Proceedings of the 36th Chinese Control Conference
July 26-28, 2017, Dalian, China

A Leader-Based Cooperation-Prompt Protocol for the Prisoner's Dilemma Game in Multi-Agent Systems

Linghui Guo[1,2], Zhongxin Liu[1,2,*], Zengqiang Chen[1,2]

1. College of Computer and Control Engineering, Nankai University, Tianjin, 300353, China
E-mail: lzhx@nankai.edu.cn

2. Tianjin Key Laboratory of Intelligent Robotics, Nankai University, Tianjin, 300353, China

Abstract: In this paper, we propose a new protocol that prompts the cooperation rate of a multi-agent system, the members of which compete based on the prisoner's dilemma game. The main idea of this algorithm is to introduce a smart leader to the system, which can be viewed as a third player in the original bilateral game. The leader is motivated by its own benefit, but as a result of its attempt to maximum the expectational personal benefit, the system's cooperation rate is actually improved. This phenomenon is common in reality in some complicated situations. We studied the condition in which the leader maximums its expectational payoff, and proved the optimal condition's existence and uniqueness. Furthermore, we adopted a feed-back control method to ensure that the system globally converges to and remains stable in the optimal point. The simulation results illustrate the practicability and efficiency of our algorithm in reality system.
Key Words: Game theory, Evolutionary game theory, Leader-follower, Prisoner's Dilemma, Multi-agent system

1 Introduction

Any well-governed society can't get rid of good management, and such management could be artificial or spontaneous. To maximum the warfare of the society, the manager should prompt individuals to cooperate rather than betray with each other. In game theory, the cooperation phenomenon, especially the spontaneous one, challenges the basic game theory rule and the Darwin's natural selection theory, but does actually exist in nature and human society. The study on this topic is a rising area that grabs the interest of many different fields [1-7].

As the coming of the information era and the Internet evolution, the online market has witnessed a booming during the last few decades, which brings about new issues. For example, the management of this new business mode is a more complicated problem than that of the traditional one. According to the traditional game theory, the individuals should trend to maximum their own interest in practice, and if betraying the partner is more beneficial than cooperating, they would rarely risk choosing cooperation. This phenomenon is more common in the Internet market, where the business connections are more free but unreliable, and people's behavior would be more complex and interesting.

One of the results of the cooperation crisis in online business is the rising of all kinds of electronic commerce platforms, like eBay, Amazon, Alibaba, etc. These Internet based enterprises are special participants in online commercials. They make connections among people, regulate the basic rules, and provide convenience for either side of business trades, while they themselves are also players in practical business. Such complicated roles in business are not easy to be interpreted by traditional game theory model.

In the field of game theory, the matrix game is a simple and efficient tool to analyze the complex situations in reality [8,9], which involves the famous models such as the prisoner's dilemma, the snowdrift game, etc. The classic prisoner's dilemma can be presented as a symmetric matrix game, in which the unilateral betray is always more favorable for one player than mutual cooperation. The outcome of the game depends on many factors, like the system's population structure [3,4,8-11], the interaction endurance[12-14],the group distribution[15-18], or the properties of the game itself [8-10,19].

In prisoner's dilemma game, there are several ways to prompt the cooperation rate of multi-agent system. Inspired by the learning theory, one well-studied way is to apply to the game players reinforce learning method [20-25]. For structured populations, the spatial connection could be another efficient factor to prompt cooperation [26-27]. Furthermore, reputation is also used to enhance the agents' trend to cooperation [28-30].

The leader-follower based control for multi-agent system is also an important area in control theory [31-33]. Based on this idea, the system is controlled not by the direct controlling of individuals but by just a few leaders in the system. The leaders are individuals with special influence to a certain amount of followers. The leader-follower based algorithms have made a great success in control problems such as consensus problem, formation control, flight control, concurrent computation, social management, etc.

In this paper, we apply the role of the leader to the evolutionary game system. Since both the evolutionary game theory and the multi-agent system are new and rapidly developing area, the idea of leadership, to our knowledge, has not been fully introduced to the study of evolutionary game theories. In this work, we explore the leader's effect as a way to improve the cooperation level of the evolutionary prisoner's dilemma game system, the outcome of which gives a novel sight of evolutionary game theory.

The rest of this paper is organized as follows. In section 2 we propose the foundation of our model. In section 3 we make theoretical analysis of our model. In section 4, we talk about the operable control method to fulfill this mode. The mode's simulation result will be show in section 5.

* This research is supported by the National Natural Science Foundation of China (Grant No.61573200,61573199), and the Tianjin Natural Science Foundation of China (Grant No.14JCYBJC18700, 14JCZDJC39300).

2 Model Introduction

In this paper, we apply the well-known model, namely the Prisoner's Dilemma Game (PDG), to an evolutionary system. Unlike the traditional cases, we introduce a smart leader that interferes in the payoff of individual games to encourage the system to get higher cooperation rate. Then we analysis both the behavior of the leader and the cooperation trend of individual players to examine the effectiveness of this leadership.

Firstly, we define the simple PDG based evolutionary system without the leader.

Suppose n players (or, prisoners) are attributed in one system, from which, at each periodical time point, two individuals are randomly selected to play a PDG game with each other, and get paid the corresponding payoff. Both the players may choose to cooperate(C-player) with or defect (D-player) from his or her partner, and the payoff of a player is denoted by the matrix M, where

$$M = \begin{bmatrix} R & S \\ T & P \end{bmatrix} \tag{1}$$

The elements, R, S, T and P, denote the different income of one player under different combination of strategies of both sides. In game theory [1], it's well known that in the PDG, the variations of M satisfy $T > R > P > S$ and $2R > T$, which indicates that the strategy of cooperation is a strictly dominated strategy.

To define the dynamic property of the system, we suppose that at each time point, a certain number of players in the system are randomly chosen to die. Then, the rest players complete to duplicate themselves to take the place of dead players. The probability they get this opportunity is proportional to their personal payoffs in their latest PDGs.

According to the evolutionary game dynamic theory [1], the change of the proportion of C-players (p_c) in the system satisfies:

$$\dot{p}_c = p_c(u_c - \bar{u}) \tag{2}$$

where u_c denotes the average payoff of the C-players and $\bar{u}$ denotes the total average payoff. Since cooperation is a strictly dominated strategy, p_c should decrease to 0 over time despite the initial state of the system.

Secondly, we introduce the leader who has the power to modify the PDG above to prompt cooperation.

Suppose we allow the leader to modify the PDG in this way: $\delta : 0 < \delta < 1$ is a value defined as a constant. At time t, $a = a(t) \geq 0$ is regulated by the leader, then the payoff of the PDG would be modified into:

$$M(a) = \begin{bmatrix} R & S + \delta a \\ T - a & P \end{bmatrix} \tag{3}$$

Then, in the modified PDG, when a C-player meets a D-player, the former is encouraged by δa while the latter is punished in a . Meanwhile, in this case, the leader is as well paid by $(1-\delta)a$. In this sense, this smart leader could be viewed as a third partner in the PDG.

At last, we define the optimization problem for the leader to determine the best parameter a .

For an altruistic leader, it's intuitive that he or she should choose a to be large enough to drive the system to uniform cooperation, which maximums the total warfare, but this is impossible in our case. Here, we suppose the motivation of the leader is the profit of himself. The his expected profit in one game can be easily calculated as

$$J(a) = 2(1-\delta)ap_c(1-p_c) \tag{4}$$

If we assume that a is not constant but varies over time, the optimal function would be the integration of the leader's expected gains. Then, the problem becomes a nonlinear dynamic programming problem and would be difficult to calculate. To simplify this question, we only optimize the system's equilibrium state, in which both a and p_c remains unchanged.

In a nutshell, the smart leader's optimization target could be written as:

$$\begin{aligned} &\max_a J(a) = 2(1-\delta)ap_c(1-p_c) \\ &s.t. \quad \mathrm{a} \geq 0 \\ &\qquad \dot{p}_c(a, p_c) = 0 \end{aligned} \tag{5}$$

Remark 1: As the formula (5) shows, in equilibrium state, the income of the leader depends both on the value of a and the current cooperation level, which should not be too high or too low. In this sense, the leader will prefer to not conducting the whole system into uniform cooperation. Besides, because $0 < \delta < 1$ is a constant, we can omit it in equation (5). But since the value of δ also affects the dynamic behavior of p_c , the optimal result still depends on it.

3 Theoretical Analysis

Based on the optimal problem we defined above, we now propose our main theory in this paper:

Theorem 1 *If the PDG and the evolutionary system satisfy the requirements defined above, and δ is defined as a constant between 0 and 1, there should be a unique solution $p_c = p^* \in (0,1)$ for the leader's optimal problem (5).*

Proof: In equilibrium condition, the cooperation rate p_c should remain unchanged, which means p_c and a must satisfy the relationship below:

$$\begin{aligned} 0 &= \dot{p}_c = p_c(u_c - \bar{u}) \\ &= p_c[Rp_c + (S+\delta a)(1-p_c) - Rp_c^2 - (S+\delta a)p_c(1-p_c) \\ &\quad -(T-a)p_c(1-p_c) - P(1-p_c)^2] \\ &= p_c(1-p_c)[(R-S+P-T)p_c + S - P - (\delta p_c - p_c - \delta)a] \end{aligned} \tag{6}$$

We can deduce that $p_c = 0$ or $p_c = 1$ or

$$a = \frac{(T+S-R-P)p_c + P - S}{(1-\delta)p_c + \delta} \tag{7}$$

must be satisfied. But the former two conditions lead to $J(a) = 0$, hence could not be the solution to the leader's optimal profit. So , we get the equilibrium condition:

$$a = \frac{Up_c + V}{(1-\delta)p_c + \delta} \tag{8}$$

where

$$U = T + S - R - P \tag{9}$$

$$V = P - S \tag{10}$$

And then, submitting a to $J(a)$ helps us get

$$J(a(p_c)) = \frac{2(1-\delta)(Up_c + V)(1-p_c)p_c}{(1-\delta)p_c + \delta} \tag{11}$$

By making derivation of J,

$$\frac{dJ}{dp_c} = 2(1-\delta)\frac{G(p_c)}{[(1-\delta)p_c + \delta]^2} \tag{12}$$

where

$$G(p_c) = -2U(1-\delta){p_c}^3 - [(1-\delta)(V-U) + 3\delta U]{p_c}^2 - 2\delta(V-U)p_c + \delta V \tag{13}$$

we solve the equation $G(p_c) = 0$ to obtain the optimal solution p^*. By noticing

$$G(0) = \delta V = \delta(P-S) > 0, \tag{14}$$

$$G(1) = -(V+U) = -(T-R) < 0 \tag{15}$$

and the continuity of $G(p_c)$, we can guarantee that $J(p_c)$ has at least one local maximum point $p^* \in (0,1)$.

To prove the uniqueness of p^*, we need to discuss under different conditions of U since $G(p_c)$ is at most a cubic equation.

For the case of $U \le 0$, we get $-2U(1-\delta) \ge 0$. Then $G(p_c)$, as a cubic or quadratic function, has at most one descending interval. Since $G(0) > 0$, $G(1) < 0$, in the interval $(0,1)$, it's easy to see that $G(p_c)$ meets zero exactly once.

For the case of $U > 0$, we get $-2U(1-\delta) < 0$, which can be divided into two conditions.

On one hand, if $V - U \le 0$, it's easy to see that $G'(0) = -2\delta(V-U) \ge 0$. Then, since $G(p_c)$ is a cubic function and $-2U(1-\delta) < 0$, $G(p_c)$ could only have one ascending interval. Besides, when $p_c > 0$, there could only be one descending interval. Then, since $G(0) > 0$, the zero point in $(0,1)$ must be unique.

On the other hand, if $V - U > 0$, we have

$$G(-x) = 2U(1-\delta)x^3 - [(1-\delta)(V-U) + 3\delta U]x^2 + 2\delta(V-U)x + \delta V \tag{16}$$

And

$$2U(1-\delta) > 0 \tag{17}$$

$$-[(1-\delta)(V-U) + 3\delta U] < 0 \tag{18}$$

Then, according to the Routh Stability Criterion, $G(-x)$ has at least one root on the right half of the complex plane, which means $G(p_c)$ should at least has one root on the left half of the complex plane. Then, according to the shape of $G(p_c)$, it can only have odd number of zero points in $(0,1)$, which could only be one. ■

Theorem 1 ensures the leader to find the stable and most profitable $p_c = p^* \in (0,1)$ as he manipulates the PDG on the evolutionary system. So, once M and δ are determined for the game, the leader first solves the equation $G(p_c) = 0$ to find an optimal point $p^* \in (0,1)$. Then, he manipulates $a(t)$ to drive the cooperation rate change until $p_c = p^*$, where he will acquire maximum stable income. At this point, $a(t)$ remains constant and

$$a(t) = a^* = \frac{(T+S-R-P)p^* + P - S}{(1-\delta)p^* + \delta} \tag{19}$$

4 Feedback Control Technique

Although we have proven the existence of the optimal equilibrium point in theory, it might not be a stable one. In fact, whenever $a(t)$ is set as constant satisfying (19), the cooperation rate p_c will inevitably deviate the designed equilibrium point along a random trajectory until it reaches $p_c = 0$ or $p_c = 1$.

In order to confine p_c near the designed equilibrium point, we adopt the feedback control technique to (19).

According to the dynamic replication equation,

$$\dot{p}_c = p_c(1-p_c)[(R-S+P-T)p_c + S - P - (\delta p_c - p_c - \delta)a] \tag{20}$$

Taking a as a control variant and setting the feedback coefficient k>0, we get a feedback control law:

$$a = \frac{(T+S-R-P)p_c + P - S - k(p_c - p^*)}{(1-\delta)p_c + \delta} \tag{21}$$

Then we have $\dot{p}_c = -kp_c(1-p_c)(p_c - p^*)$. It's obvious that, under this control policy (21), p^* is a stable point when $0 < p < 1$, and (19) is satisfied at this point, which can be summarized as theorem 2:

Theorem 2 *In the leader-based PDG evolutionary system and the feedback control policy we defined above, for any initial state* $0 < p_c < 1$, *the system's cooperation rate will asymptotically converge to the designed equilibrium point* p^*.

Proof: By noting that $p_c > 0$, $1 - p_c > 0$, and with the function $V(p_c) = (p_c - p^*)^2$ always decreasing, the proof of theorem 2 is trifle. ■

In the following text, we will apply this feedback control policy to the experiment to check our theory and the optimal value we get.

5 Simulation Results

Based on the model we introduced above, we now set the parameters of PDG as:

$$M = \begin{bmatrix} 0.7 & 0 \\ 1 & 0.1 \end{bmatrix} \tag{21}$$

Furthermore, we set $\delta = 0.6$ and the feedback coefficient k = 0.8. The total number of individuals in the system is set to n= 5000.

In every iteration, we arrange for each individual to go and find a partner, play the PDG, and get a payoff. Then we randomly pick 2% of the individuals to die so that the other individuals compete to take their places.

According to our optimal condition, we set $G(p_c)$ equals to zero and get the ideal cooperation rate: $p^* = 0.4874$. Then, we launch the iteration, record the leader's total acquirement in every circle, and averaged it to every game played in that circle.

Fig. 1 illustrates the efficiency of the feedback control policy on this dynamic evolutionary system. Although the initial cooperation rate is relatively low, under the influence of the leader, the cooperation rate finally comes to the desired level and so does the leader's own profit.

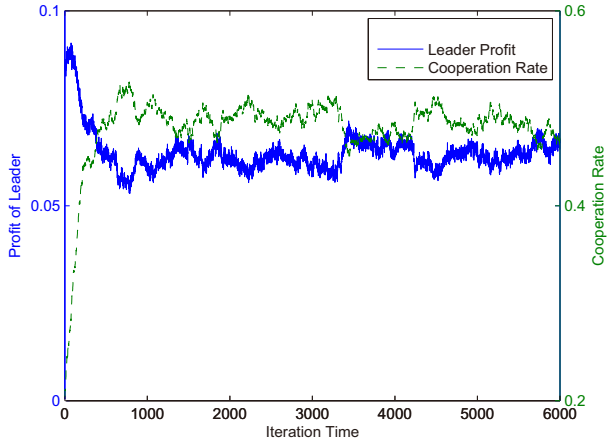

Fig. 1: The variation of the profit of the leader (left axis) and the cooperation rate (right axis) of the system. Both of them converge to a relatively steady level after a shot time interval.

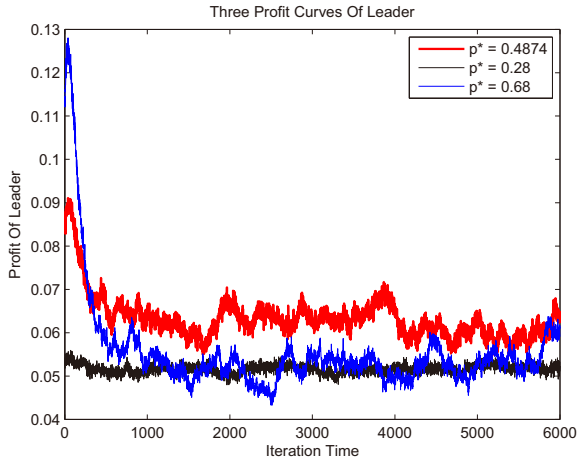

Fig. 2: The variation of the leader's profit (Y-axis) under three different levels of p^*. Only the red curve corresponds to the optimal value of p^*, The simulation illustrates that in the stable state, only the optimal level we calculated makes the leader acquire the highest income.

In Fig.2, we applied two extra values of p^*, which is either higher or lower than the optimal value. When p^* is set higher, although it leads to a higher payoff at the very early stage of the game, as the cooperation rate converges to p^*, leader's personal payoff decreases very fast and could only be maintained in a relatively low level. On the other hand, when p^* is set lower, both the rate of cooperation and the leader's profit in a single game gets very low, both of which lead to the lower level of the leader's income.

We illustrate the optimal condition more impressively in fig.3, in which we set the desired cooperation value switch from 0 to 1 with an interval of 0.02, and it's easy to see that the leader's own profit reaches its climax near the theoretical optimal value.

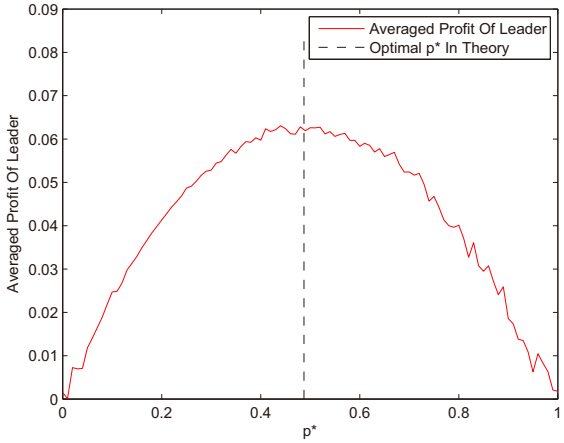

Fig. 3: The leader's profit on different designed cooperation rate is shown above, which arrives its climax around $p^* = 0.4874$, the theoretical optimal value.,

6 Conclusion

In this paper we proposed a new model based on the classic PDG game and the evolutionary system. In this model, a new player, the leader, is introduced to improve the cooperation rate of the system to relatively higher level. With the feedback control protocol we proposed, the higher cooperation rate can be reached and maintained, but the value usually is not 100 percent, which is determined by the leader's consideration of its own interest.

Since in recent years, the online business platforms is developing rapidly, the business regulators in online market is becoming a server rather than a governor. So our future work includes how to dig up this model to study the regular of market activity in reality. By refining this model, we can study what kind of leader is better to improve the market cooperation. Based on this model, we can also study the competing and bargaining behavior between different kinds of leaders in the same market which will open a new area to study the economy phenomenon.

References

[1] D. Fudenberg and D. Levine. The theory of learning in games[B]. *The MIT Press*, 1996.

[2] M. A. Nowak, K. M. Page, and K. Sigmund. Fairness Versus Reason in the Ultimatum Game[J], *Science*, 289(5485): 1773-1775, 2000.

[3] A. Traulsen and M. A. Nowak. Evolution of cooperation by multilevel selection[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 103(29): 10952-10955, 2006.

[4] M. A. Nowak. Five Rules for the Evolution of Cooperation[J]. *Science*, 314(5805): 1560-1563, 2006.

[5] M. A. Nowak and K. E. Sigmund. Evolutionary Dynamics of Biological Games[J]. *Science*, 303(5659): 793-799, 2004.
[6] L. I. Aming, W. U. Bin, and W. Long. A simple rule for robust stabilization of evolutionary dynamics[C]. *conference on computational complexity*, 2013: 1170-1175, 2013.
[7] D. Cheng, H. Qi, F. He, T. Xu, and H. Dong. Semi-tensor product approach to networked evolutionary games[J]. *Control Theory and Technology*, 12(2):198-214, 2014.
[8] C. E. Tarnita, N. Wage, and M. A. Nowak. Multiple strategies in structured populations[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6):2334, 2011.
[9] F. Fu and M. A. Nowaks. Global migration can lead to stronger spatial selection than localmigration[J]. *Journal of Statistical Physics,* 151(3):637-653, 2013.
[10] M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations[J]. *Philosophical Transactions of the Royal Society B*, 365(1537): 19-30, 2010.
[11] M. G. Zimmermann, and V. M. Eguíluz. Cooperation, social networks, and the emergence of leadership in a prisoner's dilemma with adaptive local interactions[J]. *Physical Review E,* 72(5), 2005.
[12] L. A. Imhof and M. A. Nowak. Stochastic Evolutionary Dynamics of Direct Reciprocity[J]. *Proceedings of The Royal Society B: Biological Sciences*, 277(1680): 463-468, 2010.
[13] H. Ohtsuki and M. A. Nowak. Direct reciprocity on graphs[J]. *Journal of Theoretical Biology*, 247(3): 462-470, 2007.
[14] Y. C. Zhang, M. A. Aziz-Alaoui, C. Bertelle, S. Zhou and W. T. Wang. Emergence of Cooperation in Non-scale-free Networks[J]. *Journal of Physics A*, 47(22), 2014.
[15] R. Boyd and P. J. Richerson. Group selection among alternative evolutionarily stable strategies.[J]. *Journal of Theoretical Biology*, 145(3): 331-342, 1990.
[16] A. Traulsen and M. A. Nowak, Evolution of cooperation by multilevel selection[J]. *Proceedings of the National Academy of Sciences of the United States of America,* 103(29): 10952-10955, 2006.
[17] S. D. Wilson, V. M. Van, and R. Ogorman. Multilevel Selection Theory and Major Evolutionary Transitions Implications for Psychological Science[J]. *Current Directions in Psychological Science*, 17(1): 6-9, 2008.
[18] J. A. Fletcher and M. Zwick. The Evolution of Altruism: Game Theory in Multilevel Selection and Inclusive Fitness[J]. *Journal of Theoretical Biology*, 245(1): 26-36, 2007.
[19] J. Zhang, Z. Chen, Z. Liu, and C. Zhang. Evolutionary dynamics and individual heterogeneity in multi-agent networking systems[C]. *Conference On Computational Complexity*: 7640-7645, 2016.
[20] S. Dridi and L. Lehmann. A model for the evolution of reinforcement learning in fluctuating games[J]. *Animal Behaviour* 2015: 87-114, 2015.
[21] M. Bowling. An Analysis of Stochastic Game Theory for Multiagent Reinforcement Learning[J]. *Computer Science Department Carnegie Mellon University*, 2000.
[22] M. I. Abouheaf, F. L. Lewis, K. G. Vamvoudakis, S. Haesaert, and R. Babuska. Multi-agent discrete-time graphical games and reinforcement learning solutions[J]. *Automatica*, 50(12): 3038-3053, 2014.
[23] S. Tanabe and N. Masuda. Evolution of cooperation facilitated by reinforcement learning with adaptive aspiration levels[J]. *Journal of Theoretical Biology*, 2011:151-160, 2011.
[24] S. Kar, J. M. F. Moura, and H. V. Poor. QD-Learning : A Collaborative Distributed Strategy for Multi-Agent Reinforcement Learning Through[J]. *IEEE Transactions on Signal Processing*, 61(7): 1848-1862, 2013.
[25] N. Masuda and H. Ohtsuki. A Theoretical Analysis of Temporal Difference Learning in the Iterated Prisoner's Dilemma Game[J]. *Bulletin of Mathematical Biology*, 71(8): 1818-1850, 2009.
[26] M. E. J. Newman. The Structure and Function of Complex Networks[J]. *Siam Review*, 45(2): 167-256, 2003.
[27] H. H. Nax and M. Perc. Directional learning and the provisioning of public goods.[J]. *Scientific Reports*, 2015: 8010-8010, 2015.
[28] J. Li, C. Zhang, Q. Sun, Z. Chen, and J. Zhang. Changing Interaction Intensity via Evaluating Individual Behavior in Iterated Prisoner's Dilemma[J]. *IEEE Transactions on Evolutionary Computation*, 2016: 1-1, 2016.
[29] H. Brandt and K. Sigmund. Indirect reciprocity, image scoring, and moral hazard[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7): 2666-2670, 2005.
[30] U. Berger. Learning to cooperate via indirect reciprocity[J]. *Games and Economic Behavior,* 72(1): 30-37, 2011.
[31] W. Ren. Second-order Consensus Algorithm with Extensions to Switching Topologies and Reference Models[C]. *American Control Conference*, 2007: 1431-1436, 2007.
[32] W. Li, Z. Chen, and Z. Liu. Leader-following formation control for second-order multiagent systems with time-varying delay and nonlinear dynamics[J]. *Nonlinear Dynamics,* 72(4): 803-812, 2013.
[33] W. Qin, Z. Liu, and Z. Chen. A novel observer-based formation for nonlinear multi-agent systems with time delay and intermittent communication[J]. *Nonlinear Dynamics*, 79(3): 1651-1664, 2015.