

取り違えのある繰り返し囚人のジレンマにおける協力のダイナミクス

村井 伸一郎*
Shinnichiro Murai

五十嵐 瞭平†
Ryohei Igarashi

岩崎 敦†
Atsushi Iwasaki

1 はじめに

繰り返しゲームは、長期的関係にあるプレイヤー間の（暗黙の）協調を説明するためのモデル [3] であり、主に経済学分野で企業間の談合といった協調行動を分析するために発展してきた。2人がまったく見間違えない「完全」観測下では、常に裏切り（ALLD）や一度でも裏切られたら許さない（GRIM）といった非協力的な戦略しか生き残らないことが知られている [4]。しかし、実際の人間はしばしば行動を取り違えたり、見間違えることがある。例えば、協力したが、失敗してしまったり、サボったつもりがうまくいってしまったりといったことが起こると考えるのは自然である。こうした行動の取り違えは進化ゲーム理論における重要な仮定であると考えられている。実際、こうした間違いがないと、お互いに協力することが進化的安定性を満たさないことが知られている [1]。また、行動の見間違えは、自分が行動を取り違えたかどうかかわからない（自分の行動が相手にどう見えているかわからない）状況での行動の取り違えのモデルである。

繰り返しゲームの戦略表現として、1期記憶戦略（Memory-One Strategies）がよく使われる。しかしこの表現は、行動の取り違えにおいて、いずれかのプレイヤーが一度でも裏切ったら二度と協力することがない戦略である GRIM を表現できない。つまり、お互いが行動を取り違えたとき、相互協力を回復することがある。そこで本研究では戦略を有限状態機械（Finite State Automaton, FSA）で表現する。FSA もよく使われる表現であるが、多くの先行研究で、プレイヤーが行動を取り違えるか否かで、その後の振る舞いを区別していない。このため、実質的には行動の見間違えと同じ戦略空間しか扱えていなかった。

そこで本研究では、プレイヤーが行動を取り違えた後の振る舞いを区別した戦略空間の下で、自然淘汰と突然変異がどのような戦略が選択するかを吟味し、行動の見間違えの結果と比較する。その結果、見間違えの下で生き残りやすい GRIM は、取り違えの下では、お互いが行動を取り違えた後に相互協力に戻るようになり生き残ることがわかった。また、見間違えの下で生き残りにくかったしつぱ返し戦略（Tit-For-Tat, TFT）は、取り違えの下の方が生き残りやすいことを明らかにした。

表 1: プレイヤ 1 の利得 $g_1(\hat{a})$

	$\hat{a}_2 = C$	$\hat{a}_2 = D$
$\hat{a}_1 = C$	1	-l
$\hat{a}_1 = D$	1+g	0

表 2: 同時確率分布 $o((\hat{a}_1, \hat{a}_2)|(a_1, a_2))$

	$\hat{a}_2 = a_2$	$\hat{a}_2 \neq a_2$
$\hat{a}_1 = a_1$	0.95	0.01
$\hat{a}_1 \neq a_1$	0.01	0.03

2 モデル

本章では行動の取り違えのある無限回繰り返しゲームをモデル化する。ここでプレイヤー $i \in \{1, 2\}$ はステージゲームを無限期間 $t = 0, 1, 2, \dots$ に渡って繰り返す。割引因子は $\delta \in (0, 1)$ とする。各期においてプレイヤー i は有限集合 $A_i = \{C, D\}$ から行動 a_i を選択し、その行動の組を $\mathbf{a} = (a_1, a_2) \in A^2$ とする。

このとき、意図した行動を a 、実現した行動を \hat{a} とする。プレイヤー 1 の利得 $g_1(\hat{a})$ は表 1 のように与える。ここで、意図した行動の組 (a_1, a_2) に対して、実現した行動の組 (\hat{a}_1, \hat{a}_2) が起きる確率を同時確率分布 $o((\hat{a}_1, \hat{a}_2)|(a_1, a_2))$ で定義する。

プレイヤーの戦略は、そのプレイヤーが意図した行動から実際にとった行動への写像で表現される。本研究では戦略を状態数 2 以下の非同相な 482 個の FSA を戦略空間とする。FSA の状態は、 R (reward, 報酬) と P (punishment, 処罰) の 2 つに区別され、プレイヤー i は状態 R で行動 $a_i = C$ を選び、状態 P で行動 $a_i = D$ を選ぶ。それぞれの状態でプレイヤーは自分と相手がつとった行動で次にどの状態に遷移するかが決まる。例えば、状態 R からは 4 つの行動の組に対して状態遷移が決まる。図 1 や図 2 において、状態 R の CC や CD は自分が行動を取り違えなかったとき、 DC や DD は自分が行動を取り違えたときを表す。行動の見間違えの下では、自分が行動を取り違えたかどうかを区別できないので、状態 R であれば CC と CD 、状態 P であれば DC と DD が将来の状態遷移を決定する。その場合の非同相な FSA は 26 個に限定される。

このような数ある戦略の中から有効な戦略を発見する方法の 1 つとして、突然変異付きレプリケータダイナミクス [2] がある。本論文では、その方程式を

$$\dot{x}_i = x_i [f_i(\vec{x}) - \phi(\vec{x})] + u \left(\frac{1}{n} - x_i \right), \quad i = 1, \dots, n \quad (1)$$

と定義する。 $\phi(\cdot)$ を全ての戦略の利得の平均 $\sum_j x_j f_j(\vec{x})$ 、 $f_j(\cdot)$ を $\sum_m x_m a_{jm}$ とする。ただし、 a_{jm} は戦略 j をとるプレイヤーが戦略 m を取るプレイヤーと無限回プレイしたときの割

* 電気通信大学情報理工学域

† 電気通信大学大学院情報理工学域研究科

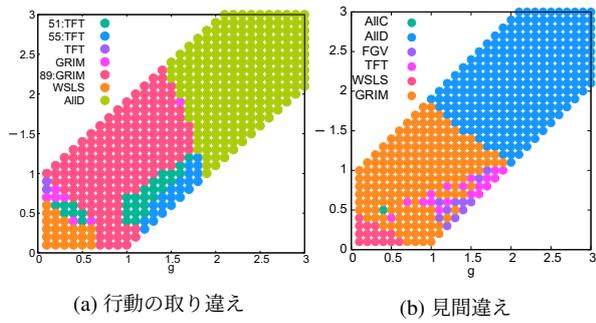


図 1: 最大多数プロット

引利得和である。

数値実験では、50000 期の帰結を分析した。同時確率分布 $\alpha((\hat{a}_1, \hat{a}_2)|(a_1, a_2))$ は表 2 に従う。割引因子 δ を 0.90, 突然変異率 u を 0.01 とした。 g と l は $[0.1, 3.0]$ の範囲で 0.1 刻みで変化させ、 $|g - l| < 1$ となる組のみを使用した。

3 取り違えがある環境下のダイナミクス

図 1a に取り違えのある環境下、図 1b に見間違えのある環境下における最大多数戦略を示す。最大多数戦略とは、収束時に最も多くの人口を獲得した戦略である。図の横軸は自分の裏切りによる利得の増分 g 、縦軸は相手の裏切りによる損失 l に対応している。図 1a, および図 1b が示すように、行動の取り違えと見間違えでは類似した戦略が生き残ることがわかった。行動の取り違えと見間違えの違いとして、見間違えにおける GRIM は協力に戻ることはないが、行動の取り違えにおける GRIM は協力に戻る方が生き残りやすい。また、見間違えにおける TFT は生き残りにくいが、取り違えでは $g - l = 1$ に近い領域で生き残りやすい。

まず、 g と l がどちらも大きいとき、取り違え、見間違えどちらの環境下においても常に裏切る AIID が最大多数となる。次に、 g と l が中程度のとき、見間違えでは最初に協力し、相手の裏切りを見るとそれ以降永遠に裏切り続ける GRIM が最大多数となる。取り違えにおいて同じ領域で最大多数となる 89 番 (図 2) は行動の取り違えが起こらないときは、見間違えにおける GRIM と同じ振る舞いをするが、再度、協力に戻ることが可能な戦略である。89 番は、各プレイヤーがお互いに裏切ろうとし

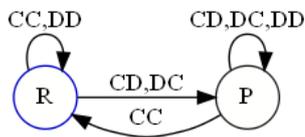


図 2: 89 (GRIM)

ているときに同時に取り違えることで、結果的にお互いに協力を取ることができるようになる。そのため、同時に取り違えた後もお互いに協力を取ろうとすることで、お互いに裏切り合い続けるより大きな利得が得られ、生き残りやすくなる。

また、 g と l がどちらも中程度かつ、 $g > l$ の場合、見間違えではしつぱ返し戦略 (Tit-For-Tat, TFT) は常に最大多数とはならないが、取り違えでは 51 番 (図 3a)、55 番 (図 3b) が最大多数となる。TFT は、状態 R からスタートし、相手が協力した次の期では協力を、裏切った次の期には裏切りを行う戦略である。51 番と 55 番は行動の取り違えが起こらないときは、見間違えにおける TFT と同じ振る舞いをする戦略である。ここで、

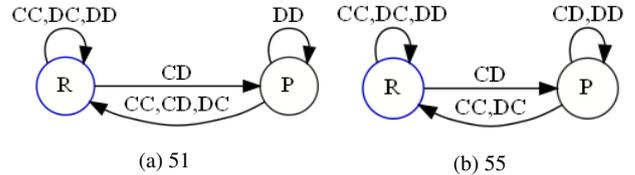


図 3: TFT

TFT はお互いに裏切りと協力を繰り返す報復の連鎖に陥ることがある。報復の連鎖の際に各プレイヤーが 2 期で獲得できる利得は、 $1 + g - l$ となる。この利得が 2 期とも協力した時の利得 (2) と一致したとすると、 $1 + g - l = 2$ 、つまり、 $g - l = 1$ となる。図 1a において $g - l = 1$ の直線を考えると、行動の取り違えがあるとき、TFT はこの直線に近い領域で生き残りやすい。

また、この領域において、51 番と 55 番では g の値が大きいときほど、55 番が生き残りやすく、この 2 戦略では状態 P において自分が取り違え、相手が正しく行動したときの遷移、つまり、図 3a と図 3b における状態 P から CD の遷移が異なる。ここでは、プレイヤー 1 が取り違えて C を取り、プレイヤー 2 は正しく行動して D を取ったとする。プレイヤー 2 は状態 P から DC の遷移により、次の期に状態 R で協力を取ろうとするため、プレイヤー 1 が状態 P にとどまると報復の連鎖が起こる。まず、 g の値が l よりも非常に大きいとき、先述した $g - l = 1$ に近い関係が成り立つ。したがって、プレイヤー 1 が状態 P にとどまる 55 番が生き残りやすくなる。次に、 g の値が少し小さくなると、 $g - l = 1$ の関係が成り立ちにくくなる。そのため、プレイヤー 1 は状態 P にとどまっても生き残りにくくなる。したがって、プレイヤー 2 は次の期に状態 R で協力を取ろうとするため、51 番のようにプレイヤー 1 も状態 R で協力を取ろうとすることでお互いに協力することができ、生き残りやすくなる。

参考文献

- [1] D. Fudenberg and E. Maskin. Evolution and Cooperation in Noisy Repeated Games. *The American Economic Review*, 80(2):274-279, 5 1990.
- [2] B. Zagorsky, J. Reiter, K. Chatterjee, and M. Nowak. Forgiver triumphs in alternating prisoner's dilemma. *PLOS ONE*, pp. 1-8, 2013.
- [3] 神取. 人はなぜ協調するのか—くり返しゲーム理論入門—. 三菱経済研究所, 2015.
- [4] 西野上, 五十嵐, 岩崎. 私的観測下の繰り返し囚人のジレンマにおける協力のダイナミクス. 第 19 回情報科学技術フォーラム, 2020.