SCIENCE CHINA
Information Sciences

• LETTER •

June 2021, Vol. 64 169202:1–169202:3
https://doi.org/10.1007/s11432-019-9926-2

Profile-dynamic based fictitious play

Xiao ZHANG[1] & Daizhan CHENG[1,2*]

[1]*School of Control Science and Engineering, Shandong University, Jinan* 250061, *China;*
[2]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190, *China*

Received 1 February 2019/Accepted 30 May 2019/Published online 15 May 2020

Citation Zhang X, Cheng D Z. Profile-dynamic based fictitious play. Sci China Inf Sci, 2021, 64(6): 169202, https://doi.org/10.1007/s11432-019-9926-2

Dear editor,

The fictitious play (FP) was initially proposed by Brown [1] in 1951 as a handy learning rule. The main idea of FP can be described as follows: In a game process, the player who adopts FP assumes that the others choose their strategies randomly and independently, according to a fixed distribution. Based on this assumption, the player tends to choose his or her own strategy empirically to counter the opponents and ultimately earn more benefits. Many distinguished studies have been springing up after the introduction of FP.

In this study, a new type of FP called the profile-dynamic based fictitious play (PDBFP) is proposed. In a PDBFP process, players assume that their opponents choose the strategies depending on the previous profiles. One application of PDBFP is learning the opponents' dynamic evolutions in evolutionary games (EGs). By the acquired empirical transition matrix, players who adopt PDBFP are able to improve their payoffs. The motivation of this study is to find a new variant of FP, which intends to predict other players' dynamic evolutions in EGs. Inspired by the algebraic expression of finite EGs [2,3], an empirical transition matrix is constructed. The empirical matrix will be updated and used to estimate the opponents' joint strategy at each step of the game. The key improvement of PDBFP is that not only the opponents' joint strategies but also their strategy updating rules are estimated.

The fundamental tool used in our approach is the algebraic state space representation of finite games based on the semi-tensor product (STP) of matrices. This technique was firstly proposed in [4], and it has been successfully applied to investigate Boolean networks and logical systems subsequently [5]. For more details about STP and its applications in games and logical systems, please see [6–8].

The main contribution of this study is that in PDBFP, the players are allowed to make and modify their strategies simultaneously, in contrast to many EGs where the players update the strategies one by one or in a random sequential order.

Definition 1 (Normal form game)**.** A normal form game (NFG) is defined by a triple $G = (N, S, C)$, where

(i) $N = \{1, 2, \ldots, n\}$ is the set of players;

(ii) $S_i = \mathcal{D}_{k_i}$ is the set of strategies for player i, $i = 1, \ldots, n$, where $\mathcal{D}_{k_i} := \{1, 2, \ldots, k_i\}$ (Moreover, $S = \prod_{i=1}^{n} S_i$ is called the profile of G, and $S_{-i} = \prod_{j \neq i} S_j$ is the joint strategy of all the players except player i);

(iii) $c_i : S \to \mathbb{R}$ is the payoff function of player i, $i = 1, \ldots, n$, and $C = \{c_1, \ldots, c_n\}$.

Definition 2 ([3])**.** The strategy dynamics of an NFG is

$$\begin{cases} x_1(t+1) = f_1(x_1(t), \ldots, x_n(t)), \\ x_2(t+1) = f_2(x_1(t), \ldots, x_n(t)), \\ \qquad\qquad \vdots \\ x_n(t+1) = f_n(x_1(t), \ldots, x_n(t)), \end{cases} \tag{1}$$

where $x_i(t) \in \mathcal{D}_{k_i}$ is the strategy of player i at time t, $f_i : \prod_{j=1}^{n} \mathcal{D}_{k_j} \to \mathcal{D}_{k_i}$, $i = 1, \ldots, n$ are mix-valued logical functions, and f_i is determined by the strategy updating rule (SUR) of the EG.

Using STP, the algebraic express of (1) can be obtained, that is

$$x(t+1) = Lx(t), \tag{2}$$

where $x(t) = \ltimes_{j=1}^{n} x_j(t)$, $L \in \mathcal{L}_{k \times k}$ is a logical matrix, and $k = \prod_{i=1}^{n} k_i$.

Definition 3 (Myopic best response adjustment)**.** Construct a set of best response set of strategies at time t as

$$\overline{\beta}_i(t) := \underset{s_i \in S_i}{\operatorname{argmax}}\, c_i(s_i, s_{-i}(t)). \tag{3}$$

(1) If $s_i(t) \in \overline{\beta}_i(t)$, then $s_i(t+1) = s_i(t)$;

(2) If $s_i(t) \neq \overline{\beta}_i(t)$, then choose any strategy $j \in \overline{\beta}_i(t)$, with equal probability $p = \frac{1}{|\overline{\beta}_i(t)|}$.

Definition 4. PDBFP is a variant of FP consisting of the following factors:

(1) The exogenous initial weight function of all the players except player i (utilized by player i), represented as

$$\kappa_0^{-i}(s_{-i}, s) : (S_{-i} \times S) \to \mathbb{N}; \tag{4}$$

* Corresponding author (email: dcheng@iss.ac.cn)

(2) The exogenous weight function utilized by player i at time $t \geqslant 2$, represented as

$$\begin{aligned}\kappa_t^{-i} &= \kappa_{t-1}^{-i} + I\{s_{-i}(t-1) = s_{-i} | s(t-2) = s\} \\ &= \sum_{\tau=2}^{t} I\{s_{-i}(\tau-1) = s_{-i} | s(\tau-2) = s\}. \end{aligned} \tag{5}$$

For $\forall i \in N$ and $\forall (s_{-i}, s) \in S_{-i} \times S$, $\kappa_1^{-i} = \kappa_0^{-i} = 0$.

Definition 5. In PDBFP processes, the probability that player i assigns to other players' playing s_{-i} at time t, while the profile is s at time $t-1$, is given by a time-varying function $\gamma_t^{-i} : (S_{-i} \times S) \to \mathbb{R}_+$, that[1)]

$$\gamma_t^{-i}(s_{-i}, s) := \begin{cases} \dfrac{\kappa_t^{-i}(s_{-i}, s)}{\sum_{s^*_{-i} \in S_{-i}} \kappa_t^{-i}(s^*_{-i}, s)}, & s \in \tilde{S}_{t-1}, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where κ_t^{-i} is the exogenous weight function defined in (4), s is the joint strategy of all the players, and $s_{-i} \in S_{-i}$.

Definition 6. At time $t \geqslant 2$,

(1) The empirical frequency of an arbitrary joint strategy $\overline{s}_{-i} \in S_{-i}$, which is acquired by player i, is

$$p_{-i}^{\overline{s}_{-i}}(t) = \gamma_t^{-i}(\overline{s}_{-i}, s(t-1)); \tag{7}$$

(2) The empirical frequency vector of player i for all the other players' profiles is

$$p_{-i}(t) = (p_{-i}^{s_1}, p_{-i}^{s_2}, \ldots, p_{-i}^{s_{k/k_i}})^{\mathrm{T}}. \tag{8}$$

Definition 7. Construct two kinds of probability distributions over the sample space $S_i \times S$. Player i chooses s_i with the probability $p_i^{s_i}(t)$ at time t as

$$p_i^{s_i}(t) := \begin{cases} \dfrac{\mathrm{e}^{\frac{1}{\theta} c_i(s_i, s_{-i}(t-1))}}{\sum_{\overline{s_i} \in S_i} \mathrm{e}^{\frac{1}{\theta} c_i(\overline{s}_i, s_{-i}(t-1))}}, & s(t-1) \notin \tilde{S}_{t-1}, \\ \dfrac{\mathrm{e}^{\frac{1}{\theta} E[c_i(s_i, p_{-i}(t))]}}{\sum_{\overline{s_i} \in S_i} \mathrm{e}^{\frac{1}{\theta} E[c_i(\overline{s}_i, s_{-i}(t-1))]}}, & \text{otherwise}, \end{cases} \tag{9}$$

where $\theta > 0$ is a parameter which determines the likelihood that player i chooses a suboptimal strategy: (i) $\theta \to 0$ means that player i will choose a best response to $p_{-i}(t)$ or $s_{-i}(t-1)$; (ii) $\theta \to \infty$ means that player i will choose any strategy $s_i \in S_i$ with equal probability.

Theorem 1. In 2-person finite games where the evolutionary process of player i is a Markov chain and player j takes his or her strategies by adopting PDBFP, player j will finally acquire player i's evolution dynamics.

Using STP, the best response PDBFP processes can be expressed as

$$x_{-i}^e(t) = L_p^{-i}(t) \ltimes_{j=1}^n x_j(t-1),$$

$$x_i(t) = \begin{cases} \underset{x_i \in \Delta_{k_i}}{\operatorname{argmax}} V_i^c W_{[k_i \times_{j=1}^{i-1} k_j]} x_i x_{-i}^e(t), \\ \quad \text{if} \quad x(t-1) \in \tilde{S}_{t-1}, \\ \underset{x_i \in \Delta_{k_i}}{\operatorname{argmax}} V_i^c W_{[k_i \times_{j=1}^{i-1} k_j]} x_i x_{-i}(t-1), \\ \quad \text{if} \quad x(t-1) \notin \tilde{S}_{t-1}, \end{cases} \tag{10}$$

where $L_p^{-i}(t) \in \Upsilon_{k/k_i \times k}$ is a probabilistic matrix with its ℓ-th column being

$$[\gamma_t^{-i}(\delta_{k/k_i}^1, \delta_k^\ell), \ldots, \gamma_t^{-i}(\delta_{k/k_i}^{k/k_i}, \delta_k^\ell)]^{\mathrm{T}}, \ \ell = 1, \ldots, k.$$

We call it the "profile empirical matrix" of player i. It is obvious that $x_{-i}^e(t) = p_{-i}(t)$; $x_{-i}^e(t) \in \Upsilon_{k/k_i \times 1}$ is player i's prediction of the others' joint strategy at time t; $V_i^c \in \mathbb{R}^k$ is the structure vector of c_i; and $\Upsilon_{m \times n}$ is the set of $m \times n$ probabilistic matrices where $\Upsilon_{i,j} \geqslant 0$ and $\sum_{i=1}^m \Upsilon_{i,j} = 1$, $j = 1, \ldots, n$.

Theorem 2. In n-person finite games, if at any time $t > 0$, the profile $s(t)$ generated by a best response PDBFP process is a Nash equilibrium, then for $\forall \tau > 0$, $s(t+\tau) = s(t)$.

Theorem 3. In n-person finite games where all the players adhere to best response PDBFP, any pure strategy steady state must be a Nash equilibrium.

The following example shows the advantages of PDBFP, comparing with FP.

Example 1. In a 2-player rock-paper-scissors game, player 1 adopts best response PDBFP to counter player 2, while player 2 uses a transition probability matrix L_2 to generate strategies. At each period, the winner gets 1 point, the loser loses 1 point, and both get 0 point for a tie.

We use $x_i(t) \in \Delta_3$ to express the strategies used by player i at period t: δ_3^1 (rock), δ_3^2 (scissors), and δ_3^3 (paper). Then the payoff vectors are

$$\begin{aligned} V_1^c &= [0, 1, -1, -1, 0, 1, 1, -1, 0]; \\ V_2^c &= [0, -1, 1, 1, 0, -1, -1, 1, 0]. \end{aligned}$$

The evolution dynamics of player 1 is

$$x_2^e(t+1) = L_p^2(t) \ltimes x(t);$$

$$x_1(t+1) = \begin{cases} \underset{x_1 \in \Delta_3}{\operatorname{argmax}} V_1^c x_1 x_2^e(t+1), \\ \quad \text{if} \quad x(t) \in \{x(0), \ldots, x(t-1)\}; \\ \underset{x_i \in \Delta_{k_i}}{\operatorname{argmax}} V_1^c x_1 x_2(t), \\ \quad \text{if} \quad x(t) \notin \{x(0), \ldots, x(t-1)\}. \end{cases}$$

The evolution dynamics of player 2 is

$$x_2(t+1) = L_2 x(t),$$

where

$$L_2 = \begin{bmatrix} 0.7, 0.1, 0.2, 0.1, 0.2, 0.7, 0.1, 0.7, 0.2 \\ 0.1, 0.2, 0.7, 0.2, 0.7, 0.1, 0.2, 0.1, 0.7 \\ 0.2, 0.7, 0.1, 0.7, 0.1, 0.2, 0.7, 0.2, 0.1 \end{bmatrix}.$$

We simulated player 1's accumulated payoffs, comparing with the case where player 1 uses FP. The outcome is shown in Figure 1, where we can easily find that after 1000 times' play, the payoffs gained by adopting best response PDBFP are obviously higher than the one gained by FP.

1) Actually, if one profile $\widehat{s}$ has not appeared in the past, the values of $\gamma_t^{-i}(s_{-i}, \widehat{s})$ can be defined arbitrarily. Because it will not be used as an empirical frequency vector, at each time t, there are only k/k_i values which will be utilized by player i, out of k^2/k_i values from $\gamma_t^{-i}(s_{-i}, s)$, and these selected values are corresponding to profile $s(t-1)$ which appeared at time $t-1$.

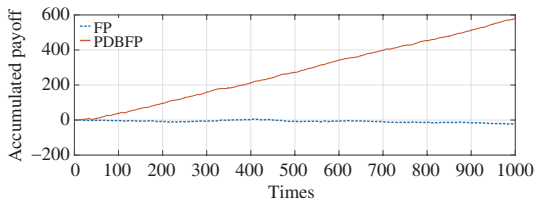

Figure 1 (Color online) Player 1's payoffs via using different rules.

The point is that FP cannot distinguish the evolution dynamics of player 2, because of the strong correlation of player 2's strategies. This can be found in the following experimental data: (i) the empirical frequency vector acquired by FP: $q_2(1001) = [0.334, 0.334, 0.332]^{\mathrm{T}}$; (ii) the profile empirical matrix acquired by PDBFP:

$$L_p^2(1001) = \begin{bmatrix} 0.69, 0.10, 0.22, 0.11, 0.20, 0.69, 0.11, 0.71, 0.21 \\ 0.07, 0.22, 0.68, 0.22, 0.71, 0.10, 0.20, 0.12, 0.69 \\ 0.24, 0.68, 0.10, 0.67, 0.09, 0.21, 0.69, 0.17, 0.10 \end{bmatrix}.$$

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61773371, 61733018).

References

1 Brown G W. Iterative solution of games by fictitious play. Activ Anal Prod Allocation, 1951, 13: 374–376
2 Cheng D Z, Xu T T, Qi H S. Evolutionarily stable strategy of networked evolutionary games. IEEE Trans Neural Netw Learn Syst, 2014, 25: 1335–1345
3 Cheng D Z, He F H, Qi H S, et al. Modeling, analysis and control of networked evolutionary games. IEEE Trans Automat Contr, 2015, 60: 2402–2415
4 Cheng D Z, Qi H S, Li Z. Analysis and Control of Boolean Networks: A Semi-tensor Product Approach. London: Springer, 2011
5 Lu J Q, Li H T, Liu Y, et al. Survey on semi-tensor product method with its applications in logical networks and other finite-valued systems. IET Control Theory Appl, 2017, 11: 2040–2047
6 Cheng D Z, Qi H S, Liu Z Q. From STP to game-based control. Sci China Inf Sci, 2018, 61: 010201
7 Li C X, He F H, Liu T, et al. Symmetry-based decomposition of finite games. Sci China Inf Sci, 2019, 62: 012207
8 Li H T, Ding X Y. A control Lyapunov function approach to feedback stabilization of logical control networks. SIAM J Control Opt, 2019, 57: 810–831