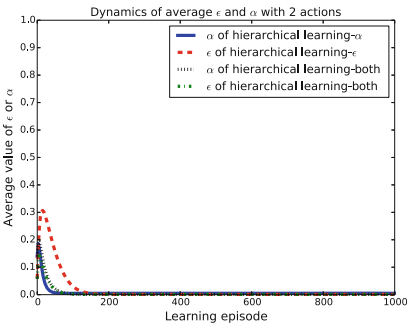


Fig. 3. Dynamics of ϵ and α (2 actions).

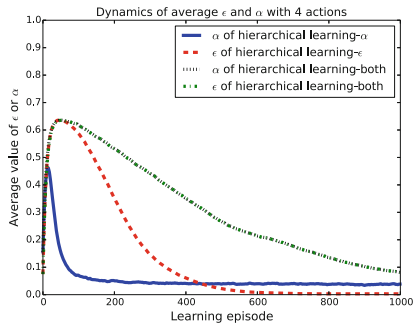


Fig. 4. Dynamics of ϵ and α (4 actions).

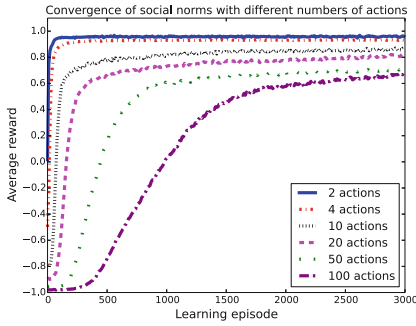


Fig. 5. Influence of number of actions with hierarchical learning- α .

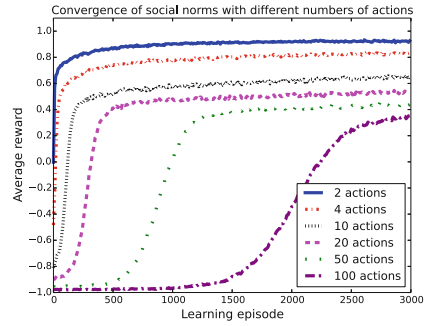


Fig. 6. Influence of number of actions with traditional social learning.

the exploration probability ϵ and the learning rate α accordingly. Hierarchical learning- α is more adaptive in the scenario of a larger action space. This is because agents can find the correct action easily when ϵ is increased (agents are more likely to choose the other action with lower Q-value which is correct) in a two-action scenario. However, in a multi-action scenario, the probability to find the correct action is reduced with the increasing of exploration probability ϵ , and failing to find a correct action can lead to a further increase of ϵ . As a result, agents may be trapped in a vicious circle. But if we choose the mechanism of adapting learning rate α , the agents can learn more experience from unsuccessful interactions, which accordingly speeds up the learning process.

Influence of Number of Actions. The emergence of norms with a large action space is a difficult problem in the field of norm emergence. Figures 5 and 6 show the comparison of hierarchical learning approach (i.e., hierarchical learning- α) and conventional social learning approach. We can see that the hierarchical learning approach performs better than the social learning approach in terms of a faster convergence speed and a higher convergence level. This distinction is more apparent when the action space becomes larger. This is because the agents can receive support from the wide view of their supervisors and get access to the state of the whole society. With this information they can understand whether they have complied with the norm in the society and decide to learn/explore more or to keep their state. It is the process of imitating others with higher rewards that reduces the time to explore unnecessary actions. As a result, the norm can emerge efficiently and subnorms can be removed to a large extent.

Influence of Population Size and Cluster Size. The influence of population size on norm emergence is presented in Fig. 7. We can see that efficiency is nearly unaffected by different population sizes, but effectiveness (i.e., the level of emergence) is obviously reduced in a larger population. This is because in the hierarchical learning framework, with the same size of clusters, each supervisor

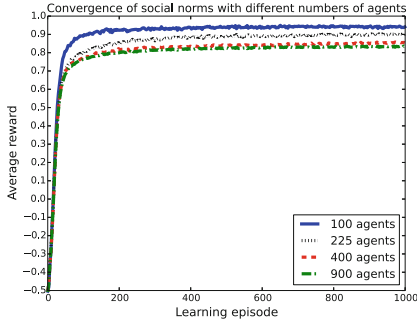


Fig. 7. Influence of population size when separated into 4×4 clusters.

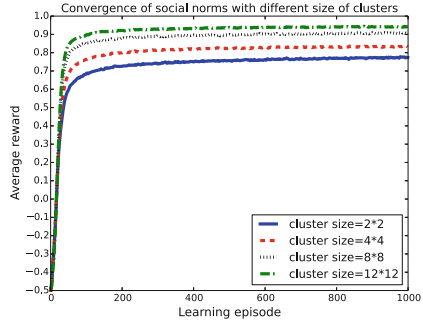


Fig. 8. Influence of cluster size in a 30×30 grid network.

can have a wider view when the population size gets smaller. As a result, each supervisor can have a more powerful control force over the cluster comparatively, and the system tends to be more close to centralized control, which leads to higher efficiency. Similarly, Fig. 8 shows how the cluster size influences the level of emergence with the same size of population. As we can see, a larger cluster size can result in a higher level of emergence. This is because when the cluster size is larger, each supervisor can have a more powerful control force over the cluster comparatively, and therefore the norm can emerge more efficiently.

5 Related Work

Many studies have been done to investigate the mechanisms of norm emergence in the literature. Shoham and Tennenholtz [8] proposed a rule based on the highest cumulative reward to study the emergence of rule. Sen *et al.* [10] presented an emergence model, in which individuals learn from another agents from the group randomly. Then Sen *et al.* [2] extended this model by assessing the influence of heterogeneous agent systems and space-constrained interactions on norm emergence. Savarimuthu [1, 16] recapped the research on the multiagent-based emergence mechanisms, and presented the role of three proactive learning methods in accelerating norm emergence. Airiau and Sen [11, 17] extended group-based random learning model [10] to complicated networks to study the influence of topologies on norm emergence. Villatoro *et al.* [18] proposed a reward learning mechanism based on interaction history. Later, they established two rules (i.e., re-building links with neighborhood and observing neighbors' behaviors) to overcome the subnorm problems [7, 19]. Mihaylo [20] came up with a learning mechanism of Win-Stay Lose-Probabilistic-Shift to make 100% emergence of norms in complicated networks. Mahmoud *et al.* [21] further extended Axelrod's seminal model by considering topological structures among agents. Unlike these studies, which focused on norm emergence either in a fully mixed agent population or in a network agent society, our work captures the organizational

characteristic of hierarchical supervision among learning agents in networked MASs. This feature sets our work apart from all these existing studies.

From another perspective, the problem that has been addressed in this paper is to employ heuristics to guide the policy search to speed up the multi-agent learning process. There is plenty of work in the research of multi-agent learning to properly address this problem. For example, Zhang *et al.* [22] defined a multi-level organizational structure for automated supervision and a communication protocol for exchanging information between lower-level agents and higher-level supervising agents, and Bianchi *et al.* [23] raised Heuristically Accelerated Minimax-Q (HAMMQ) and incorporated heuristics into the Minimax-Q algorithm to speed up convergence rate. Our work is different from the above studies in terms of targeting a different problem (i.e., norm emergence in networked MASs). Nevertheless, the main principle embodied in the proposed framework can shed some light on understanding the learning dynamics in MAL during an efficient convergence to a predefined (i.e., Nash) equilibrium.

6 Conclusions and Future Work

Norm emergence is a crucial issue in understanding and controlling complex systems. Assuming a fully centralized controller to govern the process of norm emergence is not only infeasible for large systems, but also expensive in terms of manageable cost. Therefore, it is more applicable for a norm to emerge on its own through agents' local interactions. This kind of pure distributed way of norm emergence, however, has another limitation in terms of low efficiency especially when the system becomes complex. The hierarchical learning framework proposed in this paper makes a balance between centralized control and distributed interactions during the emergence of social norms by integrating hierarchical supervision into distributed learning of agents. Experiments have indicated that this compromising solution is indeed robust and efficient for evolving stable norms in networked systems, especially when the norm space is large.

Much work remains to be done in the future. For example, the hierarchical learning framework can be conducted in some complex networks, like the small-world networks and the scale-free networks, and the influence of network structure on norm emergence can be explored in depth. Moreover, in this work, we use two-level network for experiments, and it is promising that multilevel networks can lead to a more efficient emergence of social norms.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grant 61502072, Fundamental Research Funds for the Central Universities of China under Grant DUT14RC(3)064, and Post-Doctoral Science Foundation of China under Grants 2014M561229 and 2015T80251.

References

1. Savarimuthu, B.T.R., Arulanandam, R., Purvis, M.: Aspects of active norm learning and the effect of lying on norm emergence in agent societies. In: Kinny, D., Hsu, J.Y., Governatori, G., Ghose, A.K. (eds.) PRIMA 2011. LNCS, vol. 7047, pp. 36–50. Springer, Heidelberg (2011)
2. Mukherjee, P., Sen, S., Airiau, S.: Norm emergence under constrained interactions in diverse societies. In: Proceedings of 7th AAMAS, pp. 779–786 (2008)
3. Hao, J., Sun, J., Huang, D., Cai, Y., Yu, C.: Heuristic collective learning for efficient and robust emergence of social norms. In: Proceedings of 14th AAMAS, pp. 1647–1648 (2015)
4. Yu, C., Zhang, M., Ren, F.: Collective learning for the emergence of social norms in networked multiagent systems. *IEEE Trans. Cybernet.* **44**(12), 2342–2355 (2014)
5. Yu, C., Zhang, M., Ren, F., Luo, X.: Emergence of social norms through collective learning in networked agent societies. In: Proceedings of the 12th AAMAS, pp. 475–482 (2013)
6. Yu, C., Zhang, M., Ren, F., Hao, J.: Emergence of social norms through collective learning in networked agent societies. In: Proceedings of MFSC@AAMAS 2014 (2014)
7. Villatoro, D., Sabater-Mir, J., Sen, S.: Social instruments for robust convention emergence. In: Proceedings of 22nd IJCAI, pp. 420–425 (2011)
8. Shoham, Y., Tennenholtz, M.: On the emergence of social conventions: modeling, analysis, and simulations. *Artif. Intell.* **94**(1), 139–166 (1997)
9. Young, H.P.: The economics of convention. *J. Econ. Perspect.* **10**(2), 105–122 (1996)
10. Sen, S., Airiau, S.: Emergence of norms through social learning. In: Proceedings of the 20th IJCAI, pp. 1507–1512 (2007)
11. Sen, O., Sen, S.: Effects of social network topology and options on norm emergence. In: Padget, J., Artikis, A., Vasconcelos, W., Stathis, K., da Silva, V.T., Matson, E., Polleres, A. (eds.) COIN@AAMAS 2009. LNCS, vol. 6069, pp. 211–222. Springer, Heidelberg (2010)
12. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An introduction. MIT Press, Cambridge (1998)
13. Weibull, J.W.: Evolutionary Game Theory. MIT Press, Cambridge (1997)
14. Pacheco, J.M., Traulsen, A., Nowak, M.A.: Coevolution of strategy and structure in complex networks with dynamical linking. *Phys. Rev. Lett.* **97**(25), 1–9 (2006)
15. Bowling, M., Veloso, M.: Multiagent learning using a variable learning rate. *Artif. Intell.* **136**(2), 215–250 (2002)
16. Savarimuthu, B.: Norm Learning in Multi-agent Societies. University of Otago, Dunedin (2011)
17. Airiau, S., Sen, S., Villatoro, D.: Emergence of conventions through social learning. *Auton. Agent. Multi-Agent Syst.* **28**(5), 779–804 (2014)
18. Villatoro, D., Sen, S., Sabater-Mir, J.: Topology and memory effect on convention emergence. In: Proceedings of the 2009 WI and IAT, pp. 233–240 (2009)
19. Villatoro, D., Sabater-Mir, J., Sen, S.: Robust convention emergence in social networks through self-reinforcing structures dissolution. *ACM Trans. Auton. Adapt. Syst.* **8**(1), 2–21 (2013)
20. Mihaylov, M., Tuyls, K., Now, A.: A decentralized approach for convention emergence in multi-agent systems. *Auton. Agent. Multi-Agent Syst.* **15**(2), 1–30 (2013)

21. Mahmoud, S., Griffiths, N., Keppens, J.: Norm emergence: overcoming hub effects in scale free networks. In: Proceedings of COIN, pp. 136–150 (2012)
22. Zhang, C., Sherief, A., Lesser, V.: Integrating organizational control into multi-agent learning. In: Proceedings of the 7th AAMAS, pp. 757–764 (2009)
23. Bianchi, R.A.C., Ribeiro, C.H.C., Costa, A.H.R.: Heuristic selection of actions in multiagent reinforcement learning. In: Proceedings of 2007 IJCAI, pp. 690–695 (2007)