Achieving Social Optimality with Influencer Agents

Jianye Hao and Ho-fung Leung

Department of Computer Science and Engineering
The Chinese University of Hong Kong
jyhao,lhf@cse.cuhk.edu.hk

Abstract. In many multi-agent systems (MAS), it is desirable that the agents can coordinate with one another on achieving socially optimal outcomes to increase the system level performance, and the traditional way of attaining this goal is to endow the agents with social rationality [7] - agents act as system utility maximizers. However, this is difficult to implement when we are facing open MAS domains such as peer-to-peer network and mobile ad-hoc networks, since we do not have control on all agents' behaviors in such systems and each agent usually behaves individually rationally as an individual utility maximizer only. In this paper, we propose injecting a number of influencer agents to manipulate the behaviors of individually rational agents and investigate whether the individually rational agents can eventually be incentivized to coordinate on achieving socially optimal outcomes. We evaluate the effects of influencer agents in two common types of games: prisoner's dilemma games and anti-coordination games. Simulation results show that a small proportion of influencer agents can significantly increase the average percentage of socially optimal outcomes attained in the system and better performance can be achieved compared with that of previous work.

1 Introduction

In certain multi-agent systems (MASs), an agent needs to coordinate effectively with other agents in order to achieve desirable outcomes, since the outcome not only depends on the action it takes but also the actions taken by others. How to achieve effective coordination among agents in multi-agent systems is a significant and challenging research topic, especially when the interacting agents represent the interests of different parties and they have generally conflicting interests.

It is well-known in game theoretic analysis that, if each agent behaves as an individual utility maximizer in a game, and always makes its best response to the behaviors of others, then the system can result in a Nash equilibrium such that no agent will have the incentive to deviate from its current strategy [1]. The equilibrium solution has its merits considering its desirable property of stability, however, it can be extremely inefficient in terms of the overall utilities the agents receive. The most well-known example is the prisoner's dilemma (PD) game: the

K. Glass et al. (Eds.): COMPLEX 2012, LNICST 126, pp. 140–151, 2013.

agents will reach the unique Nash equilibrium of mutual defection if the agents act as individually rational entities, while the outcome of mutual cooperation is the only optimal outcome in terms of maximizing the sum of both agents' payoffs. In certain domains, nevertheless, a more desirable alternative is a social optimal solution [7], in which the utilitarian social welfare (i.e., the sum of all agents's payoffs) is maximized.

One approach of addressing this problem is to enforce the agents to behave in a socially rational way - aiming at maximizing the sum of all agents' utilities when making their own decisions. However, as mentioned in previous work [4], this line of approach suffers from a lot of drawbacks. It will greatly increase the computational burdens of individual agents, since each agent needs to consider all agents' interests into consideration when it makes decisions. Besides, it may become infeasible to enforce the agents to act in a socially rational manner if the system is open in which we have no control on the behaviors of all agents in the system. To solve these problems, one natural direction is considering how we can incentivize the individually rational agents to act towards coordinating on socially optimal solutions. A number of work [4,9,8,5] has been done in this direction by designing different interaction mechanisms of the system while the individual rationality of the interacting agents is maintained and respected at the same time. One common drawback of previous work is that certain amount of global information is required to be accessible to each individual agent in the system and also there still exists certain percentage of agents that are not able to learn to coordinate on socially optimal outcomes (SOs).

In this work, we propose inserting a number of influencer agents into the system to incentivize the rest of individually rational agents to behave in a socially rational way. The concept of influencer agent is first proposed by Franks et al [2] and has been shown to be effective in promoting the emergence of high-quality norms in the linguistic coordination domain [11]. In general, an influencer agent is an agent with desirable convention or prosocial behavior, which is usually inserted into the system by the system designer aiming at manipulating those individually rational agents into adopting desirable conventions or behaviors. To enable that the influencer agents exert effective influences on individually rational agents, we consider an interesting variation of sequential play by allowing entrusting decision to others similar to previous work [6]. During each interaction between each pair of agents, apart from choosing an action from its original action set, each agent is also given the option of choosing to entrust its interacting partner to make a joint decision for both agents. It should be noted that such a decision to entrust the opponent is completely voluntary, hence the autonomy and rationality of an agent are well respected and maintained. The influencer agents are socially rational in the sense that they will always select an action pair that corresponds to a socially optimal outcome should it becomes the joint decision-maker. Besides, each agent is allowed to choose to interact with an influencer agent or an individually rational agent, and then it will interact with a randomly chosen agent of that type. Each agent (both individually rational and influencer agents) uses a rational learning algorithm to make its decisions

in terms of which type of agent to interact with and which action to choose, and improves its learning policy based on the reward it receives from the interaction. We evaluate the performance of the learning framework in two representative types of games: PD game and anti-coordination (AC) game. Simulation results show that a small proportion of influencer agents can efficiently incentivize most of the purely rational agents to coordinate on the socially optimal outcomes and better performance in terms of the average percentage of socially optimal outcome attained can be achieved compared with that of previous work [5].

The remainder of the paper is organized as follows. An overview of related work is described in Section 2. In Section 3, we give a description of the problem we investigate in this work. In Section 4, the learning framework using influencer agents we propose is introduced. In Section 5, we present our simulation results to compare the performance under the learning framework using influencer agents with previous work. Lastly conclusion and future work are given in Section 6.

2 Related Work

Hales and Edmonds [4] first introduce the tag mechanism originated in other fields (e.g., artificial life and biological science) into multi-agent systems research to design effective interaction mechanism for autonomous agents to achieve desirable outcomes in the system. They focus on the Prisoner's Dilemma game (called PD game hereafter), and each agent is represented by $1 + L$ bits. The first bit indicates the agent's strategy (i.e., playing C or D), and the remaining L bits are the tag bits, which are used for biasing the interaction among the agents and are assumed to be observable by all agents. In each generation each agent is allowed to play the PD game with another agent with the same tag string. The agents in the next generation are formed via fitness proportional reproduction scheme together with low level of mutations on both the agents' strategies and tags. This mechanism is demonstrated to be effective in promoting high level of cooperation among agents when the length of tag is large enough. However, there are some limitations of this tag mechanism. Since the agents mimic the strategy and tags of other more successful agents and the agents choose the interaction partner based on self-matching scheme, an agent can only play the same strategy as that of the interacting agent. Thus socially rational outcome can be obtained if and only if the agents need to coordinate on identical actions.

McDonald and Sen [8][9] propose three different tag mechanisms to tackle the limitations of the model of Hales and Edmonds. Due to space limitation, here we only focus on the third mechanism called paired reproduction mechanism since this is the only one that is both effective and practically feasible. It is a special reproduction mechanism which makes copies of matching pairs of individuals with mutation at corresponding place on the tag of one and the tag-matching string of the other at the same time. The purpose of this mechanism is to preserve the matching between this pair of agents after mutation in order to promote the survival rate of cooperators. Simulation results show that this mechanism can help sustaining the percentage of coordination of the agents at a high level in

both PD games and Anti-Coordination games (AC games hereafter), and this mechanism can be applied in more general multi-agent settings where payoff sharing is not allowed. However, similar to Hales and Edmonds' model [4], these mechanisms all heavily depend on mutation to sustain the diversity of groups in the system. Accordingly this leads to the undesired result that the variation of the percentage of coordination is very high.

Considering the disadvantages of evolutionary learning (heavily depending on mutation), Hao and Leung [5] develop a tag-based learning framework in which each agent employs a reinforcement learning based strategy to make its decisions. Specifically, they propose a Q-learning based strategy in which each agent's learning process is augmented with an additional step of determining how to update its Q-values, i.e., update its Q-values based on its own information or information of others in the system. Each update scheme is associated with a weighting factor to determine which update scheme to use each time and these weighting factors are adjusted adaptively based on a greedy strategy. They evaluate their learning framework in both PD game and AC games and simulation results show that better performance can be achieved in terms of both average percentage of socially optimal outcomes attained and the stability of the system compared with the paired reproduction mechanism [8].

3 Problem Description

The general question we are interested in is how individually rational agents can learn to coordinate with one another on desirable outcomes through repeated pairwise interactions. In particular, we aim to achieve socially optimal outcomes, under which the utilitarian social welfare (i.e., the sum of all agents' payoffs) is maximized. At the same time, we desire that the rationality and autonomy of individual agents be maintained. In other words, the agents should act independently in a completely individually rational manner when they make decisions. This property is highly desirable particularly when the system is within an open, unpredictable environment, since the system implemented with this kind of property can largely withstand the exploitations of selfish agents designed by other parties.

Specifically, in this paper we consider studying the learning problem in the context of a population of agents as follows. In each round each agent chooses to interact with another agent (i.e., to play a game with that agent), which is constrained by the interaction protocol of the system. Each agent learns concurrently over repeated interactions with other agents in the system. The interaction between each pair of agents is formulated as a two-player normal-form game, which will be introduced later. We assume that the agents are located in a distributed environment and there is no central controller for determining the agents' behaviors. Each agent can only know its own payoff during each interaction and makes decisions autonomously.

Following previous work [4,8,5], we focus on two-player two-action symmetric games for modeling the agents' interactions, which can be classified into two

different types. For the first type of games, the agents need to coordinate on the outcomes with identical actions to achieve socially rational outcomes. One representative game is the well-known PD game (see Fig. 1), in which the socially optimal outcome is (C, C), however, choosing action D is always the best strategy for any individually rational agent. The second type of games requires the agents to coordinate on outcomes with complementary actions to achieve socially optimal outcomes. Its representative game is AC game (see Fig. 2), in which either outcomes (C, D) or (D, C) is socially optimal. However, the row and column agents prefer different outcomes and thus it is highly likely for individually rational agents to fail to coordinate (achieving inefficient outcomes (C, C) or (D, D)). For both types of games, we are interested in investigating how the individually rational agents can be incentivized to learn to efficiently coordinate on the corresponding socially optimal outcomes.

A's payoff, B's payoff		Agent B's action	
		C	D
Agent A's action	C	R,R	S,T
	D	T,S	P,P

Fig. 1. Prisoner's dilemma (PD) game satisfying the constraints of $T > R > P > S$ and $2R > T + S > 2P$

A's payoff, B's payoff		Agent B's action	
		C	D
Agent A's action	C	L,L	H,H
	D	H,H	L,L

Fig. 2. Anti-coordination (AC) game satisfying the constraints of $H > L$

4 Learning Framework

We first give a background introduction on the concept of influencer agent and how it can be applied for solving our problem in Section 4.1. Then we describe the interaction protocol within the framework in Section 4.2. Finally the learning strategy the agents adopt to make decisions is introduced in Section 4.3.

4.1 Influencer Agent

The concept of influencer agent is firstly termed by Franks et al. [2], and there is also a number of previously work with similar ideas. In general, an influencer agent (IA) is an agent inserted into the system usually by the system designer in order to achieve certain desirable goals, e.g., emergence of efficient convention or norms [2]. Sen and Airiau [10] investigate and show that a small portion of agents with fixed convention can significantly influence the behavior of large group of selfish agents in the system in terms of which convention will be adopted in the system. Similarly Franks et al. [2] investigate the problem of how a small set of influencer agents adopting pre-fixed desirable convention can influence the rest of individually rational agents towards adopting the convention the system designer desires in the linguistic coordination domain [11].

Since we are interested in incentivizing individually rational agents to behave in the socially rational way, here we consider inserting a small number of influencer agents, which are socially rational, into the system. To enable the influencer agents to exert effective influence on individually rational agents' behaviors, we consider an interesting variation of sequential play by allowing entrusting decision to others similar to previous work [6]. During each interaction between each pair of agents, apart from choosing an action from its original action set, each agent is also given an additional option of asking its interacting partner to make the decision for both agents (denoted as choosing action F). If an agent A chooses action F while its interacting partner B does not, agent B will act as the leader to make the joint decision for them. If both agents choose action F simultaneously, then one of them will be randomly chosen as the joint decision-maker. The influencer agents are socially rational in that they will always select the socially optimal outcome as the joint action pair to execute whenever it becomes the joint decision-maker. If there exist multiple socially optimal outcomes, then these socially optimal outcomes will be selected with equal probability. For those individually rational agents, we simply assume that they will always choose the outcome under which their own payoffs are maximized as the joint action for execution, whenever they are entrusted to make joint decisions.

4.2 Interaction Protocol

From previous description, we know that there exist two different types of agents in the system: influencer agents (IA) and individually rational (or 'selfish') agents (SA). In each round, each agent is allowed to choose which type of agent to interact with, and then it will interact with an agent randomly chosen from the corresponding set of agents. This is similar to the commonly used interaction model that the agents are situated in a fully connected network in which each agent randomly interacts with another agent each round [10,12]. The only difference is that in our model the population of agents are divided into two groups and each agent is given the freedom to decide which group to interact with first but the specific agent to interact with within each group is still chosen randomly. Our interaction model can better reflect the realistic scenarios in human society, since human can be classified into different groups according to their personality traits and different persons may have different preferences regarding which group of people they are willing to interact with.

Each agent uses a rational learning algorithm to make its decisions in terms of which type of agent to interact with and which action to choose, and improves its policy based on the rewards it receives during the interactions. Besides, each agent chosen as the interacting partner also needs to choose an action to respond accordingly depending on the type of its interacting agent. We assume that during each interaction each agent only knows its own payoff and cannot have access to its interacting partner's payoff and action. The overall interaction protocol is shown in Algorithm 1.

Algorithm 1. Interaction Protocol

```
for a fixed number of rounds do
    for each agent i in the system do
        determine which type of agents to interact with
        interact with one agent randomly chosen from the corresponding set of agents
        update its policy based on the reward received from the interaction
    end for
end for
```

4.3 Learning Algorithm

For the individually rational agents, it is natural that they always choose the strategy which is a best response to its partner's current strategy in order to maximize its own payoff. If a learning algorithm has the property that it can converge to a policy that is a best response to the other players' policies when the other players' policies converge to stationary ones, then it is regarded as being rational [1]. A number of rational learning algorithms exist in the multi-agent learning literature and here we adopt the Q-learning algorithm [13], which is the most commonly used.[1] Specifically, each individually rational agent maintains two different set of Q-tables: one corresponding to the estimates of the payoffs for actions for interacting with influencer agents, Q_{IA}, and the other corresponding to the estimates of the payoffs for the set of actions for interacting with individually rational agents, Q_{SA}. In the following discussion, a_{IA} refers to an action when interacting with an influencer agent and a_{SA} refers to an action when interacting with an individually rational agent. In each round t, an individually rational agent i makes its decision (which type of agent to interact and which specific action to choose) based on the Boltzmann exploration mechanism as follows. Formally any action a_{IA} belonging to the set of actions available for interacting with influencer agents is selected with probability

$$\frac{e^{Q_{IA}(a_{IA})/T}}{\sum_{a_{IA}} e^{Q_{IA}(a_{IA})/T} + \sum_{a_{SA}} e^{Q_{SA}(a_{SA})/T}} \tag{1}$$

Any action a_{SA} belonging to the set of actions available for interacting with individually rational agents is selected with probability

$$\frac{e^{Q_{SA}(a_{SA})/T}}{\sum_{a_{IA}} e^{Q_{IA}(a_{IA})/T} + \sum_{a_{SA}} e^{Q_{SA}(a_{SA})/T}} \tag{2}$$

The temperature parameter T controls the exploration degree during learning, and initially it is given a high value and decreased over time. The reason is that initially the approximations of the Q-value functions are inaccurate and the

[1] Other rational learning algorithms such as WOLF-PHC [1] and Fictitious Play [3] will be investigated as future work.

agents have no idea of which action is optimal, thus the value of T is set to a relatively high value to allow the agents to explore potential optimal actions. After enough explorations, the exploration has to be stopped so that the agents can focus on exploiting the actions that has shown to be optimal before.

An individually rational agent selected as the interacting partner makes decisions depending on which type of agent it will interact with. If it interacts with another individually rational agent, then it will choose an action from the set of actions available for interacting with individually rational agents, and any action a_{SA} is chosen with probability

$$\frac{e^{Q_{SA}(a_{SA})/T}}{\sum_{a_{IA}} e^{Q_{SA}(a_{SA})/T}} \tag{3}$$

If it is chosen by an influencer agent, then it will pick an action from the set of actions available for interacting with influencer agents, and any action a_{IA} is selected with probability

$$\frac{e^{Q_{IA}(a_{IA})/T}}{\sum_{a_{IA}} e^{Q_{IA}(a_{IA})/T}} \tag{4}$$

After the interaction in each round t, each agent updates its corresponding Q-table depending on which type of agent it has interacted with. There are two different learning modalities available for performing update [12]: 1) multi learning approach: for each pair of interacting agents, both agents update their Q-tables based on the payoffs they receive during interaction, 2) mono learning approach: only the agent who initiates the interaction updates its Q-table and its interacting partner does not update. In mono learning approach, each agent updates its policy in the same speed, while in the multi learning approach, some agents may learn much faster than others due to the bias of partner selection. We investigate both updating approaches and the effects of both updating approaches on the system level performance will be shown in Section 5. Formally, each agent updates its Q-tables during each interaction as follows depending on which type of agent it has interacted with,

$$Q^{t+1}_{IA/SA}(a) = \begin{cases} Q^t_{IA/SA}(a) + \alpha_i(r^t_i - Q^t_{IA/SA}(a)) & \text{if } a \text{ is chosen in round } t \\ Q^t_{IA/SA}(a) & \text{otherwise} \end{cases} \tag{5}$$

where r^t_i is the reward agent i obtains from the interaction in round t by taking action a. α_i is the learning rate of agent i, which determines how much weight we give to the newly acquired reward r^t_i, as opposed to the old Q-value. If $\alpha_i = 0$, agent i will learn nothing and the Q-value will be constant; if $\alpha_i = 1$, agent i will only consider the newly acquired information r^t_i.

For influencer agents, we do not elevate their learning abilities above the rest of the population. They make decisions in the same way as individually rational agents. The only difference is that the influencer agents behave in a socially rational way in that they will always select the socially optimal outcome(s) as the joint action pair(s) to execute whenever it is selected as the joint decision-maker as described in Section 4.1.

5 Simulation

In this section, we present the simulation results showing how the influencer agents can significant influence the population's behaviors towards socially optimality in two types of representative games: PD game and AC game. All the simulations are performed in a population of 1000 agents. Following previous work [2], we consider 5% of influencer agents in a population (i.e., 50 agents out of 1000) to be an appropriate upper bound of how many agents can be inserted into a system in practical application domains. However, for evaluation purpose, we perform simulations with the percentage of influencer agents up to 50 % in order to have a better understanding of the effects of influencer agents on the dynamics of the system. We first give an analysis of the effects of the number of influencer agents and different update modalities on the system level performance in each game in Section 5.1 and 5.2 respectively, and then compare the performance of our learning framework using influencer agents with that of previous work [5] in Section 5.3.

5.1 Prisoner's Dilemma Game

For the PD game, we use the following setting: $R = 3$, $S = 0$, $T = 5$, $P = 1$. Fig. 3 shows the average percentage of socially optimal outcome in the system when the number of influencer agents (IAs) inserted varies using mono-learning update. When no IAs are inserted into the system (i.e., a population of individually rational agents (SAs)), the percentage of socially optimal outcomes (SOs) achieved quickly reaches zero. This is obvious since choosing action D is always the best choice for each individually rational agent when it interacts with another individually rational entity in PD game. By inserting a small amount of IAs with proportion of 0.002 (2 IAs in the population of 1000 agents), we can see a significant gain in terms of the percentage of SOs attained up to 0.8. The underlying reason is that most of SAs are incentivized to voluntarily choose to interact with IAs and also select action F. Further increasing the number of IAs (to 10 IA agents) can significantly improve the speed of increase of percentage of SOs and also bring in small improvement of the percentage of SOs finally attained. We hypothesize that it is because some IAs' behaviors against SAs are not optimal when the number of IAs is small, and more IAs can successfully learn the optimal action against SAs when the number of IAs becomes larger. However, as the number of IAs is further increased, the increase in the final value of the proportion of SOs attained becomes less obvious. The reason is that with the number of IAs increasing, there is little additional benefit on the behaviors of IAs and the small amount of increase in the percentage of SOs is purely the result of the increase of the percentage of IAs itself.

Fig. 4 shows the differences between updating using multi-learning and mono-learning approach on the system level performance, i.e., the average percentage of SOs attained in the system, in PD game. From previous analysis, we have known that most SAs learn to interact with IAs and choose action F, thus IAs are given much more experience and opportunities to improve their policies against SAs

under multi-learning approach. Accordingly it is expected that higher level of SOs can be achieved compared with that under mono-learning approach. It is easy to verify that the simulation results in Fig. 4 are in accordance with our predication.

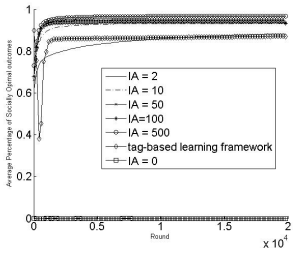

Fig. 3. Average percentage of SOs with different number of IAs under mono-learning update (PD game)

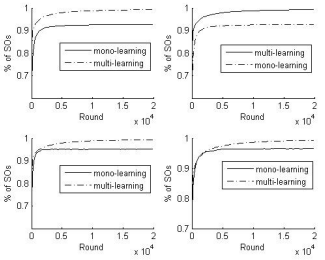

Fig. 4. Mono-learning approach v.s. multi-learning approach in PD game (IA = 2, 20, 100, 500)

5.2 Anti-coordination Game

In AC game, the following setting is used to evaluate the learning framework: $L = 1$, $H = 2$. Fig. 5 shows the average proportion of socially optimal outcomes attained in the system when the number of influencer agents varies. Different from the PD game, when there is no influencer agents inserted in the system, most of the selfish agents (up to 85%) can still learn to coordinate with each other on socially optimal outcomes. Initially the percentage of socially optimal outcomes is very high because most of the agents learn that choosing action F is the best choice which prevents the occurrence of mis-coordination. However, gradually the agents realize that they can benefit more by exploiting those peer agents choosing action F by choosing action C or D and thus this inevitably results in mis-coordination (i.e., achieving outcome (C, C) or (D, D)) when these exploiting agents interact with each other. Thus the average percentage of socially optimal outcomes gradually drops to around 85%. Besides, the mis-coordination rate converges to around 15 %, which can be understood as the dynamic equilibrium that the system of agents has converged to, i.e., the percentages of agents choosing action C, D, and F are stabilized.

Significant increase in the average percentage of socially optimal outcome attained (up to almost 100%) can be observed when a small amount of influencer agents (2 IAs out of 1000 agents) is inserted into the system. This can be explained by the fact that the SAs learn to entrust those IAs to make the joint decisions and also the IAs choose between the two socially optimal outcomes randomly. There is little incentive for the SAs to deviate since most of SAs have learned to interact with IAs and respond to SAs with action C or D, thus there is no benefits to exploit other SAs due to high probability of mis-coordination. When the number of IAs is further increased (the number of IAs is 50), there

is little performance increase in terms of the percentage of socially optimal outcomes achieved, and we only plot the case of $IA = 50$ for the purpose of clarity.

Fig. 6 shows the differences between updating using multi-learning and mono-learning approach on the system level performance, i.e., the average percentage of SOs attained in the system, in AC game with different number of IAs. Different from PD game, we can observe that slightly lower percentage of SOs is attained under multi-learning approach. We hypothesize that it is due to the fact that there exist two different socially optimal outcomes in AC game and thus it becomes easier for the agents to switch their policies between choosing these two outcomes and increase the chances of mis-coordination when they update their policies more frequently under multi-learning approach.

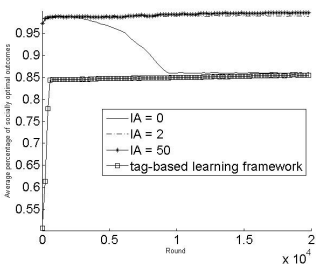

Fig. 5. Average percentage of SOs with different number of IAs under mono-learning update (AC game)

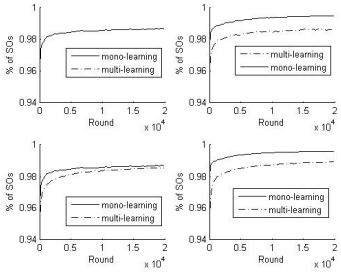

Fig. 6. Mono-learning approach v.s. multi-learning approach in AC game (IA = 2, 10, 100, 500)

5.3 Comparison with Previous Work

We compare the performance of our learning framework using influencer agents with that of the tag-based learning framework [5] in both PD game and AC game. The number of influencer agents are set to 50 (5 % of the total number of agents) in our learning framework. The experimental setting for the tag-based learning framework follows the setting given in [5].

Fig. 3 and 5 show the performance comparisons with the tag-based learning framework in PD game and AC game respectively. For both cases, we can observe that there is a significant increase in the average percentage of SOs under our learning framework using influencer agents. Besides, the rate in which the agents converge to SOs is higher than that using the tag-based learning framework. The underlying reason is that under the tag-based learning framework, the agents learn their policies in a periodical way, and the coordination towards socially optimal outcomes requires at least two consecutive periods' adaptive learning between individual learning and social learning. Also our learning framework using IAs can better prevent the exploitations from SAs since the IAs act in the same way as SAs if their interacting partners do not choose action F. Accordingly, higher percentage of SAs can be incentivized to cooperate with IAs and thus higher percentage of socially optimal outcomes can be achieved.

6 Conclusion and Future Work

In this paper, we propose inserting influencer agents into the system to manipulate the behaviors of individually rational agents towards coordination on socially optimal outcomes. We show that a small percentage of influencer agents can successfully incentivize individually rational agents to cooperate and thus achieve socially optimally outcomes.

As future work, we are going to give detailed analysis of the learning dynamics of both types of agents (IAs and SAs) in order to better understand the effects of influencer agents. Another interesting direction is to apply this learning framework to other MAS domains (e.g., other types of games), and investigate the effects of influencer agents on the learning dynamics of individually rational agents as well.

References

1. Bowling, M., Veloso, M.: Multiagent learning using a variable learning rate. Artificial Intelligence 136, 215–250 (2002)
2. Franks, H., Griffiths, N., Jhumka, A.: Manipulating convention emergence using influencer agents. In: AAMAS (2012)
3. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games. MIT Press (1998)
4. Hales, D., Edmonds, B.: Evolving social rationality for mas using "tags". In: AAMAS 2003, pp. 497–503. ACM Press (2003)
5. Hao, J.Y., Leung, H.F.: Learning to achieve social rationality using tag mechanism in repeated interactions. In: ICTAI 2011, pp. 148–155 (2011)
6. Hao, J., Leung, H.-F.: Learning to achieve socially optimal solutions in general-sum games. In: Anthony, P., Ishizuka, M., Lukose, D. (eds.) PRICAI 2012. LNCS, vol. 7458, pp. 88–99. Springer, Heidelberg (2012)
7. Hogg, L.M., Jennings, N.R.: Socially rational agents. In: Proceeding of AAAI Fall Symposium on Socially Intelligent Agents, pp. 61–63 (1997)
8. Matlock, M., Sen, S.: Effective tag mechanisms for evolving coordination. In: AAMAS 2007, pp. 1–8 (2007)
9. Matlock, M., Sen, S.: Effective tag mechanisms for evolving coperation. In: AAMAS 2009, pp. 489–496 (2009)
10. Sen, S., Airiau, S.: Emergence of norms through social learning. In: IJCAI 2007, pp. 1507–1512 (2007)
11. Steels, L.: A self-organizing spatial vocabulary. Artificial Life 2(3), 319–392 (1995)
12. Villatoro, D., Sen, S., Sabater-Mir, J.: Topology and memory effect on convention emergence. In: WI-IAT 2009, pp. 233–240 (2009)
13. Watkins, C.J.C.H., Dayan, P.D.: Q-learning. Machine Learning, 279–292 (1992)