Dynamic Analysis of Cell interactions in Biological Environments under Multiagent Social Learning Framework

Chengwei Zhang[1], Xiaohong Li[1], Shuxin Li[1] and Jianye Hao[2]

Abstract— **Biological environment is uncertain and its dynamic is similar to the multiagent environment, thus the research results of the multiagent system area are of great significance and can provide valuable insights to the understanding of biology. Learning in a multiagent environment is highly dynamic since the environment is not stationary anymore and each agent's behavior changes adaptively in response to other coexisting learners, and vice versa. The dynamics becomes more unpredictable when we move from fixed-agent interaction environments to multiagent social learning framework. Analytical understanding of the underlying dynamics is important and challenging. In this work, we consider a social learning framework with homogeneous learners (e.g., Policy Hill Climbing (PHC) learners), to model the behavior of players in the social learning framework as a hybrid dynamical system. By analyzing the dynamical system, we obtain some conditions about convergence or non-convergence. It can be used to predict the convergence of the system. At last, we experimentally verify the predictive power of our model using a number of representative games.**

I. INTRODUCTION

All living systems live in environments that are uncertain and dynamically-changing. It is remarkable that these systems survive and achieve their goals by exhibiting intelligent features such as adaption and robustness. Biological system behaviors[1] are often the outcome of complex interactions among a large number of cells and their environments.

Similarly, in the multiagent system[2], [3], [4], [5], [6], an important ability of an agent is to adjust its behavior adaptively to facilitate efficient coordination among agents in unknown and dynamic environments. If we regard the cells in the biological system as the agents in the multiagent system, we can analyse the cells' behavior using multiagent system. So understanding collective decision made by such intelligent multiagent system is an interesting research topic not only for artificial intelligent but also for biology. The conclusion of the theoretical analysis can be applied to the research of biology, for example, the results of convergence can be used for explaining the phenomenon of cell's group behaviour.

Now, many researchers have investigated biological systems which are composed of cells and their environment via modeling and simulation[1], [7]. There are two principal approaches: population based modeling and discrete agent based modeling. Population based modeling approximates the cells within any grid box by a set of variables associated with the grid box[8], [9]. Discrete agent based modeling maps each cell to a discrete simulation entity[8], [10], [11].

We use multiagent learning techniques to model the behaviors of each cell agent, which is an important technique to achieve efficient coordination in multiagent system area[6], [12], [13]. Until now, significant amount of efforts have been devoted to develop effective learning techniques for different multiagent interaction environments[14], [15], [16]. In the multiagent environments, each agent interacts with an agent selected from its neighborhood randomly each round, and updates its strategy based on the feedback in the current round. To describe the behavior of an agent, one common line of researches is to extend existing reinforcement learning techniques in single-agent environment to multiple-agent interaction environment. However, due to the violation of Markov property, the existing theoretical guarantees do not hold any more in multiagent environment. It is challenging for us to model the multi-agent environment and understand the learning dynamics of multiagent environments.

This paper presents a social learning framework model to simulate the dynamics of multi-agent system in biological environment, as well as a theoretical analysis of the learning dynamics of this model. The analysis results shed lights on how and when the consistent knowledge in terms of equilibrium can be evolved or not among the population of agents. In the social learning framework, all agents play PHC strategy[17] for decision making, and use a weighted graph model for neighbor selection. In the part of theoretical analysis, we present a theoretical model to analyze the learning dynamics of the learning framework. The purpose of analysing the learning dynamics is to judge whether the learning algorithm that the agent adopt can converge or can not. The intention behind is that convergence to an equilibrium has been the most commonly accepted goal to pursue in multiagent learning literature. Firstly, we model the overall dynamics among agents as a system of differential equations. Then, some conditions are proved to be the sufficient condition of convergence or non-convergence. It can be used to predict the convergence of the system. Finally, we estimate the prediction through simulation experiment. It shows that our theoretical analysis well predicts the experimental results.

The remainder of the paper is organized as follows. Section II reviews normal-form game and the basic gradient ascent approach. Section III introduces the multiagent learning framework of our model. Theoretically analyzation of the proposed framework and its corresponding experimental

[1] Chengwei Zhang, Xiaohong Li and Shuxin Li are with School of Computer Science and Technology, Tianjin University, China. Email:{chenvy,lishuxin,xiaohongli}@tju.edu.cn

[2] Jianye Hao is with School of Computer Software, Tianjin University, China. Email:jianye.hao@tju.edu.cn; Corresponding author

978-1-5090-1610-5/16/$31.00 ©2016 IEEE

simulation are proposed in Section IV and V respectively. Lastly Section VI concludes the paper.

II. Background

A. Normal-form games

In a two-player, two-action, general-sum normal-form game, the payoff for each player $i \in \{k, l\}$ can be specified by a matrix as follows,

$$R_i = \begin{bmatrix} r_i^{11} & r_i^{12} \\ r_i^{21} & r_i^{22} \end{bmatrix} \tag{1}$$

Each player i simultaneously selects an action from its action set $A_i = \{1, 2\}$, and the payoff of each player is determined by their joint actions. For example, if player k selects the pure strategy of action 1 while player l selects the pure strategy of action 2, then player k receives a payoff of r_k^{12} and player l receives the payoff of r_l^{12}.

Apart from pure strategies, each player can also employ a mixed strategy to make decisions. A mixed strategy can be represented as a probability distribution over the action set and a pure strategy is a special case of mixed strategies. Let $p_k \in [0,1]$ and $p_l \in [0,1]$ denote the probability of choosing action 1 by player k and player l respectively. Given a joint mixed strategy profile (p_k, p_l), the expected payoffs of player l and player r can be computed as follows,

$$\begin{aligned} V_k(p_k, p_l) = & r_k^{11} p_k p_l + r_k^{12} p_k (1-p_l) + r_k^{12} (1-p_k) p_l \\ & + r_k^{22} (1-p_k)(1-p_l) \\ V_l(p_k, p_l) = & r_l^{11} p_k p_l + r_l^{12} p_k (1-p_l) + r_l^{21} (1-p_k) p_l \\ & + r_l^{22} (1-p_k)(1-p_l) \end{aligned} \tag{2}$$

A strategy profile is a Nash Equilibrium (NE) if no player can get a better expected payoff by changing its current strategy unilaterally. Formally, $\left(p_k^*, p_l^*\right) \in [0,1]^2$ is a NE, iff $V_k\left(p_k^*, p_l^*\right) \geq V_k\left(p_k, p_l^*\right)$ and $V_l\left(p_k^*, p_l^*\right) \geq V_l\left(p_k^*, p_l\right)$ for any $(p_k, p_l) \in [0,1]^2$.

B. Gradient Ascent (GA) and PHC algorithm

When a game is repeatedly played, an individually rational agent updates its strategy with the propose of maximizing its expected payoff. We know that the gradient direction is the fastest increasing direction, thus it is a well-deserved way to model the behavior of agent using gradient ascent algorithm. Agent i that employs GA-based algorithm updates its policy towards the direction of its expected reward gradient, which is shown in the following equations.

$$\Delta p_i^{(t+1)} \leftarrow \eta \frac{\partial V_i\left(p^{(t)}\right)}{\partial p_i} \tag{3}$$

$$p_i^{(t+1)} \leftarrow \Pi_{[0,1]}\left(p_i^{(t)} + \Delta p_i^{(t+1)}\right) \tag{4}$$

The parameter η is the size of gradient step. $\Pi_{[0,1]}$ is the projection function mapping the input value to the valid probability range of $[0,1]$, which is used for preventing the gradient from moving the strategy out of the valid probability space. Formally, we have,

$$\Pi_{[0,1]}(x) = argmin_{z \in [0,1]} |x - z| \tag{5}$$

To simplify the notation, let us define $u_i = r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21}$, $c_i = r_i^{12} - r_i^{22}$ and $d_i = r_i^{21} - r_i^{22}$. For the two-player case, the Equation 3 and 4 can be represented as follows,

$$p_k^{(t+1)} \leftarrow \Pi_{[0,1]}\left(p_k^{(t)} + \eta\left(u_k p_l^{(t)} + c_k\right)\right) \tag{6}$$

$$p_l^{(t+1)} \leftarrow \Pi_{[0,1]}\left(p_l^{(t)} + \eta\left(u_l p_k^{(t)} + d_l\right)\right) \tag{7}$$

In the case of infinitesimal size of gradient step ($\eta \to 0$), the learning dynamics of the agent can be modeled as a system of differential equations. Further, it can be analyzed using dynamic system theory [19]. It is proved that the strategies of all agents will converge to a Nash equilibrium, or if the strategies do not converge, agents' average payoff will converge to the average payoff of Nash equilibrium [18]. The policy hill-climbing algorithm (PHC) is a combination of gradient ascent algorithm and Q-learning where each agent i adjusts its policy p to follow the gradient of expected payoff (or the value function Q).

III. Modeling Multiagent Learning

Under a multiagent social learning framework with N agents, each agent interacts with one of its neighbors selected randomly from its neighborhood each round. The neighborhood of each agent is determined by its underlying network topology. The interaction between each pair of agents is modeled as a two-agent normal-form game. During an interaction, each agent selects its action following a specified learning strategy, which is updated repeatedly based on the feedback from the environment at the end of each interaction. The framework is presented in Algorithm 1.

Algorithm 1 Overall interaction protocol of the social learning framework

```
repeat
    for each agent in the population do
        Chose one of its neighbors with a certain probability.
        Play a two-player normal-form game with this neighbor and choose one of his action.
        Select a action according to its mixed strategy with suitable exploration.
    end for
    Environmental feedback.
    for each agent in the population do
        Observing reward $r$ and update its policy based on its past experience according to specific policies.
    end for
until the repeated game ends
```

We use graph $G = (V, E)$ to model the underlying neighborhood network, which is composed by $N = |V|$ agents. The edges $E = \{e_{ij}\}$, $i, j \in V$ represent social contacts among agents, where e_{ij} denotes the probability that agent i chooses agent j to interact with. And we have $\sum_{j\in V} e_{ij} = 1 \wedge e_{ii} = 0$. Here, we propose an adaptive strategy for agents to make their decisions in social learning framework with PHC learning strategy, which is shown in Algorithm 2.

Algorithm 2 Learning process in the multiagent framework for agent $i \in V$

```
Let $\alpha \in (0,1]$ and $\delta \in (0,1]$ be learning rates.
Initialize $Q_i(a) \leftarrow 0$, $p_i(a) \leftarrow \frac{1}{|A_i|}$.
repeat
    Select agent $j \in V$ according to $E$ with probability $e_{ij}$, and play a $2\times 2$ game with palyer j.
    Select action $a \in A_i$ according to mixed strategy $p_i$ with suitable exploration.
    Observe reward $r$ according to interaction between $i$ and $j$.
    Update $Q$ value
        $Q_i(a) \leftarrow (1-\alpha)Q_i(a) + \alpha r$
    Step $p$ closer to the optimal policy w.r.t. $Q$,
        $p_i(a) \leftarrow p_i(a) + \Delta_a$
        while constrained to a legal probability distribution,
        $\Delta_a = \begin{cases} -\delta_a & a \neq argmax_{a'} Q_i(a') \\ \sum_{a' \neq a} \delta_{a'} & otherwise \end{cases}$
        $\delta_a = \min\left(p_i(a), \frac{\delta}{|A_i|-1}\right)$
until the repeated game ends
```

Here, $\alpha \in (0,1]$ and $\delta \in (0,1]$ are learning rate, and Q values are maintained just as in normal Q-learning. The policy is improved by increasing the probability of selecting the highest valued action based on the learning rate δ.

IV. Analysis of the Multiagent Learning Dynamics

In this section, we present a theoretical model to estimate and analyze the learning dynamics of the above multiagent learning framework in Algorithm 2. We extend notations in section II to the multiagent environment. Without loss of generality, we consider the case with two-action only.

Assume that the payoff that an agent receives only depends on the joint action, then the payoff for agent $i \in V$ can be defined as a fixed matrix $R_i = \begin{bmatrix} r_i^{11} & r_i^{12} \\ r_i^{21} & r_i^{22} \end{bmatrix}$, where r_i^{mn} denotes the payoff agent i receives when i selects action m and its neighbor selects n. Here, we use the p_i to donate the probability that the player i selects action 1. Then the mixed strategy $(p_1, p_2, ..., p_N)$ in multiagent framework $G = (V, E)$ can be considered as a point in $\mathbb{R}^N$ constrained to the unit square. The expected payoff $V_i(p_1, p_2, ..., p_N)$ of player i can be computed as follows,

$$\begin{aligned} & V_i(p_1,p_2,...,p_n) \\ =& \sum_{j\in V} e_{ij} V_i(j)(p_i,p_j) \\ =& u_i p_i \sum_{j\in V} e_{ij} p_j + c_i p_i + (r_i^{21} - r_i^{22}) p_j + r_i^{22} \end{aligned} \tag{8}$$

where $u_i = r_i^{11} + r_i^{22} - r_i^{12} - r_i^{21}$, $c_i = r_i^{12} - r_i^{22}$ and $V_i(j)(p_i,p_j) = r_r^{11} p_i p_j + r_r^{12} p_i (1-p_j) + r_r^{12} (1-p_i) p_j + r_r^{22} (1-p_i)(1-p_j)$. And e_{ij} is the probability that the agent i selects agent j to interact with.

Each agent i updates its strategy in order to maximize the value of V_i. Recall the equation 3 and 4, we can obtain

$$\begin{aligned} p_i^{(k+1)} &= \prod\nolimits_\Delta \left[p_i^{(k)} + \eta \partial_{p_i} V_i(p_1, p_2, ..., p_N) \right] \\ &= \prod\nolimits_\Delta \left[p_i^{(k)} + \eta \left(u_i \sum_{j\in V} e_{ij} p_j + c_i \right) \right] \end{aligned} \tag{9}$$

where parameter η is the size of gradient step.

As $\eta_p \to 0$, it is straightforward that the equation 9 becomes differential equation. Considering the step size to be infinitesimal, the unconstrained dynamics of the all players' strategies can be modeled by the following model.

$$\dot{p}_i = u_i \sum_{j\in V} e_{ij} p_j + c_i, \quad i \in \{1, 2, ..., N\} \tag{10}$$

Equation 10 can be simplified into $\dot{P} = UEP + C$, where $P = (p_1, p_2, ..., p_N)^T$, $\dot{P} = (\dot{p}_1, \dot{p}_2, ..., \dot{p}_N)^T$ and $C = (c_1, c_2, ..., c_N)^T$. The matrix $U = diag(u_1, u_2, ..., u_N)$ is the diagonal matrix generated by $(u_1, u_2, ..., u_N)$.

For the constrained dynamics of the strategies, we can model it as the following equations,

$$\begin{cases} \dot{p}_i = 0 & p_i = 0 \wedge G_i \leq 0 \\ \dot{p}_i = 0 & p_i = 1 \wedge G_i \geq 0 \\ \dot{p}_i = G_i & otherwise \end{cases} \tag{11}$$

where $G_i = u_i \sum_{j\in V} e_{ij} p_j + c_i$.

Notice that equation 11 is a hybrid system composed of two parts: a series of continuous linear differential dynamic systems in the respective domain space and a switch mechanism between differential dynamic systems when dynamic touch the boundary. Generally, it is hard to obtain a complete conclusion by analyzing dynamics of a general hybrid system, even though the differential system is linear. But we can still find some convergence and non-convergence conditions under certain instances(i.e.,equation 11).

A. Non-convergence condition of the multiagent learning framework

According to the above definition, we have the following general result under which non-convergence is guaranteed.

Theorem 1: In an N agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in equation 11, if the following two conditions are met:

1) There exists a point $P^* = (p_1^*, p_2^*, ..., p_N^*) \in (0,1)^N$, that $UEP^* + C = 0$.
2) There exists a pair of pure imaginary eigenvalues of matrix UE.

Then there exists a set $\mathbb{P} \subset [0,1]^N$ such that the solution of the initial value problem of equation 11 with $P(0) \in \mathbb{P}$ that can not converge.

Proof: Considering the complexity of the hybrid functions, we begin with the unconstrained ones first. Based on the differential equations dynamical systems theorems[19], we calculate the analytic solution of equation **??**. Homogenizing the in-homogeneous equation by substituting P with

$P = X + P^*$, where $UEP^* + C = 0$, we get $\dot{X} = UEX$. Here, UE is an $N \times N$ matrix, then there is a invertible matrix $T = (v_1, ..., v_N)$ that can transform UE into J,

$$T^{-1}UET = J = \begin{bmatrix} J_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & J_m \end{bmatrix}$$

The J_i is a square matrix and its form is one of the following two:

$$(1)\begin{bmatrix} \lambda & 1 & \cdots & \\ & \lambda & 1 & \\ \vdots & & \ddots & \vdots \\ & & \cdots & \lambda \end{bmatrix} \quad (2)\begin{bmatrix} D_2 & I_2 & \cdots & \\ & D_2 & I_2 & \\ \vdots & & \ddots & \vdots \\ & & \cdots & D_2 \end{bmatrix}$$

Where $D_2 = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\alpha, \beta, \lambda \in \mathbb{R}$ and $\beta \neq 0$. Here, J is the Jordan normal form of matrix UE. J_i is the Jordan block corresponding to λ_i, which is a repeated eigenvalue of UE with multiplicity n_i. If eigenvalue λ_i is a real number, then J_i is in the form (1), else (2). Suppose that $\lambda_1, ..., \lambda_k$ are matrix UE's real eigenvalues, and $\lambda_{k+1}, ..., \lambda_m$ is matrix UE's complex eigenvalues, then we have $n_1 + ... + n_k + 2(n_{k+1} + ...n_m) = N$.

Then the analytic solution of function $\dot{X} = UEX$ with initial value $X(0)$ will be:

$$X(t) = \exp(tUE)X(0) = T \begin{bmatrix} e^{tJ_1} & & \\ & \ddots & \\ & & e^{tJ_m} \end{bmatrix} T^{-1}X(0)$$

Using the notation $Y(t) = T^{-1}X(t)$, we have

$$Y(t) = \exp(tJ)Y(0) = \begin{bmatrix} e^{tJ_1} & & \\ & \ddots & \\ & & e^{tJ_m} \end{bmatrix} Y(0)$$

Suppose that $\lambda_k = \beta i$ is a pure imaginary eigenvalue of UE with multiplicity n_k, $\bar{\lambda}_k = -\beta i$ is an eigenvalue of UE with multiplicity n_k either. Then J has a block J_k,

$$J_k = \begin{bmatrix} D_2 & I_2 & \cdots & \\ & D_2 & I_2 & \\ \vdots & & \ddots & \vdots \\ & & \cdots & D_2 \end{bmatrix}, \text{ where } D_2 = \begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}.$$

Due to $e^{tD_2} = \exp\left(t\begin{bmatrix} 0 & \beta \\ -\beta & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos\beta t & \sin\beta t \\ -\sin\beta t & \cos\beta t \end{bmatrix}$, there must exist a pair of items about vector $Y(t)$ as follows.

$$\begin{cases} y_i(t) = y_i(0)\cos\beta t + y_{i+1}(0)\sin\beta t \\ y_{i+1}(t) = -y_i(0)\cos\beta t + y_{i+1}(0)\sin\beta t \end{cases} \quad (12)$$

If $y_i(0) \neq 0 \vee y_{i+1}(0) \neq 0$, then equation 12 has a periodic solution. Let v_i and v_{i+1} to denote eigenvector of $T = (v_1, ..., v_N)$ corresponding to λ_k and $\bar{\lambda}_k$, respectively. Note that $X(t) = TY(t)$, then the solution of equation **??** with the initial value $P(0) \in S$ is cyclical, where

$$S = \left\{P \in [0,1]^N | P = k_1v_1 + k_2v_2 + P^*, k_1, k_2 \in \mathbb{R}\right\}$$

Because of $P^* \in (0,1)^N$, there must exists a $\varepsilon > 0$ for the deleted neighborhood $\mathbb{B}(P^*; \varepsilon) \subset (0,1)^N$ of P^*,

$$\mathbb{B}(P^*; \varepsilon) = \left\{x \in \mathbb{R}^N | 0 < ||x - P^*||_2 < \varepsilon\right\} \subset (0,1)^N$$

Let $\mathbb{P}$ denote $S \bigcap \mathbb{B}(P^*; \varepsilon)$, the solution of the equation 11 with any initial value belongs to $\mathbb{P}$ is cyclical, which means the algorithm corresponding to the equation 11 can not converge. ■

Theorem 1 shows that there exist some situations in which the agents fail to converge under the multiagent social learning framework.

B. Convergence condition of the multiagent learning framework

In most cases, the conditions that guarantee the convergence of a algorithm are more valuable.

Theorem 2: In an N agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in equation 11, if the following two conditions are met:

1) There exists a point $P^* = (p_1^*, p_2^*, ..., p_N^*) \in (0,1)^N$, that $UEP^* + C = 0$.
2) All of the eigenvalues of matrix UE has negative real part.

Then all the solutions of the initial value problem of equation 11 with $P(0) \in [0,1]^N$ will converge eventually.

Proof: The conclusion is obvious. It is known that the construction of the linear dynamic system is stable. If all eigenvalues of matrix UE have negative real part, then point P is a stable equilibrium point. It means that all the solutions of the initial value problem of the equation 11 with $P(0) \in [0,1]^N$ will converge to P. ■

Theorem 2 proposes a sufficient condition to identify the convergence of dynamic in equation 11. We know that it is hard to calculate eigenvalues of a matrix with high dimensional. Here, we propose a more realistic convergence condition which is suitable for multiagent learning framework shown in algorithm 2.

Theorem 3: In an N agent, two-action, integrated general sum game, every player follows the constrained dynamics of the strategy we defined in equation 11, if matrix UE is symmetrical, then all the solution of the initial value problem of equation 11 with $P(0) \in [0,1]^N$ will converge eventually.

Proof: The eigenvalues of real symmetric matrices are real numbers[20]. We analyze all the cases of equation 11 when all of the eigenvalues of matrix UE are real:

1) There exists a point $P^* = (p_1^*, p_2^*, ..., p_N^*) \in (0,1)^N$, that $UEP^* + C = 0$.
2) There are no such a point, that $UEP^* + C = 0$.

For case 1), if all eigenvalues of matrix UE are negative number, then point P is a stable equilibrium points; otherwise, all the solutions of the initial value problem of the hybrid system with $P(0) \in [0,1]^N$ will move away from P toward boundary of the hybrid system [19]. Because the domain of hybrid model 11 has boundary(i.e., $P(t) \in [0,1]^N$), then there must exists a point $P' = (p_1', ..., p_N')^T$ in the boundary of the domain, where $(p'_i = 0 \wedge G_i \leq 0) \vee (p'_i = 1 \wedge G_i \geq 0)$ for all $i \in V$. The dynamic $P(t)$ will converge to P' eventually.

Similarly, we can find a point $P' = (p'_1, ..., p'_N)^T$ in the boundary of the hybrid system domain in case 2) and the dynamic $P(t)$ will converge to P' eventually. ■

V. Experimental Simulation

In this section, we compare the empirical dynamics of the multiagent social learning framework with PHC learners with theoretical prediction of our hybrid dynamic model. We perform two experiments that satisfy the Theorem 1 and Theorem 3, respectively.

A. A non-convergence multiagent Game

In this subsection, we consider a 4-agent, two-action games. The game is defined as follows,

$$R_1 = R_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, R_3 = R_4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Metrix $R_i, i \in \{1,2,3,4\}$ is the payoff matrix of agent i, and element e_{ij} of matrix E is the probability that player i selects player j in each interaction. In this game, we have $u_1 = u_3 = 2$, $u_2 = u_4 = -2$, $c_1 = c_3 = -1$, and $c_2 = c_4 = 1$. Then the unconstrained dynamic model of this game is $\dot{P} = UEP + C$, where $UE = \begin{bmatrix} 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \end{bmatrix}$, $C = (-1, 1, -1, 1)^T$. Then this game has a $P^*(1/2, 1/2, 1/2, 1/2)^T \in (0,1)^4$, which satisfies $UEP^* + C = 0$. Matrix UE has a pair of pure imaginary eigenvalues which is $\lambda_1 = 2i$ and $\lambda_1 = 2i$. The eigenvectors are $v_1 = (0, 1/2, 0, 1/2)^T$ and $v_2 = (1/2, 0, 1/2, 0)^T$ corresponding to λ_1 and λ_2. Let $P(0) = P^* + k_1 v_1 + k_2 v_2$. As long as k_1 and k_2 are sufficiently small, according to Theorem 1, the solution of the initial value problem of game 1 with $P(0)$ can't converge.

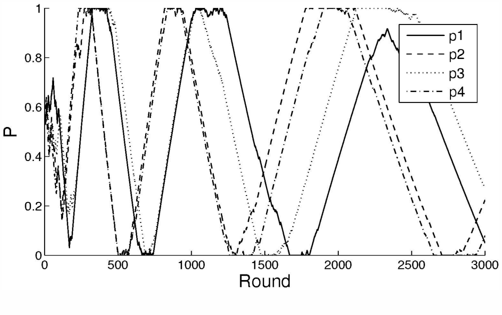

Fig. 1. Agent dynamics of game satisfying the conditions of Theorem 1

In Figure 1, the dynamic solution of the game with initial value $P(0)$ is plotted, where $k_1 = k_2 = 0.1$. Each of the four lines in figure 1 shows the strategy's dynamic changing of each agent, respectively. We can see that the strategies of those agents do not converge. Obviously, the simulation results are consistent with the theoretical prediction.

B. A convergence multi-agent Game

In this subsection, we consider a 4-agent, two-action games. The game is defined as follows,

$$R_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, i \in \{1,2,3,4\}$$

$$E = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

Metrix $R_i, i \in \{1,2,3,4\}$ is the payoff matrix of agent i, and element e_{ij} of matrix E is the probability that player i selects player j in each interaction. In this game, we have $u_i = 2$ and $c_i = -1, i \in \{1,2,3,4\}$. Then the unconstrained dynamic model of this game is $\dot{P} = UEP + C$, where $UE = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$, $C = (-1, -1, -1, -1)^T$. Because matrix UE is symmetrical, according to Theorem 3, the solution of the initial value problem of this game with any $P(0) \in [0,1]^4$ will converge eventually.

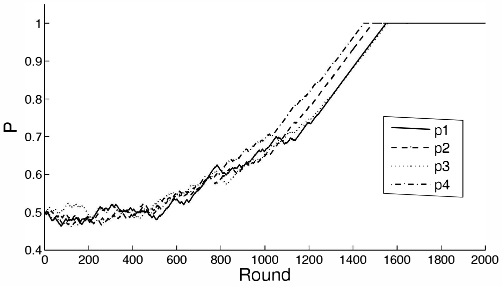

Fig. 2. Agent dynamics of game satisfying the conditions of Theorem 3

Figure 2 illustrates dynamics of the PHC learners' strategy for the game with initial value initial value $P(0) = (1/2, 1/2, 1/2, 1/2)^T$. Each of the four lines in figure 2 shows the strategy's dynamic changing of each agent, respectively. We can see that the strategies of those agents converge eventually, which are consistent with the theoretical prediction.

VI. Conclusion

In this work, we proposed a multiagent social learning framework to model the behavior of agent in biologic environment, and theoretically analyzed the dynamics of multiagent social learning framework using non-linear dynamic theories. We obtain and prove some sufficient conditions about convergence or non-convergence by the theoretically analysis. It can be used to predict the convergence of the system. Experimental results show that the predictions of our dynamic model are consistent with the simulation results.

REFERENCES

[1] Kang S, Kahan S, Mcdermott J, et al. Biocellion: accelerating computer simulation of multicellular biological system models. Bioinformatics, 2014, 30(21):3101-8.

[2] Kaelbling L P, Littman M L, Moore A W. Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 1996, 4(1):237–285.

[3] Li X, Zhang C, Hao J. Socially-Aware Multiagent Learning: Towards Socially Optimal Outcomes. European Conference on Artificial Intelligence. 2016:533-541.

[4] Hao J, Leung H F, Ming Z. Multiagent Reinforcement Social Learning toward Coordination in Cooperative Multiagent Systems. Acm Transactions on Autonomous and Adaptive Systems, 2014, 9(4):374-378.

[5] Hao J, Leung H F. The dynamics of reinforcement social learning in cooperative multiagent systems. International Joint Conference on Artificial Intelligence. 2013:184-190.

[6] Busoniu L, Babuska R, De Schutter B. A Comprehensive Survey of Multiagent Reinforcement Learning. IEEE Transactions on Systems Man & Cybernetics Part C, 2008, 38(2):156-172.

[7] Torii M, Wagholikar K, Liu H. Detecting concept mentions in biomedical text using hidden Markov model: multiple concept types at once or one at a time?. Journal of Biomedical Semantics, 2014, 5(1):3-3.

[8] Anderson A R A, Chaplain M A J. Continuous and Discrete Mathematical Models of Tumor-induced Angiogenesis. Bulletin of Mathematical Biology, 1998, 60(5):857-899.

[9] Xavier J B, Martinezgarcia E, Foster K R. Social evolution of spatial patterns in bacterial biofilms: when conflict drives disorder. American Naturalist, 2009, 174(1):1-12.

[10] Ferrer J, Prats C, Lpez D. Individual-based Modelling: An Essential Tool for Microbiology. Journal of Biological Physics, 2008, 34(1-2):19-37.

[11] Jeannin-Girardon A, Ballet P, Rodin V. An Efficient Biomechanical Cell Model to Simulate Large Multi-cellular Tissue Morphogenesis: Application to Cell Sorting Simulation on GPU. Theory and Practice of Natural Computing. Springer Berlin Heidelberg, 2013:96-107.

[12] Matignon L, Laurent G J, Fort-Piat N L. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. Knowledge Engineering Review, 2012, 27(1):1-31.

[13] Bloembergen D, Tuyls K, Hennes D, et al. Evolutionary dynamics of multi-agent learning: a survey. Journal of Artificial Intelligence Research, 2015, 53(1):659-697.

[14] Abdallah S, Lesser V. A multiagent reinforcement learning algorithm with non-linear dynamics. Journal of Artificial Intelligence Research, 2008, 33(1):521-549.

[15] Zhang C, Lesser V R. Multi-Agent Learning with Policy Prediction. Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, Usa, July. 2010.

[16] Chakraborty D, Stone P. Multiagent learning in the presence of memory-bounded agents. Autonomous Agents and Multi-Agent Systems, 2014, 28(2):182-213.

[17] Bowling M, Veloso M. Multiagent learning using a variable learning rate. Artificial Intelligence, 2002, 136(2):215-250.

[18] Singh S P, Kearns M J, Mansour Y. Nash Convergence of Gradient Dynamics in General-Sum Games. Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc. 2000:541–548.

[19] Shilnikov, Leonid P. Methods of qualitative theory in nonlinear dynamics. Methods of qualitative theory in nonlinear dynamics /. World Scientific, 1998:592.

[20] Olshevsky V, Tyrtyshnikov E. Matrix methods : theory, algorithms and applications : dedicated to the memory of Gene Golub. IEEE Trans Signal Process, 2010, 51(5):1306 - 1323.