Applied Mathematics and Computation 426 (2022) 127093

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

The influence of experienced guider on cooperative behavior in the Prisoner's dilemma game

Tao You*, Hailun Zhang, Ying Zhang, Qing Li, Peng Zhang, Mei Yang

Department of Computer Science, Northwestern Polytechnical University, Xi'an, China

ARTICLE INFO

Article history:
Received 25 November 2021
Revised 22 February 2022
Accepted 12 March 2022
Available online 1 April 2022

Keywords:
Cooperation
Reinforcement learning
Multi-layer network
Prisoner'S dilemma game

ABSTRACT

In game theory, it is an essential topic to study the emergence and maintenance of cooperative behavior in groups based on the theories of evolutionary game and complex network. Unfortunately, an in-depth analysis of cooperative behavior on maintenance and development is usually challenged by the diversity of groups in society, which is mainly caused by the single mechanism in traditional networks. More recent studies have shown that multi-layer coupled network based evolutionary game theory is promising in exploiting the transmission of cooperative behavior between individuals in the game. Meanwhile, inspired by the decisive ability of reinforcement learning in overcoming the limitation of replica, in this work, we propose to combine the game strategy of reinforcement learning with the traditional prisoner's dilemma strategy based on multiple coupled networks. The most advantage of this model is the improved capability of intelligent decision making for group behaviors. With the simulation of game evolution, the influence of individual strategy change, as well as individual ability on cooperative behavior in reinforcement learning, is also explored. Substantial validations have verified that in social dilemmas, the cooperative behavior can be maintained by adjusting the group's ability with effective guidance.

1. Introduction

Cooperative behavior is one of the most common group activities in different living systems. As a basic rule in the field of evolutionary biology [1], this selfless and altruistic behavior is able to enhance the competitive superiority of the group. From another point of view, according to Darwin's theory of natural selection [2], individuals who adopt this behavior may lose their reproductive superiority and gradually perish as a consequence. In last decades, a quantity of practical studies have shown that the co-existence between collaborators and betrayers can last for a long time, but how to maintain such cooperative behaviors, as well as to understand their emergence, are still the challenges to conduct in-depth research [3–5].

To effectively analyze the interaction between individuals in structural populations [6–10], the complex networks play an essential role in the studies of cooperation evolution and strategic competition by incorporating with evolutionary game theory [11–13]. Based on the study of evolution of group cooperative behavior on a spatial grid network, Nowak *et. al.* advocated to concluded that the spatial structure of the population can promote cooperation [14]. Followed, five fundamental theories have been proposed in the field of game theory to explain the cooperative behavior [15] of human, including group selection, gene selection, indirect reciprocity, direct reciprocity and network reciprocity [16–20]. Scholars have also designed

* Corresponding author.
E-mail address: yoytao@yeah.net (T. You).

https://doi.org/10.1016/j.amc.2022.127093
0096-3003/

a variety of game strategies based on different phenomena in nature and society, such as aspiration, reputation, imitation and voluntary participation [21–26,28]. Furthermore, Hiromu *et. al.* proposed the sixth reciprocity which is a dynamic utility rule [29]. Since the overall repeated games are dynamic in principle, the mechanism of [27] can be applied to any game. All of the theories above give an important reference for us to explain the interaction behavior of individuals in the game.

A variety of network structures have been introduced into the evolutionary game modeling, which pushed forward the study of the spatial structure of populations and the evolution of cooperative behavior. E.g. Santos *et. al.* found that the cooperation on scale-free networks can exist in a wide range, and the corresponding network heterogeneity has a direct proportional impact on the evolution of cooperation [30,31]. For the multi-layer coupled network in prisoner's dilemma game, the heterogeneity and edge weight were found to effectively promote of cooperative behavior by enriching the spatial reciprocity [32,33,35].

For the decision making in those works discussed above, the individuals only consider the benefits brought in the current situation, which means that they discard the experiences obtained from previously games that they have accomplished. But the living creatures of natural world are able to adjust their social behavior based on cues from the environment, including feedback from previous experiences. This indicates that their behaviors are information-memorizing, which is highly relevant to the principles of reinforcement learning [36–38]. As a representative simulation of individual decision making in evolutionary game, Nowak *et al.* [39] studied the reinforcement learning with dynamic expectation level in the iterative PDG. It was found in [39] that the individual's behavior in simply modified BM model has changed dramatically under the action of adjustable expectation level. In addition, if the correlation between the reinforcement signal and the next round of action became strong enough, the individuals who follow the reinforcement learning rules would have a high probability of learning from each other, and also be highly cooperative and competitive in the evolutionary dynamics.

In a similar way, T. Ezaki et al. [40] established an association between the aspiration learning and conditional cooperation. Meanwhile, a numerical convincible evidences were obtained by comparing the association with relevant experiments. This work demonstrate that the BM is able to explain the experimental results more accurately as a simple expectation-based reinforcement learning model, which is completely different from the previous studies of [41]. It was also found when making strategic choices in social dilemmas, a rule of engaging alternative evolutionary can be obtained by aspiration learning for individuals.

To simulate the group diversity and individual decision-making intelligence more realistically, a multi-layer coupled network structure is introduced in the proposed work. Also with a prisoner's dilemma model by following the reinforcement learning strategy, the evolutionary game is studied by Monte Carlo simulation [42,43] as well. Here, the individuals with reinforcement learning models are regarded as the experienced guiders, which establish a correlation between multiple rounds of game actions across the adjustable expectation level.Suppose there are groups of people with different experiences in society, it would be more effective for the individuals with rich experience to guide the individuals with lack of experience in solving problems. The reason behind is that the experienced individuals often have the learning ability of self reflection, while inexperienced individuals often only imitate the surrounding strategies [44,45]. Therefore, we describe experienced individuals with reinforcement mechanism and inexperienced individuals in the traditional game mechanism, which is to make learning guide imitation. Specifically, the probability of adopting a cooperative strategy can change adaptively according to BM model dynamics. With different reinforcement learning model parameters, the Monte Carlo simulation is employed to further verify the strategy choice of guides and the guided ordinary individuals. Moreover, the impact of ordinary players' learning ability on game evolution is also exploited based on simulation results, which are also utilized to explore the survival and evolution methods of cooperative groups in society under severe social difficulties. The remainder of the paper is organized as follows: the proposed work is introduced elaboratively in Section 2, the experiment results are presented with analysis in Section 3, and in Section 4, we conclude this work.

2. Model

Fig. 1 presents the proposed two-layer networks, which are the $L * L$ square grid networks with periodic boundary conditions. In the networks, each node corresponds to the other nodes on another layer. Players following reinforcement learning rules in PDG (as experienced guiders) are placed at each node of the upper network(the enhancement layer), while players in the traditional PDG are placed at the lower network(the traditional layer) (Fig. 1). The game evolves as each player interacts with the four neighbors around him. Each player chooses one strategy: cooperation(C) or defection(D),for each round of the game, and all the guider's (x) strategies are updated firstly, then the ordinary player (x).

The same strategy will be applied to the game against the four neighbors, which means that players are not allowed to cooperate with one neighbor but betray other neighbors in the same round.

In prisoner's dilemma game, there are three pairs of strategies that might be chosen by two players: the pairwise cooperation (C,C) with reward R, mutual defection (D,D) resulting in punishment P, the mixed strategies reward cooperator in (C,D) or (D,C) with the sucker payoff S. The reward given to the defector is the temptation value of T. The classic model of PDG, Donor & Recipient game (D & R game) in [41–48] used multiple parameters to represent the player's payoff. In order to study the influence of the proposed mechanism on the game evolution more intuitively, we take a specific boundary game as the research model. Different from D & R game, only one parameter is used in this model to describe the range of social

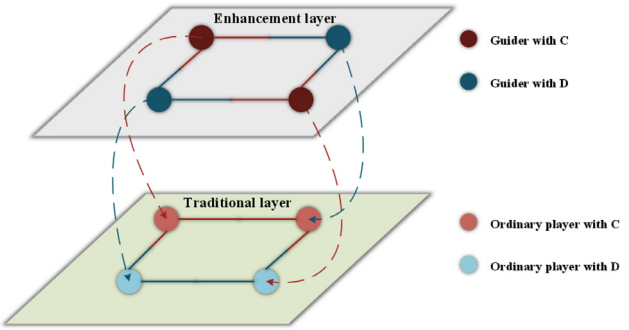

Fig. 1. Different types of players in two-layer networks, and the lower layer network has one-way dependence on the upper layer network. Players in the upper layer play evolutionary games with reinforcement learning models, with dark red indicating cooperative strategies (C) and dark blue defective strategies (D). In the lower network, players playing the traditional PDG choose C, represented by light red, or D, represented by light blue. The dashed lines with arrows indicate the influence of the upper players on the lower players, and the colors are related to the strategies the upper players employ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dilemmas in PDG. Therefore, the return matrix M of the PDG is as follows:

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

where the variables in matrix above are set as $R = 1, S = 0, T = b, P = 0$. The value of b depends the temptation to defect behavior, which also reflects the intensity of social dilemma. The higher the value of b, the stronger the social dilemma. The total payoff P_x ($P_{x'}$) of player x (x') is calculated using Eq. 2, where r_{xy} ($r_{x'y'}$) is obtained from the matrix M, and Ω_x is defined as the group of the player's neighbors.

$$P_x = \sum_{y \in \Omega_x} r_{xy} \tag{2}$$

In the enhancement network, the probability of the player choosing cooperation p_{t-1} in the $t-1$ round is calculated according to the BM model as:

$$p_t = \begin{cases} p_{t-1} + (1 - p_{t-1}) s_{t-1}(a_{t-1} = C, s_{t-1} \geq 0), \\ p_{t-1} + p_{t-1} s_{t-1}(a_{t-1} = C, s_{t-1} < 0), \\ p_{t-1} - p_{t-1} s_{t-1}(a_{t-1} = D, s_{t-1} \geq 0), \\ p_{t-1} - (1 - p_{t-1}) s_{t-1}(a_{t-1} = D, s_{t-1} < 0), \end{cases} \tag{3}$$

where a_{t-1} is the strategy of player in the $t-1$ round. s_{t-1} is the stimulus signal used to drive learning, which is obtained by inputting the activation function in the neural network following the comparison between income and expectation. When s_{t-1} is over zero, it means that the action of player in $t-1$ round has produced satisfactory results. This would result in an increased probability for this player to insist on the original policy by following computation of s_{t-1} as below:

$$s_{t-1} = tanh[\beta(r_{t-1} - A)] \tag{4}$$

where r_{t-1} is the average income of the player in the $t-1$ round. The player plays the game with four neighbor players in turn to obtain the r_{t-1} and A is the player's expected income. If $r_{t-1} - A > 0$, the average income achieves the expectation of the player which means the player is satisfied, otherwise the player is not satisfied when $r_{t-1} - A < 0$. Variable β represents the sensitivity that controls the conversion of the gap between earnings and expectations into a stimulus signal.

To the ordinary player x located in the traditional network, he will firstly calculate the income P_x after playing with the other four players according to the Eq. 2, then compute its fitness F_x according to the following equation:

$$F_x = \alpha \cdot P_{x'} + P_x \tag{5}$$

It can be found that the calculation of fitness F_x introduces $P_{x'}$, the income of the upper corresponding player $x\prime$. The variable α controls the coupling strength between two layers of networks. The larger the value, the greater the impact of guider's income on the ordinary player's fitness, which also means that the traditional layer has more references to the enhancement layer. Players randomly select a neighbor y to determine their strategy by Fermi update formula6 using fitness F_x and F_y.

$$W = \frac{1}{1 + \exp[(F_x - F_y)/K]} \tag{6}$$

where the parameter K can be utilized to measure the noise intensity, as well as the likelihood of irrational behavior and the imitation of a less-well-behaved neighbor.

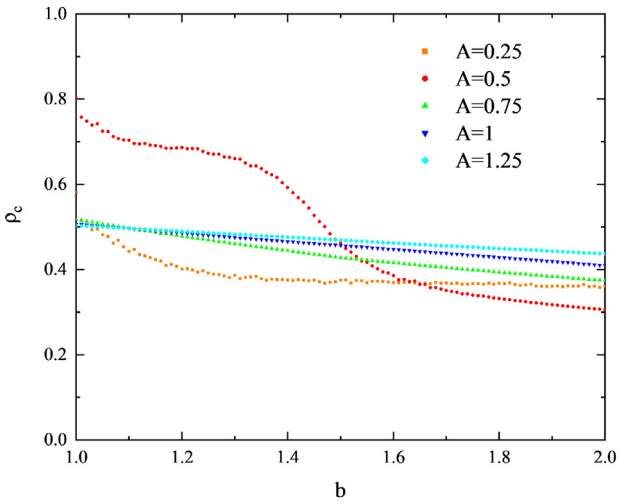

Fig. 2. The medium aspiration value can maintain the cooperative behavior well within a certain range of defective temptation. Relationship between cooperation frequency ρ_c and intensity of social dilemma b, with respect to different aspiration level values A.

In each round of evolution, all the guiders start to evolve independently. Then the ordinary players update the strategy by conditionally taking into the guiders's experience account. That is, the guider is regarded as a reference to the ordinary player but not affected by the ordinary player. For ordinary players in the model, the calculation of (F_x) does not consider the income of the guider $(P_{x'})$ when the fixed number of steps is within the threshold value (age_T) (denoted as the $\alpha = 0$ in Eq. 5). The reference level of ordinary players is controlled by this reference threshold (age_T) and coupling strength (α), which determines the influence of the guiders on ordinary players.

To ensure the stability of the simulation results, we conduct $2 * 10^4$ Monte Carlo steps on the $500 * 500$ lattices. All the results are obtained by averaging the values generated in the stabilization stage (the last 5000 steps) to ensure the accuracy.

3. Results

The evaluation of the proposed work starts from the reinforcement learning based PDG evolution on the single-layer network. The effects of different variation A on the cooperation rate are obtained with different social dilemma intensity, as shown in Fig. 2. It is demonstrated that these cooperators can survive in a relatively intense social dilemma with reinforcement learning. When the intensity of dilemma is low ($b < 1.5$), the moderate value of variation A is most conducive to cooperation. As the intensity of dilemma increases, the change of A value may have complex effect on individuals' behavior. The cooperation rate of the group with higher A value is also relative high in the strict social dilemma($b > 1.5$). According to Eq. 3, the gap between A value and game payoff will affect the possibility of maintaining the previous strategy. When the intensity of dilemma is low, the large gap brought by the high A value leads to the inability of cooperators to maintain their strategy. Once the dilemma intensity becomes high, the higher the A value is, the lower the payoff advantage of the defector will reduce, and finally leads to the increase of the cooperation rate. In following discussion, we denote the individuals who are capable of reinforcement learning as 'more adaptive guiders'. In the society, those guiders are able to provide useful reference for the ordinary players in the traditional PDG.

Next, the reinforcement rules in two-layer network are introduced according to the model part, as well as the effect of parameters on the evolution of the game. Fig. 3 shows the influence of the guiders' aspiration value A on the cooperation rate, the optimal trend appears at $A = 0.5$. This phenomenon is as same as what is in Fig. 2 for a single layer. Considering the enhancement layer network has the advantage of cooperation rate, the ordinary individuals are guided to keep cooperative more effectively. Besides,the red curve in Fig. 3(c) is below the curve in the other two subgraph. This indicates that the aspiration level of guiders has no positive correlation with encouragement effect in the low difficulty dilemma. In other words, a very high expectation of the enhancement layer leads to a decline of the cooperation rate instead, which means that the cooperation rate evolves with the aspiration value A and intensity of social dilemma b as shown in Fig. 4.

For the effect of parameter on the cooperation rate, the variable β represents the sensitivity of the enhancement layer. According to the outputs of single layer with reinforcement learning, three are three representative values of β: 0.1, 0.5 and 1, and Fig. 5 shows that β is positively correlated with the cooperation rate of the enhancement layer. In the traditional network, the proportion of cooperation at $\beta = 0.1$ (Fig. 5(a)) is significantly lower than that of 0.5 and 1 (Fig. 5(b)(c)). Besides, the point where the cooperation rate becomes 0 appears when the value of b is larger($\beta = 1$). These phenomenons indicate that the high sensitivity of the enhancement layer can increase the cooperation rate of traditional layer, and the corresponding setting also favorably supports the ordinary players to resist the strong social difficulties.

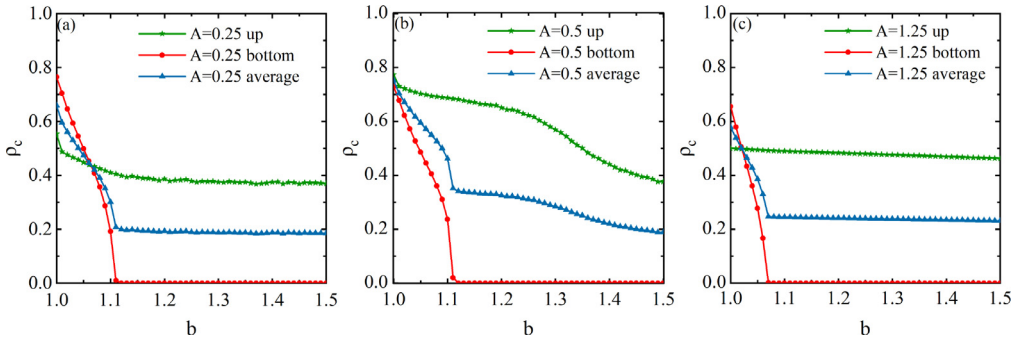

Fig. 3. Relationship between cooperation frequency ρ_c and intensity of social dilemma b on the two-layer network, with respect to different aspiration level values A (a-c). The enhancement layer is located in the upper layer (up), the traditional layer the bottom layer (bottom). The average cooperation frequency of individuals on the two layers is shown by the blue curve.. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

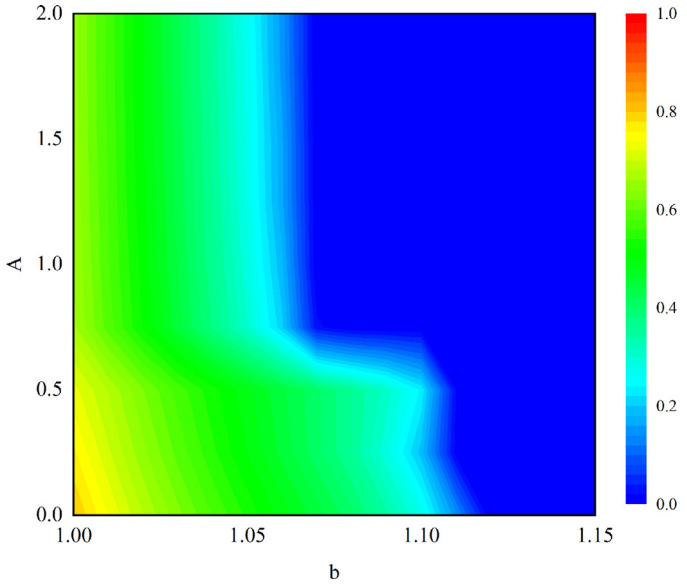

Fig. 4. Relationship between cooperation frequency ρ_c and intensity of social dilemma b on the traditional network, with respect to different aspiration level values A..

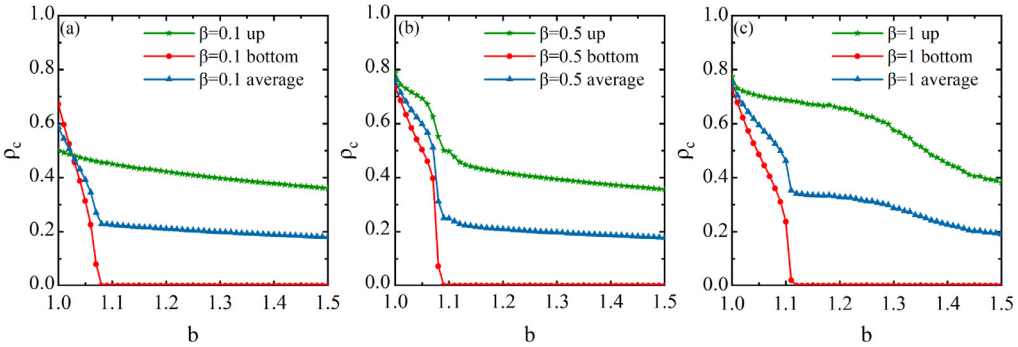

Fig. 5. Relationship between cooperation frequency ρ_c and intensity of social dilemma b on the two-layer network, with respect to different conversion sensitivity value β (a-c). The enhancement layer is located in the upper layer (up), the traditional layer the bottom layer (bottom). The average cooperation frequency of individuals on the two layers is shown by the blue curve.. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

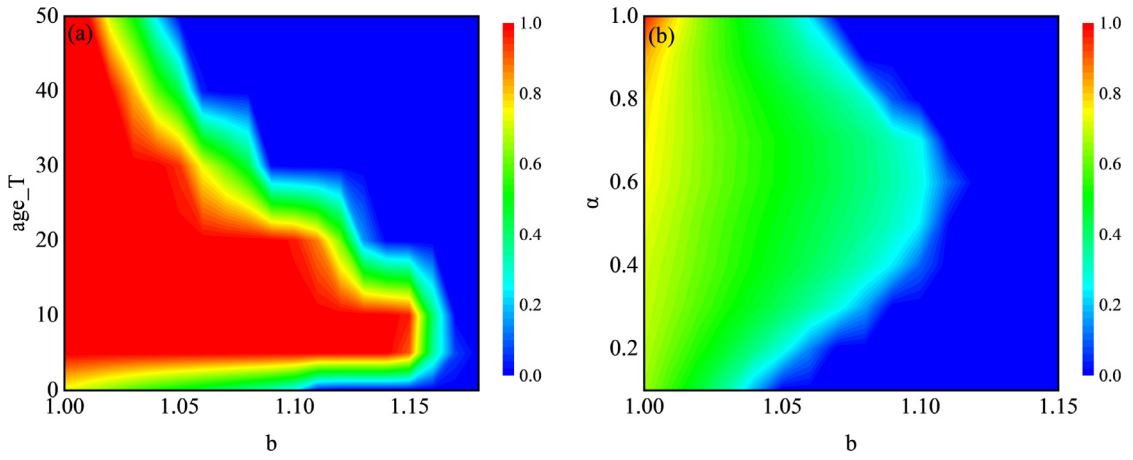

Fig. 6. Relationship between ρ_c and b on the traditional layer, with respect to different reference threshold (age_T) of the traditional players (a), and to different reference strengths (α) of the traditional players (b).

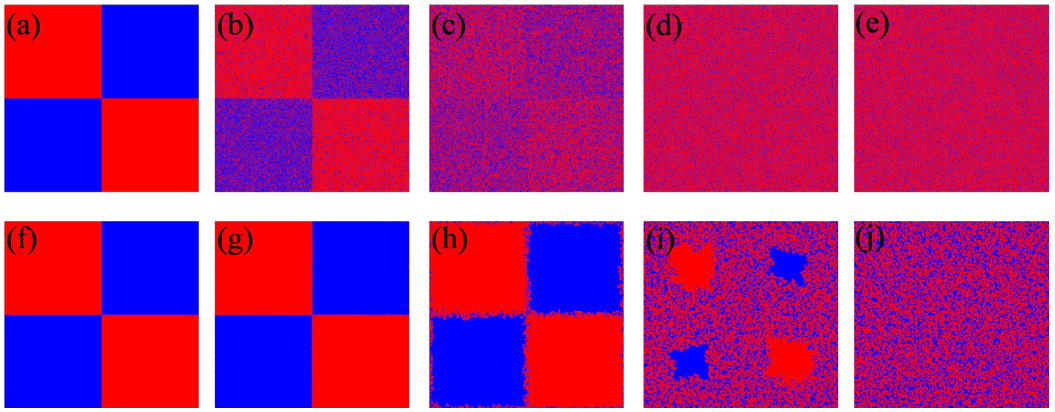

Fig. 7. Snapshots of strategy on the two layers, where red represents C and blue represents D. The snapshots in the first row are the strategy distribution on the enhancement layer, and snapshots in the second row on the traditional layer. The strategy combinations of two layers are consist of CC (the guider chooses C, the ordinary player chooses C) and DD (the guider chooses D, the ordinary player chooses D) at the beginning. From left to right the moment sequence is: step 0, step 1, step 100, step 1000 and step 10000. All results are obtained for $b = 1.03, A = 0.5, \beta = 1, age\_T = 0, \alpha = 0.5$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The reference threshold age_T is designed to control the reference level of the traditional layer, and the guiding effect of the enhancement layer would diminish with the increase of its value. To explain the cooperation rate of the traditional layer in Fig. 6 (a), we use $age\_T = 0$ to represent the income of guiders which is unconditionally introduced in the calculation of the ordinary players's fitness (F_x). Under such circumstances, the cooperation rate is either 0.6 in the lower social dilemma or 0 in the higher social dilemma. As designed in the model part, the value of b denotes the intensity of social dilemma, when it is between 1.05 and 1.15, the cooperation rate reaches 1 firstly then decreases to 0 with the advance of age_T. In other words, a moderate reference threshold value is more conducive to the maintenance of high cooperation rate. If age_T is too high, the reference level of ordinary players would be greatly limited. That will lead to the absence of guiders's participation, and consequence in the rapid disappearing of the cooperative behavior on the traditional layer.

The parameter α controls the unidirectional coupling strength between the two layers. When it becomes larger, the ordinary players refer more to the guiders, and the guiders have a greater impact on the ordinary players' strategic choice. As shown in Fig. 6(b), the optimal value of α is 0.7, which has the best effect of guidance on the cooperative behavior of the ordinary players. This phenomenon reminds us that the reference of the traditional layer to the enhancement layer should be taken moderately. It is noted that the higher coupling strength does not mean the greater support for cooperation. The extreme reference level will lead to the decrease of traditional individuals' resistance ability to social dilemmas.

Fig. 7 and Fig. 8 clearly present the strategy distribution change in the game evolution process. These figures are the strategy point diagrams under two different initial conditions, which can be used to investigate the influence of incorporated enhancement layer on the traditional layer obviously.

Fig. 9 (a) is the time sequence diagram corresponding to the point diagrams Fig. 7 and Fig. 8. The evolution of cooperation rate is reflected by the curve in this figure, which also shows that the proportion of cooperators in the enhancement layer is

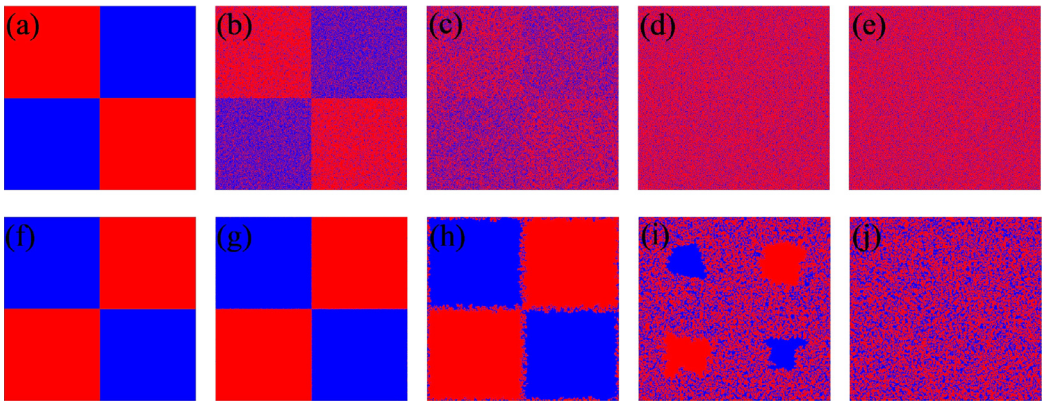

Fig. 8. Snapshots of strategy on the two layers, where red represents C and blue represents D. The snapshots in the first row are the strategy distribution on the enhancement layer, and snapshots in the second row the traditional layer. The strategy combinations of two layers are consist of CD (guiders choose C, the ordinary players choose D) and DC (guiders choose D, the ordinary players choose C) at the beginning. From left to right the moment sequence is: step 0, step 1, step 100, step 1000 and step 10000. All results are obtained for $b = 1.03, A = 0.5, \beta = 1, age\_T = 0, \alpha = 0.5$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

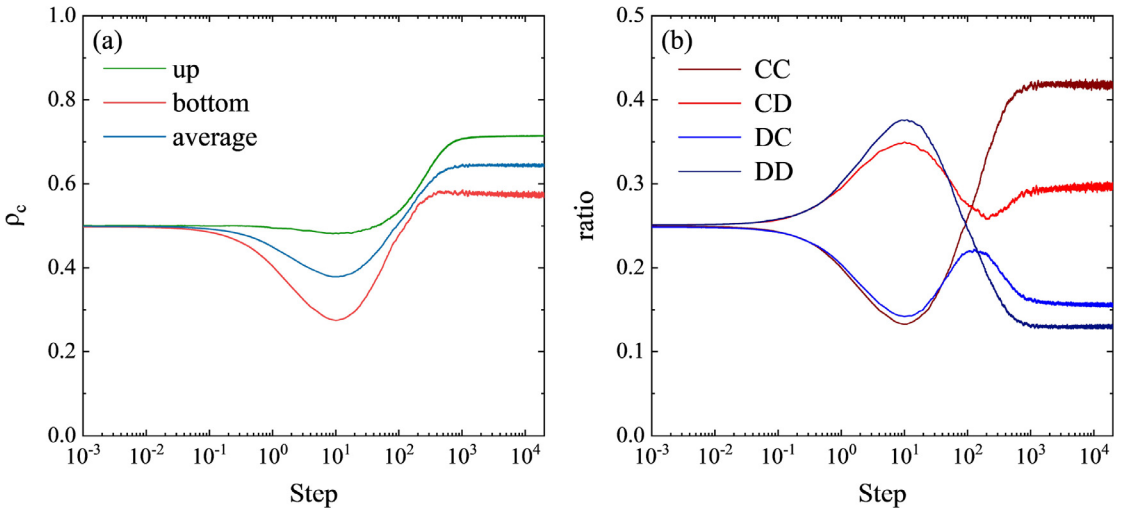

Fig. 9. The evolution of the cooperation rate for different layers (a), the evolution of the ratio of the four strategy combinations on the two layers (b). CC stand for strategy combination which the guider chooses C, the ordinary player chooses D, and the combination CD, DC, DD can explain similarly. The results are obtained under the same conditions as Fig. 8.

always higher than that of the traditional layer. Compared to the drastic descendence in cooperation rate of the traditional layer around the 10 steps, the changing tendency of cooperation rate between two layers basically keep consistent.

In the spot maps, the initial distribution is four clusters with clear boundaries (Fig. 7(a), Fig. 7(f), Fig. 8(a), Fig. 8(f)). On the enhancement layer, the distribution of two strategies achieves relatively homogeneous at 100 steps due to mutual intrusion (Fig. 7(c), Fig. 8(c)). As influenced by the guiders, the ordinary players' strategies at the boundaries become reversed (Fig. 7(h), Fig. 8(h)) and constantly concentrate toward the center (Fig. 7(i), Fig. 8(i)). Eventually on the traditional layer, the homogeneous distribution of the two strategies is presented (Fig. 7(j), Fig. 8(j)).

The subgraphs (h) and (i) of Fig. 7 and Fig. 8 show that the change of clusters' strategy shifts from the boundary to the center on the traditional layer. The reason for this outside-in invasion phenomenon lies in the strategy difference of individual neighbors. As shown in subgraphs (c) and (d) of Fig. 7 and Fig. 8, two strategies on the enhancement layer basically obey the uniform mixed distribution. This indicates that, the income of the guiders $P_{x'}$ in Eq. 5 would not be different for the ordinary individuals' position difference. When the individuals at the boundary of the cluster are surrounded by partial neighbors with opposite strategies, the changing in their income P_x make it different to the individual at the center of the cluster.

According to Eq. 5, the change of the individuals' income P_x may influence the fitness of individuals F_x. Moreover, the variation of F_x increases the changing possibility of individual strategy in the light of Eq. 6. Therefore, the strategy difference

between the surrounding neighbors leads to the active boundary and transmits to the center gradually, which finally form a mixed distribution of the two strategies.

With different conditions in initialization, the equilibrium distribution and stable cooperation rate obtained are consistent. This further verify that in different cases, the guiding mechanism with experience will promote the achievement of cooperative behavior and improve the cooperation rate of game groups, which also indicates that the guiders' boost of cooperative behavior has a strong adaptability.

By analyzing the proportion of different strategy combinations, the influence of guidance on the cooperation rate is obtained from the changing point of strategy combinations. The CC curve in Fig. 9(b) starts to rise around the 10th step, and such a transition is consistent with the cooperation rate in Fig. 9(a). This consistency also validates the pull-up effect of the CC strategy combination on cooperation rate, and demonstrates the effectiveness of guiding on the traditional layer the enhancement layer.

As the overall evaluation, the CC strategy combination with the highest proportion reached 41.9%, which is accounting for 72.9% of the cooperation rate of traditional layer 57.7%. The strategy combination has shown its effectiveness in holding the cooperative rate of the traditional layer, and proves that most ordinary players choose to cooperate because they are consistent with the guide.

4. Conclusion

In view of the low cooperation rate in the traditional single-layer network, this paper proposes a game model referring to experienced guider. Individuals participating in different games are distributed on the double-layer coupled network. In the game evolution, the individuals in the PDG with reinforcement rules guide the individuals in the traditional PDG through the coupling relationship between networks. We study the game evolution by the Monte Carlo simulation method. The results show that this mechanism can effectively improve the adaptability of cooperators and increase the group cooperation rate. For better promoting the development of cooperative behavior, the expectation of the guide should be set to a moderate value. And, the reference intensity of ordinary players should also be maintained at an appropriate level. We hope that the proposed model can bring more enlightenment to the future research on improving cooperation rate, so as to promote the development of social dilemma settlement mechanism.

References

[1] W.W. Powell, D.R. White, K.W. Koput, J. Owen-Smith, Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences1, Am. J. Sociol. 110 (2005) 1132–1205.
[2] C. Darwin, On the origin of species, Soil Sci. 71 (6) (1915).
[3] May, M. Robert, More evolution of cooperation, Nature 327 (6117) (1987) 15–17.
[4] M.A. Nowak, Evolutionary dynamics: Exploring the equations of life, 2006.
[5] Z. Wang, M. Jusup, R.W. Wang, L. Shi, J. Kurths, Onymity promotes cooperation in social dilemma experiments, Sci. Adv. 3 (3) (2017) e1601444.
[6] G. Szabó, G. Fáth, Evolutionary games on graphs, Phys. Rep. 446 (2007) 97–216.
[7] D. Cheng, F. He, H. Qi, T. Xu, Modeling, analysis and control of networked evolutionary games, IEEE Trans. Automat. Contr. 60 (9) (2015) 2402–2415.
[8] M.A. Nowak, C.E. Tarnita, T. Antal, Evolutionary dynamics in structured populations, Philos. Trans. R. Soc. Lond. 365 (1537) (2010) 19–30.
[9] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L.M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: a review, J. R. Soc. Interface (2013).
[10] D. Jia, X. Wang, Z. Song, I. Romić, X. Li, M. Jusup, Z. Wang, Evolutionary dynamics drives role specialization in a community of players, J. R. Soc. Interface 17 (2020).
[11] J.M. Smith, G.R. Price, The logic of animal conflict, Nature 246 (5427) (1973), 15–18.
[12] J.M. Smith, Evolution and the theory of games (1982).
[13] K.M.A. Kabir, J. Tanimoto, Z. Wang, Influence of bolstering network reciprocity in the evolutionary spatial prisoners dilemma game: a perspective, Eur. Phys. J. B 91 (2018) 1–10.
[14] M.A. Nowak, R.M. May, Evolutionary games and spatial chaos, Nature 359 (1992) 826–829.
[15] M.A. Nowak, Five rules for the evolution of cooperation, Science 314 (2006) 1560–1563.
[16] J.M. Smith, Group selection and kin selection, Nature 201 (1964) 1145–1147.
[17] M.A. Nowak, K. Sigmund, Evolution of indirect reciprocity, Nature 437 (2005) 1291–1298.
[18] S. Suzuki, H. Kimura, Indirect reciprocity is sensitive to costs of information transfer, Sci. Rep. 3 (3) (2013) 1435.
[19] X. Li, M. Jusup, Z. Wang, H. Li, L. Shi, B. Podobnik, H.E. Stanley, S. Havlin, S. Boccaletti, Punishment diminishes the benefits of network reciprocity in social dilemma experiments, Proc. Natl. Acad. Sci. U.S.A. 115 (2017) 30–35.
[20] H. Ito, J. Tanimoto, Scaling the phase-planes of social dilemma strengths shows game-class changes in the five rules governing the evolution of cooperation, R. Soc. Open Sci. 5 (2018).
[21] C. Chu, C. Chu, C. Mu, J. Liu, C. Liu, S. Boccaletti, L. Shi, Z. Wang, Aspiration-based coevolution of node weights promotes cooperation in the spatial prisoners dilemma game, New J. Phys. (2019).
[22] F. Fu, C. Hauert, M.A. Nowak, L. Wang, Reputation-based partner choice promotes cooperation in social networks, Phys. Rev. E Stat. Nonlin Soft. Matter Phys. 78 2 Pt 2 (2008) 026117.
[23] G. Szabó, C. Hauert, Evolutionary prisoner's dilemma games with voluntary participation, Phys. Rev. E Stat. Nonlin. Soft Matter. Phys. 66 6 Pt 1 (2002) 062903.
[24] J. Zhang, W.-Y. Wang, W.-B. Du, X. Cao, Evolution of cooperation among mobile agents with heterogenous view radii, Physica A (2011).
[25] M.R. Arefin, J. Tanimoto, Evolution of cooperation in social dilemmas under the coexistence of aspiration and imitation mechanisms, Phys. Rev. E 102 3-1 (2020) 032120.
[26] K.M.A. Kabir, J. Tanimoto, The role of pairwise nonlinear evolutionary dynamics in the rock-paper-scissors game with noise, Appl. Math. Comput. 394 (2021) 125767.
[27] P. Zhu, X. Wang, D. Jia, Y. Guo, Sh. Li, C. Chu, Investigating the co-evolution of node reputation and edge-strategy in prisoner's dilemma game, Appl. Math. Comput. 386 (2020).
[28] H. Ito, J. Tanimoto, Dynamic utility: the sixth reciprocity mechanism for the evolution of cooperation, R. Soc. Open Sci. 7 (2020).
[29] F.C. Santos, J.M. Pacheco, Scale-free networks provide a unifying framework for the emergence of cooperation, Phys. Rev. Lett. 95 (9) (2005) 098104.

[30] F.C. Santos, J.M. Pacheco, T. Lenaerts, Evolutionary dynamics of social dilemmas in structured heterogeneous populations, Proc. Natl. Acad. Sci. 103 (9) (2006) 3490–3494.
[31] C.Y. Xia, X.K. Meng, W. Zhen, L.L. Jiang, Heterogeneous coupling between interdependent lattices promotes the cooperation in the prisoner's dilemma game, PLoS ONE 10 (6) (2015) e0129542.
[32] X.-K. Meng, S. Sun, X. Li, L. Wang, C. Xia, J. Sun, Interdependency enriches the spatial reciprocity in prisoners dilemma game on weighted networks, Phys. A-stat. Mech. Appl. 442 (2016) 388–396.
[33] R.S. Sutton, A.G. Barto, Reinforcement learning: an introduction, IEEE Trans. Neural Netw. 9 (5) (1998) 1054.
[34] X. Wang, S. Gao, P. Zhu, J. Wang, Roles of different update strategies in the vaccination behavior on two-layered networks, Physics Letters A 384 (2020).
[35] D. Jia, T. Li, Y. Zhao, X. Zhang, Z. Wang, Empty nodes affect conditional cooperation under reinforcement learning, Appl. Math. Comput. 413 (2022) 126658.
[36] D. Jia, H. Guo, Z. Song, L. Shi, X. Deng, M. Perc, Z. Wang, Local and global stimuli in reinforcement learning, New J. Phys. 23 (8) (2021) 083020, doi:10.1088/1367-2630/ac170a.
[37] N. Masuda, M. Nakamura, Numerical analysis of a reinforcement learning model with the dynamic aspiration level in the iterated prisoner's dilemma, J. Theor. Biol. 278 (1) (2011) 55–62.
[38] T. Ezaki, Y. Horita, M. Takezawa, N. Masuda, Reinforcement learning explains conditional cooperation and its moody cousin, PLoS Comput. Biol. 12 (7) (2016) e1005034.
[39] G. Cimini, A. Sánchez, Learning dynamics explains human behaviour in prisoner's dilemma on networks, J. R. Soc. Interface 11 (94) (2014).
[40] C. Lanave, G. Preparata, C. Sacone, G. Serio, A new method for calculating evolutionary substitution rates, J. Mol. Evol. 20 (1) (1984) 86–93.
[41] A.W. Kemp, B.F.J. Manly, Randomization, bootstrap and monte carlo methods in biology, Biometrics 53 (1997) 1560.
[42] U. Alvarez-Rodriguez, F. Battiston, G.F. de Arruda, Y. Moreno, M. Perc, V. Latora, Evolutionary dynamics of higher-order interactions in social networks, Nat. Hum. Behav. (2021) 1–10.
[43] J. De Freitas, K.A. Thomas, P. DeScioli, S. Pinker, Common knowledge, coordination, and strategic mentalizing in human social life, Proc. Natl. Acad. Sci. U.S.A. 116 (2019) 13751–13758.
[44] J. Tanimoto, H. Sagara, Relationship between dilemma occurrence and the existence of a weakly dominant strategy in a two-player symmetric game, BioSystems 90 (1) (2007) 105–114.
[45] Z. Wang, S. Kokubo, M. Jusup, J. Tanimoto, Universal scaling for the dilemma strength in evolutionary games, Phys. Life Rev. 14 (2015) 1–30.
[46] H. Ito, J. Tanimoto, Scaling the phase-planes of social dilemma strengths shows game-class changes in the five rules governing the evolution of cooperation, R. Soc. Open Sci. 5 (2018).
[47] M.R. Arefin, K.M.A. Kabir, M. Jusup, H. Ito, J. Tanimoto, Social efficiency deficit deciphers social dilemmas, Sci. Rep. 10 (2020).
[48] J. Tanimoto, Sociophysics Approach to Epidemics, 2021. doi:10.1007/978-981-33-6481-3.