Analysis of the Causes and Evolution of Class Solidification Based on Q-Learning and Hawk-Dove Game

Xuan Zhou[*1], Yicheng Gong [1,2], Yuqiang Feng[1], Ningjing Yang[1]

1School of Science, Wuhan University of Science and Technology, Wuhan 430065, China;
2Hubei Province Key Laboratory of Systems Science in Metallurgical Process, Wuhan 430065, China

Abstract

If the phenomenon of class solidification intensifies, it will lead to the lack of impetus for social production. Class solidification can be understood as the situation of low mobility formed by the interaction of aggressive and concession strategies among classes of different classes, and this can be regarded as a kind of equilibrium of repeated hawk-dove game among classes. To find measures to promote mobility between classes, firstly, a repeated class competition game model is constructed and theoretically analyzed.Then, the Q-learning method is used to simulate the characteristics of limited rational classes seeking the best decision through trial and error in reality, to test the explanatory power of the model.And the simulation experiment get the conclusion that the mobility of classes is inversely proportional to the strength ratio between classes and the income generated.Finally,the model is improved by transforming the benefits of both parties into their own strength, the experiment shows that the class mobility has increased by 11.6% compared to before the improvement.

Keywords

class solidification; hawk-dove game; reinforcement learning; Q-learning

1. Introduction

Class solidification has always been one of the continuous topics in the study of social stratification. Without class mobility, society can't form an effective incentive mechanism to make individual members believe that their social and economic status can be improved through their own efforts. In the long run, the lack of class mobility will hinder the sustainable development of society [1-2].Although there are few empirical studies on class solidification with a long time span, it is found that people's attention to class solidification is closely related to the degree of social stability [3-4].

Asymmetric Hawk-Dove Game is often used to study the strategic choice and balance of strengths and weaknesses among classes with unequal strengths, and its theoretical basis is gradually developing. Song Bo and Huang Jing established an hawk-dove game model from the perspective of asymmetric cooperation when studying the stability analysis of strategic alliances[5], and obtained its Nash equilibrium solution of mixed strategy. At the same time,Bu Zhenxing interpreted the Sino-Japanese relationship by using the infinite game model of hawk and dove[6], and put forward that self-strength, competition for interests, etc. are all important factors that affect the game strategy. At present, the research on repeated hawk-dove game generally adopts the evolutionary game perspective, which is based on the fact that the strength of both players in the game is asymmetric and fixed, which is in line with the actual situation of class solidification. However, there is no good starting point for the changing strength of the two sides of the game.

With the continuous development of reinforcement learning, the theoretical model of reinforcement learning model is basically consistent with the framework of game theory. Masuda & Nakamura et al.under the background of iterative prisoner's dilemma, conducted a numerical test on the performance of reinforcement learning model, and explored the relationship between learning and

evolution through it[7]. Q-learning is the most commonly used RL algorithm when dealing with two-party game problems. It can be used online without the model of its environment, and it is very suitable for repeated games with unknown opponents[8] . Zhang Chunyang et al.realized the Q-learning algorithm in the prisoner's dilemma, and compared the Q-learning method with the algorithm[9]; Liu Weibing, et al.combined reinforcement learning with evolutionary game, and established a multi-agent reinforcement learning model in evolutionary game[10]. The simulation results show that the multi-agent reinforcement learning model can make the players keep learning and seek the best strategy.

In this paper, Q-learning in reinforcement learning and asymmetric repeated hawk-dove game model are combined to model, and the evolution and causes of this phenomenon are discussed by simulating the changes of social classes' strategy choices from the perspective of class solidification, and some effective suggestions are obtained through simulation exploration.

2. Main framework

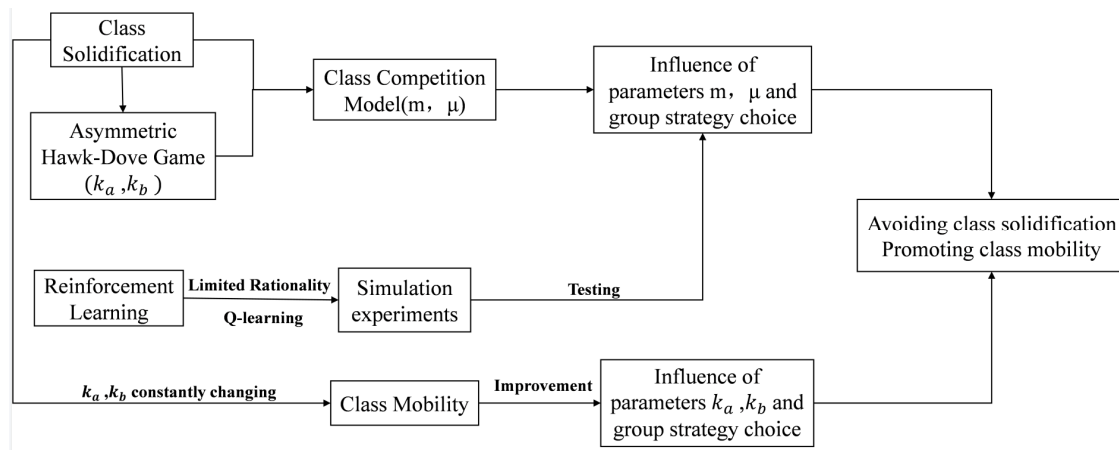The research framework of this paper is shown in Figure 1：

Figure 1 main framework flow chart

According to Figure 1, the thesis is divided into three parts:

The first part abstracts the phenomenon of class solidification as a less mobile equilibrium formed by the constant strategic interaction of different classes in social practice, while different social classes choose aggressive and concessionary strategies in line with the characteristics of the hawk-dove game.Thus, an asymmetric repeated hawk-dove game model with parameters is constructed, where the parameters are the asymmetry factor $\mu = k_a : k_b$ and the unit benefit of the conflict $m = V/C$ between the parties to the game (where V is the benefit obtained and C is the cost paid by both parties when the conflict arises),and named the class competition game model.Meanwhile, the theoretical solution of the game model is obtained.

In the second part, under the premise of limited rationality of players, the model is simulated through reinforcement learning to test the explanatory power of game theory in practice, and to explore the relationship between different types of strategy choices and parameters.

The third part improves and simulates the parameters of the game model of class competition, compares the experimental results before and after the improvement, and proposes solutions to improve the solidification of the class and realize the fluidity of the class by revealing the phenomenon of class mobility.

3. Analysis of asymmetric repeated class competition game based on hawk-dove game

3.1 An asymmetric class competition model based on the hawk-dove game

The construction of the class competitive game model consists of three parts：
(1) Strategy space of game subjects: assume that both social class A and social class B have two strategies to choose from, i.e., the hawk strategy and the dove strategy；
(2) The parameters of the game model are generally the benefits of the game V, the costs paid by both parties C when a conflict arises (in general $V < C$)；
(3) The asymmetry factor $\mu = k_a : k_b$ between the two sides of the game (here $k_a + k_b = 1$, the resources are allocated by social class A and social class B).

Table 1 Class competition model revenue matrix

		Social class B	
		H	**D**
	H	$(V-C)/4k_a , (V-C)/4k_b$	V,0
Social class A	**D**	0,V	$k_a V , k_b V$

Using the line drawing method can get Table 1 there are two pure strategy Nash equilibrium: (H, D) and (D, H). And there is also a mixed strategy Nash equilibrium, the probability p_0 is the probability that the social class chooses the dove strategy

$$(p_0, 1-p_0) = \left(\frac{(C-V)}{(C-V)+4k_a k_b V}, 1 - \frac{(C-V)}{(C-V)+4k_a k_b V} \right) \tag{1}$$

For the parameters of the above asymmetric hawk-dove game $p_0 = \dfrac{(C-V)}{(C-V)+4k_a k_b V}$, defined as the

unit gain of the conflicting parties $m = V/C$ to the game, then

$$p_0 = \frac{(1-m)(1+\mu)^2}{(1+\mu)^2 - m(\mu-1)^2} \tag{2}$$

From Eq.2, it can be seen that the probability p_0 is related to the parameters μ and m: when μ is constant, the larger m, the smaller p_0, i.e. the larger the unit gain at the time of conflict, the smaller the probability of choosing the dove strategy；When m is constant, the greater μ, the greater p_0, the p_0 is minimized when $\mu = 1$, i.e. the probability of choosing the dove strategy is the smallest for classes of equal strength, the greater the difference in strength, the greater the probability of choosing the dove strategy.

3.2 A class competition model based on Q-learning

The purpose of this paper is to study the equilibrium reached by social classes with different power ratios after a long and repeated game process, so the environment of the Q-learning method for simulating game experiments is set as follows：
(1) States $S_t = \{"hh", "hd", "dh", "dd"\}$, using the strategy choices of the previous round of social classes A and B as the state of this round, e.g. "hh"；
(2) Social class A and social class B have the same action set A, set to $A = \{h, d\}$, i.e. both A and B have two actions to choose from: the "hawk" strategy (h) and the "dove" strategy (d)；

(3)　　The reward of the model R, representing the reward of social class A R_a and the reward of social class B R_b after adopting different actions , where the reward set R_a denoted is $\{(V - C/4k_a^t), V, 0, k_a^t V\}$; the reward set R_b denoted is $\{(V - C/4k_b^t), V, 0, k_b^t V\}$;

(4)　　Transfer probability of the model P: $p = p(s', r|s, a) = 1$;

(5)　　Strategy $P = P(A_t|S_t)$: Repetition game is not just a simple repetition game, we give the social class to choose a certain randomness of strategy, so choose the Boltzmann distribution to represent the probability of the social class to choose the action: $p(\alpha_i) = \dfrac{e^{Q(s,\alpha_i)/\lambda}}{\sum_{\alpha \in A} e^{Q(s,\alpha_i)/\lambda}}$;

This gives the Q-value update formula:

$$Q_{t+1}^A(s, a_1, a_2) = (1 - \alpha_t) Q_t^A(s, a_1, a_2) + \alpha_t \left(r_t^A + \gamma \max_b Q_t^A(s', b) \right) \tag{3}$$

$$Q_{t+1}^B(s, a_1, a_2) = (1 - \alpha_t) Q_t^B(s, a_1, a_2) + \alpha_t \left(r_t^B + \gamma \max_b Q_t^B(s', b) \right) \tag{4}$$

4. Simulation experiments on class competition model based on Q-learning

In this section, the laws of social class strategy selection are discovered through simulation experiments.

4.1 Parameter Setting

The parameters in the Q-learning model in this paper are 0.9 for the γ and 0.1 for the α, considering the realistic implications, the following two parameters are set:

(1)　　The unit benefit of conflict between the two sides of the game $m = V/C$, where the conflict cost C is set to 50 and the benefit is set to 20;

(2)　　Asymmetry factor of both sides of social classes μ: According to the difference of strength values of different classes, the simulation experiment is divided into 5 cases: $k_a : k_b = 0.5 : 0.5$, $k_a : k_b = 0.4 : 0.6$, $k_a : k_b = 0.3 : 0.7$, $k_a : k_b = 0.2 : 0.8$, $k_a : k_b = 0.1 : 0.9$. Considering the stability of the model, the initial total value of the strength of both sides is set to 10000, the initial strength ratio is proportionally distributed, and the asymmetry factors μ of class A and class B do not change in each experiment.

4.2 A test of the explanatory power of game theory in practice

After 20,000 iterations of the model, it is obtained that different classes are stable in the combination of (H, D), (D, H) and (D, D) strategies at about 10,000 iterations, but the probability of choosing these three strategies combinations is related to parameters μ and m. Therefore, the control variables method is used to verify the frequency change of different social classes when the strength ratio μ is different but m unchanged (taken $m = 20/50$) versus when the strength ratio m is different but μ unchanged (taken $\mu = 0.5 : 0.5$), and the model is repeated 100 times to count the frequency change when the strategies are stable to obtain Figure 2:
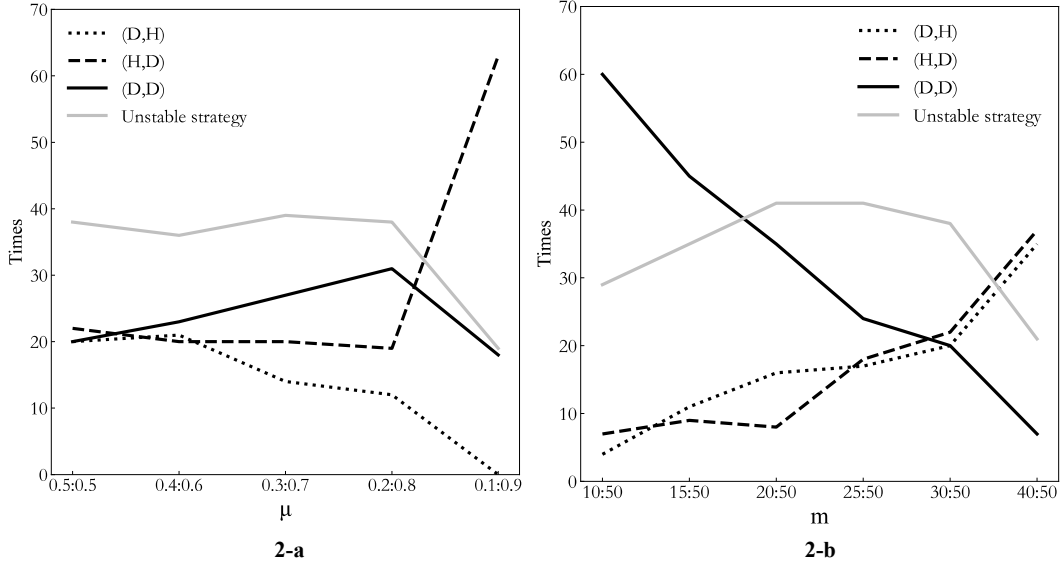
Figure 2 Plot of the variation of strategy with μ and m

From the Figure 2-a, it can be seen that with the ratio $\mu = k_a : k_b$ decreasing, that is, the strength gap between the class classes becomes smaller, the more the two sides tend to choose (D, D) strategy combination, and the probability of choosing (D, H) strategy combination and (H, D) strategy combination gradually decreases, which accords with the results given by theoretical values.Also, the experimental result means that the established Q-learning game model can basically represent the real game situation. However, when $\mu = k_a : k_b = 0.1 : 0.9$,the weak side will always choose to compromise and give in, the strong side will always choose the conflict strategy, and the probability of cooperation between the two sides is reduced.

From the 2-b diagram, we can see that as the gain of the game V becomes larger and m increases, the unit gain of conflict between the two sides increases, the chance of both sides choosing the (D,D) strategy combination becomes smaller, the choice of (D,H) and (H,D) in general becomes more and more, and the chance of both social classes choosing to settle for the status quo becomes smaller.

5. Simulation experiments on class competition model with variable parameters based on Q-learning

5.1 Parameter Setting

This section considers breaking class solidification under the premise of class solidification. The strength of different social classes in the game process will change because of some factors, which can be internal or external. The model sets the cumulative income after the game as the strength ratio of both sides after the game: i.e., the strength value k_a^t of social class A and the strength value k_b^t of

social class B are determined by the cumulative gains of the class：$k_a^t = \dfrac{R_a^{t-1}}{R_a^{t-1} + R_b^{t-1}}$, $k_b^t = \dfrac{R_b^{t-1}}{R_a^{t-1} + R_b^{t-1}}$,

the rest of the parameter settings are the same as those used for class solidification, defined as a class competition model with variable parameters.

5.2 Comparison of experimental results of class competition model simulation before and after parameter change

The strategies of both sides of the society are Q-learning with 20,000 iterations. During the experiment, it was found that after about 10,000 games, both sides of the game are stable in the

combination of three strategies (H, D), (D, H) and (D,D), the probability of the combination of the three strategies is also related to the parameters μ and m. At the same time, there are obvious differences in the choice of strategy combination between participants before and after the model improvement. The two models are simulated for 100 times, and the probability of test results (D, D) is compared, as shown in Figure 3:
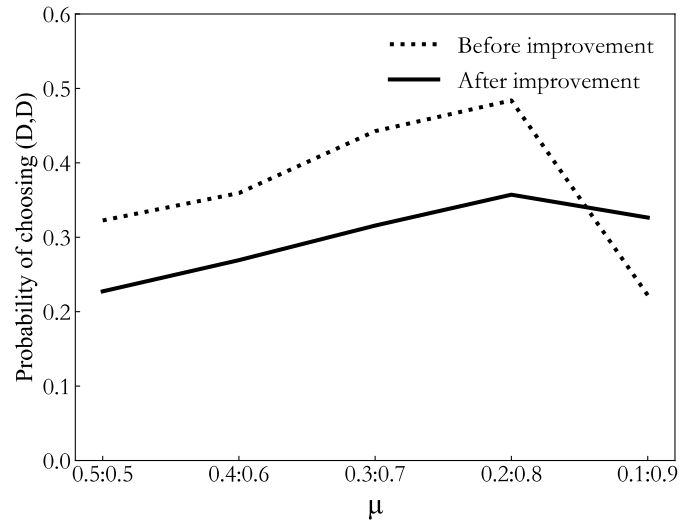

Fig. 3 Plot of the probability of choosing the (D,D) strategy combination with μ value for the social class before and after parameter improvement

It can be seen from Figure 3 that as the value or initial value μ keeps getting smaller, i.e., as the class strength gap gets bigger, the probability of the class choosing (D,D) will gradually increase, representing that the bigger the class gap is, the easier it is for the class class to lie flat. Particularly, when the class classes have different strengths $(\mu = 0.1 : 0.9)$, the lower class classes when the μ value remains unchanged basically do not choose the aggressive aggressive strategy, but choose to lie flat, while about 26% of the lower class classes when the value changes are willing to choose the aggressive aggressive strategy.

The difference is that the improved variable-parameter class competition model, on the whole, reduces the probability of the class choosing (D,D) to 11.6% compared with the class competition model under the phenomenon of class solidification , while the probability of choosing (H,D) and (D,H) has an average growth rate of 6.3% and 4.6%. It can be seen that the immutability of class status makes the disadvantaged classes more inclined to lie flat, while the variability of class status effectively stimulates the motivation of different class classes, constantly promotes the mobility between social classes and improves the dynamics of social development.

6. Conclusion and Outlook

6.1 Conclusion

In this paper, the asymmetric repeated hawk-dove game is used to model and theoretically analyze the class competition phenomenon, and the simulation experiments based on Q-learning algorithm show that: different class classes are likely to stabilize in either aggressive or comfortable strategies, but the parameters μ and m are roughly inversely proportional to the probability of class choice (D, D), and the model can be improved by changing the values of k_a and k_b to reduce the probability of class choice (D,D) to 11.6%. Finally, based on the results of the simulation experiment, this paper discusses three suggestions to avoid the phenomenon of class solidification and improve social mobility：

(1)Reduce the gap between rich and poor. In this study, we found that the probability of both sides of the class being stable at (D,D) is inversely proportional to the strength ratio μ of both sides. The

greater the strength ratio μ, the greater the probability of choosing to rest on one's laurels in social competition, and the probability of choosing to forge ahead is generally decreasing.Meanwhile,when the difference in strength between the two sides of the class is too great, the side with low strength will hardly adopt the strategy of forging ahead but lie flat to face challenges and opportunities, which will increase the degree of class solidification. Therefore, in order to avoid the aggravation of class solidification, it is necessary to promote measures to reduce the gap between the rich and the poor, so as to increase the motivation of different classes to advance up the class；

(2) Appropriate load shedding mechanisms. Studies have shown that the probability of both classes settling at (D,D) is also inversely proportional to the unit benefit m of conflict between the two parties. The increase in the unit gain m of conflict between the two sides of the class can be interpreted as the gain obtained by the class becomes larger under the condition of a certain conflict cost, when the probability of both sides of the class settling in the status quo is reduced, effectively increasing the motivation of the class to continuously improve in social competition. When the gain cost is higher, the class will be more willing to take the initiative to improve themselves under the same conflict cost condition, instead of losing motivation in the continuous internal conflict. Therefore, in order to avoid the phenomenon of class solidification, it is necessary to appropriately increase the unit gain in the conflict to encourage people of different classes to continuously advance and achieve a greater possibility of class leap；

(3) A fair and just social environment. The study shows that the changing k_a, k_b. that translates gains into their own strength values effectively reduces the probability of both sides of the game choosing to settle for the status quo by 11.6% on average, and increases the probability of choosing to be aggressive by 11%. This indicates that social classes find that their own strength values can be improved through acquired effort factors in continuous strategic interaction, which increases the motivation of classes to change themselves. Therefore, establishing a fair and just social environment for different social classes to establish a reasonable strategic interaction environment, so that the interests of both sides can be effectively protected, is an effective way to avoid the phenomenon of social solidification and improve social mobility.

7. Acknowledgments

8. References

[1] Jaafar Alloul, 'Traveling habitus' and the new anthropology of class: proposing a transitive tool for analyzing social mobility in global migration,Mobilities, 16(2):178-193(2021).
[2] Yang X , Zhou P . Wealth Inequality and Social Mobility: A Simulation-Based Modelling Approach[J]. Cardiff Economics Working Papers,196:307-329(2022).
[3] Acemoglu D,Egorov G,Sonin K.Social Mobility and Stability of Democracy: Re-evaluating De Tocqueville[J].CEPR Discussion Papers,133(2):1041-1105(2016).
[4] Lipset S M.Political Man:The Social Bases of Politics[J].Political Science Quarterly,75(2):326-328(1960).
[5] Song bo, Huang Jing. Stability analysis of strategic alliance from the perspective of asymmetric cooperation-based on the game model of hawk and dove [J]. Soft Science,27(2):28-31(2013).
[6] Bu Zhenxing. Using infinite game model of hawk and dove to study international relations-taking Sino-Japanese relations as an example [J]. Journal of Sichuan Provincial Party School of CPC,(01):99-104(2016).
[7] Masuda N,Nakamura M.Numerical analysis of a reinforcement learning model with the dynamic aspiration level in the iterated Prisoner's dilemma[J]. Journal of Theoretical Biology, 278(1):55-62.(2011)
[8] Levy N,Klein I,Ben-Elia E. Emergence of cooperation and a fair system optimum in road

networks:A game-theoretic and agent-based modelling approach[J].Research in Transportation Economics,(68):46-55(2018).

[9] Zhang Chunyang, Chen Xiaoping, Liu Guiquan, Cai Qingsheng. Q-learning algorithm and its implementation in prisoner's dilemma [J]. Computer engineering and application, (13):121-122+128(2001).

[10] Liu Weibing, Wang Xianjia. Multi-agent reinforcement learning model in evolutionary game [J]. System engineering theory and practice,29(03):28-33(2009).