A Novel Dynamic Decision Model in 2-player Symmetric Repeated Games

Liu Weibing[1], Wang Xianjia[1], Wang Guangmin[2]
(1. Institute of Systems Engineering, Wuhan University, Wuhan 430072, China;
2. School of Management, China University Of Geosciences, Wuhan 430074, China)

Abstract: Considering the dynamic character of repeated games and Markov process, this paper presented a novel dynamic decision model for symmetric repeated games. In this model, players' actions were mapped to a Markov decision process with payoffs, and the Boltzmann distribution was introduced. Our dynamic model is different from others', we used this dynamic model to study the iterated prisoner's dilemma, and the results show that this decision model can successfully be used in symmetric repeated games and has an ability of adaptive learning.
Key words: game theory; evolutionary game; repeated game; Markov process; decision model

1 Introduction

In 1970s the fundamental notion of evolutionary game theory was introduced by Maynard Smith. In his book *Evolution and Theory of Games* [1] he presented an evolutionary approach in classical game theory, and he gave the definition of evolutionary stable strategy (ESS) with Price[2]. Evolutionary game theory has extensive applications in many fields, such as mathematics, biology, ecology, economics and sociology.

Contrary to classical game theory, in evolutionary game theory the individuals of a population are not assumed to act consciously and rationally. So the theoretic frame of classical game theory and the notion of equilibrium can not be adapted to evolutionary game theory. The last decade has witnessed a number of publications on the building of models and analytic methods for evolutionary and repeated games. Yao and Darwen presented a method that introduced genetic algorithm into evolutionary games (iterated prisoner's dilemma) [3]. Yuce Thlol and Adnan Acan investigated an ant colony optimization approach for iterated prisoner's dilemma [4]. This method provided game strategies of better quality than genetic algorithms, but needed longer running times. Takafumi Kanazawa and Toshimitsu Ushio gave a multi-population replicator dynamic with erroneous perceptions [5].

During last decades there are many scholars who used stochastic process in evolutionary and repeated games. Amir et al. proposed a dynamic model for symmetric games using birth and death process [6]. Tadj and Touzene extended the work of Amir et al. and gave a QBD approach for evolutionary games (non-symmetric and symmetric games) [7]. Recently the Moran process was successfully introduced into evolutionary games by Nowak and Fudenberg and a new dynamic model was presented [8].

In this paper, we built a new dynamic model for evolutionary and repeated games using the Markov decision process. In this model, the Boltzmann distribution was introduced. We used the dynamic model to study the iterated prisoner's dilemma, and the results show that the model can successfully simulate the actions of players and has the ability of adaptive learning.

2 Some Preliminaries

2.1 Markov Process

Supposing . $I=\{0,1,2\cdots\}$ is the space of state, $T=\{0,1,2,\cdots\}$ is a set of time. A stochastic process $\{X_t, t\in T\}$ is called Markov if for arbitrary and state $i_0, i_1, \cdots, i_{n-1}, i_n, j$, we have

$$\begin{aligned} &P\{X_{t+1}=j \mid X_t=i, X_{t-1}=i_{n-1}, \cdots, \\ &X(1)=i_1, X(0)=i_0\} \\ &=P\{X_{t+1}=j \mid X_t=i\} \end{aligned}$$

i. e. , the Markov process whose future probabilities are determined by its most recent values.

The conditional probability

$$P_{ij}(1)=P\{X_{t+1}=j \mid X_t=i\}$$

is called one-step transition probability. It means the probability of state transition from i to j, so $P_{ij}(1)$ satisfies the following:

Received 29 *January* 2007

$P_{ij}(1) \geqslant 0 \ i,j \in I;$

$\sum_j P_{ij}(1) = 1 \ i \in I.$

Therefore, we have the one-step Markov transition probability matrix

$$\boldsymbol{P}_1 = \begin{bmatrix} P_{00}(1) & P_{01}(1) & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ P_{n0}(1) & P_{n1}(1) & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

2.2 Game Theory

We restrict our view to the class of finite games in strategic form. Generally a normal form game consists of three key components:

1) Players: Let $I = \{1,2,\cdots,n\}$ be a set of players, where n is a positive integer.

2) Strategies space: For each player $i \in I$, let S_i denote a set of allowable actions, called the pure strategies set. The choice of a specific action $s_i \in S_i$ of a player $i \in I$ is called a pure strategy. The vector $\boldsymbol{s} = (s_1, s_2, \cdots, s_n)$ is called a pure strategies profile.

3) Payoff function: For any strategies profile $\boldsymbol{s}$ and any player $i \in I$, let $u_i(s_1, \cdots, s_n)$ be the payoff to player .

Evolutionary game theory is the extension of classical game theory, and evolutionary game is a dynamic game. The evolutionary stable strategy (ESS) is the key notion in evolutionary game theory. $(s_i^* \cdots, s_n^*)$ is called an ESS, if it adheres to the following constraints:

1. $u_i(s_1^*, \cdots s_{i-1}^*, s_i^*, s_{i+1}^*, \cdots s_n^*) \geqslant u_i(s_1^*, \cdots, s_{i-1}^*, s_i, s_{i+1}^*, \cdots, s_n^*)$

2. If $u_i(s_1^*, \cdots s_{i-1}^*, s_i^*, s_{i+1}^*, \cdots s_n^*) = u_i(s_1^*, \cdots, s_{i-1}^*, s_{ij}, s_{i+1}^*, \cdots, s_n^*)$, then $u_i(s_1, \cdots s_{i-1}, s_i^*, s_{i+1}, \cdots s_n) \geqslant u_i(s_1, \cdots, s_{i-1}, s_i, s_{i+1}, \cdots, s_n)$.

3 Dynamic Decision Model

3.1 The Idea of Model

Our research field focuses on the repeated games. In repeated games, players update strategies by their payoffs. In this paper, we regarded iteration times of games as a set of time $t = 1,2,\cdots$, in Markov process $\{X(t), t \in T\}$, strategy space was mapped to the state space I of Markov process, and the payoffs in repeated games were regarded as rewards for state transition. Therefore, the repeated games were mapped to a Markov process with rewards. For example, the iterated prisoner's dilemma assumes a binary choice for the players, either cooperation (C) or defection (D). So the iterated prisoner's dilemma was looked as a Markov process with two-state transition. More generally, if at time t (the t^{th} repeated game) the player is in state i ($i = C, D$)). Then at time $t+1$ the probability that the player chooses strategy C or D can be obtained by Markov transition probability matrix (Fig. 1). Where $p_{ij} = P\{X_{t+1} = j | X_t = i\}$, $i,j = C$ or D.

$$\begin{array}{ccc} & & X_{t+1} \\ & & \begin{array}{cc} \mathrm{D} & \mathrm{C} \end{array} \\ X_t \quad \mathrm{D} & & \begin{bmatrix} p_{\mathrm{DD}} & p_{\mathrm{DC}} \\ p_{\mathrm{CD}} & p_{\mathrm{CC}} \end{bmatrix} \\ \quad \mathrm{C} & & \end{array}$$

Fig. 1 The Markov process for iterated prisoner's dilemma

3.2 The Decision Model in Symmetric Repeated Games

As a Markov process with Nstates, supposing u_{ij} is the payoff when the state transferred from i to j, u_{ij} can also be regarded as a reward for state transition. With the transition of states, the system will generate a series of rewards. So that the rewards u_{ij} are stochastic variables, and its probability distribution is determined by Markov transition probability matrix.

For 2-players symmetric repeated games, supposing now the player is in state i(namely the player's strategy is i), we can get the total expected payoff after n times' repeated games using Markov process. Firstly, we assume $v_i(t)$ is the total expected payoff after times' transition from state . Easily, we can get the following equation:

$$v_i(t) = \sum_{j=1}^{N} p_{ij}[u_{ij} + v_j(t-1)]$$
$$(i = 1,2,\cdots,N; t = 1,2,3,\cdots) \quad (1)$$

equation(1)can be simplified as follows:

$$v_i(t) = \sum_{j=1}^{N} p_{ij}u_{ij} + \sum_{j=1}^{N} p_{ij}v_j(t-1) \quad (2)$$

If define $$q_i = \sum_{j=1}^{N} p_{ij}u_{ij} \quad (i = 1,2,\cdots,N) \quad (3)$$

q_i can be explained as the expected reward after one time's transition from state i. So that we can obtain

$$v_i(t) = q_i + \sum_{j=1}^{N} p_{ij}v_j(t-1) \quad (4)$$

Using the form of vectors equation (4) will be denoted as the following:

$$V(t) = Q + \boldsymbol{P}\boldsymbol{V}(t-1) \quad (n = 1,2,3,\cdots) \quad (5)$$

Where, $\boldsymbol{V}(t)$ is the column vector of$v_i(t)$, $\boldsymbol{Q}$ is the column vector of q_i, and $\boldsymbol{P}$ is the Markov transition probability matrix.

3.3 Markov Transition Probability Matrix

From equation (5), if we want to obtain the total expected payoff $V(t)$, firstly we should get the Markov transition probability matrix $\boldsymbol{P}$. In this paper, we adopt the Boltzmann distribution. The Boltzmann distribution provides one method, where the probability of state transition from state i to j at time t is

$$p_{ij} = \exp(uj(t)/\lambda) / \sum_k \exp(uk(t)/\lambda \quad (k = 1, 2, \cdots, N) \qquad (6)$$

Where, $u_j(t)$ is the payoff when the player chooses strategy j. The parameter λ has an important role in the learning process. By increasing λ, we can increase the randomness of decisions. On the other hand, decreasing λ will result in decreasing randomness, which enables the player to choose the optimal strategy more accurately.

4 Examples and Numerical Results

In this section, we presented an illustrative example for 2-players symmetric repeated games. As the example , consider the iterated prisoner's dilemma as an example, prisoner's dilemma is a non-cooperative, non-zero sum game, played between two players. Each player has two choices; either cooperation or defection, and the payoffs they got for their choices are calculated according to Table 1.

Table 1 The general form of prisoner's dilemma

		Cohmm player	
		Cooperation	Defection
Rowplayer	Cooperation	(3,3)	(0,5)
	Defection	(5,0)	(1,1)

From Table. 1 we can easily get that defection is a dominant strategy, so any rational player will choose defection no matter what the other player chooses. But if they cooperate, they would get more. This is the dilemma of prisoners. To overcome this problem, Axelrod presented the thought of iterated prisoner's dilemma [9-10], which repeats the conventional game numerous times with the number of repetitions unknown to both players. Repeating the games this way can give players the hope of cooperation. In this paper, we do not try to explain whether cooperation will be emergent. Rather, we intend to develop a theoretic model for repeated games.

Supposing present time $t = 30$, namely two players have completed 30 times' repeated games. Now we analyze repeated games when $t > 30$. Before $t = 30$ the games was observed, therefore we can assume the times that row player chooses strategy C and D is 21 and 9 respectively in last 30 times' repeated games. Furthermore, the total payoff obtained by choosing C and D is supposed 49 and 9 respectively. Using equation (6) we get the Markov transition probability matrix:

$$\boldsymbol{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.7 & 0.3 \end{bmatrix}$$

According to Von Neumann-Morgenstem expected utility function, we can obtain the reward matrix of state transition:

$$\boldsymbol{U} = \begin{bmatrix} 1 & 7/3 \\ 1 & 7/3 \end{bmatrix}$$

In addition, from equation (3) we have

$$\boldsymbol{Q} = \begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}$$

From the beginning of the 30 th repeated game, if we initialize $v_i(0) = 0$, by using equation (5), we can get the total expected payoff (Table 2).

Table 2 The total expected payoffs of row player in iterated prisoner's dilemma

$t =$	30	31	32	33	34	35	36	37	38	39	40	…
$v_c(t)$	0	1.4	2.8	4.2	5.6	7.0	8.4	9.8	11.2	12.6	14.0	…
$v_p(t)$	0	1.4	2.8	4.2	5.6	7.0	8.4	9.8	11.2	12.6	14.0	…

For further study, we can get the following conclusions:

1) From transition probability matrix

$$\boldsymbol{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.7 & 0.3 \end{bmatrix},$$

we know the Markov process is ergodic. By computation, the steady-state distribution of the Markov process is the same as the transition probability matrix $\boldsymbol{P}$. That is to say, the example has an evolutionary outcome that both players will choose strategy C and D with probability 0.7 and 0.3 respectively.

2) The value of parameter λ in equation (6) can be chosen adaptively according to environment, in this paper, $\lambda = 0.05$. Note that, when $\lambda \to 0$, players have no preference to choices, namely players will assign identical probabilities to the different strategies. When λ is high, the randomness of decisions is increasing, players are inclined to choose strategy D. In addition, when the randomness is increasing, the probability for cooperation emergence will be decreasing. Therefore, we can update the value of λ according to environment in evolutionary and repeated games, so the dynamic model in this paper has the ability of adaptive learning.

5 Conclusions

As an extension of classical game theory, evolutionary and repeated games have attracted much attention in recent years. Especially, the study on learning

model of players with bounded rationality in evolutionary or dynamic games. In this paper, we introduced Markov process into repeated games and presented a dynamic model for symmetric repeated games. This model used Markov process with payoff to simulate the actions of players. Experiments show that the model is effective and can learn adaptively with exoteric environment.

Acknowledgements

We would like to thank the editors and the reviewers for their consideration. We also acknowledge the support by the National Natural Science Foundation of China (Grant No. 60574071).

References

[1] Maynard S J. Evolution and the Theory of Games [M]. Cambridge University Press, New York, 1982.

[2] Maynard S J, Price G R. The logic of animal conflict [J]. Nature, 1973, 246: 15 - 18.

[3] Yao X, Darwen P. Genetic algorithms and evolutionary games [A]. In Barnett W, Chiarella C, Keen S, et al, eds. Commerce, Complexity and Evolution [C]. Cambridge University Press, 2000, 313 - 333.

[4] Thlol Y, Acan A. Ants can play prisoner's dilemma [A]. 2003 IEEE International Conference on Systems, Man and Cybernetics [C]. 2003, 1348 - 1354.

[5] Kanazawa T, Ushio T. Multi-population replicator dynamics with erroneous perceptions [A]. 2004 IEEE International Conference on Systems, Man and Cybernetics [C]. 2004, 1006 - 1011.

[6] Amir M, Berninghaus S K. Another approach to mutation and learning [J]. Games and Economic Behavior. 1996, 14: 19 - 43.

[7] Tadj L, Touzene A. A QBD approach to evolutionary game theory [J]. Applied Mathematical Modelling, 2003, 27: 913 - 927.

[8] Nowak M A, Sasaki A, Taylor C, et al. Emergence of cooperation and evolutionary stability in finite populations [J]. Nature (London), 2004, 428: 646 - 650.

[9] Axelrod R. The evolution of strategies in the iterated prisoner's dilemma [A]. In Davis L, ed. Genetic Algorithms in Simulated Annealing [C]. Pitman, London, 1987, 32 - 41.

[10] Axelrod R, Hamilton W D. The evolution of cooperation. science [J]. 1981, 211: 1390 - 1396.

Author

Liu Weibing, male, born in 1978, graduated from Wuhan University and now is a doctor student in Systems Engineering at Wuhan University, Wuhan, China. Mr. Liu has published over 7 papers. His current research is game theory, decision and control theory, evolutionary algorithm etc. He can be reached by E-mail: liuweibing2002 @ sohu. com