限定合理的なプレイヤーの戦略分析

東京工業大学　石井慎也　ISHII Shinya

1　はじめに

近年，プレイヤーの合理性に関心が持たれるようになり，限定合理性という視点を取り入れた分析も盛んに行われている．限定合理性を考慮した分析を行うため，マシン（有限オートマトン）を用いたプレイヤーモデルがよく用いられる．(Rubinstein [3], Abreu and Rubinstein[1])

本研究では，計算機シミュレーションを用い，繰り返し囚人のジレンマ型ゲームにおける限定合理的なプレイヤーの戦略を分析する．

アクセルロッドの選手権 (Axelrod [2]) 以来，計算機シミュレーションによる分析はゲーム理論においても利用されているが，その難点としてパラメータへの依存性が指摘されている．そこで本研究では，実験パラメータが結果に及ぼす影響も考慮し，戦略を分析する．

2　マシンゲーム

プレイヤーは戦略を表現するマシンを選択し，それに基づいて，繰り返しゲームをプレイする．プレイヤー i のマシン M_i は次の4つ組で定義される．

$$M_i = \langle Q_i, q_i^0, f_i, \tau_i \rangle$$

ただし，Q_i は有限集合，$q_i^0 \in Q_i$，$f_i : Q_i \to A_i$（ただし，A_i はプレイヤー i の行動の集合），$\tau_i : Q_i \times A_j \to Q_i \, (j \neq i)$ である．Q_i は状態の集合，q_i^0 は初期状態，f_i は出力関数，τ_i は遷移関数をそれぞれ意味する．今期の行動を決定するのが出力関数であり，次期の状態を決定するのが遷移関数である．

3　設定

計算機実験では，プレイヤーは有限繰り返しの囚人のジレンマ型ゲームをプレイする．設定は総当たり型，進化型，ノイズ型の3つのモデルに分類される．

総当たり型では，各プレイヤーはそのプレイヤー自身も含めたすべてのプレイヤーと対戦する．ここでは，利

表5　利得行列

		Player2 C	Player2 D
Player1	C	α, α	β, γ
	D	γ, β	δ, δ

$\alpha = 3, \beta = 0, \gamma = 5, \delta = 1$ とする．

得を変えることにより，各戦略のパラメータに対する依存についても調べる．

進化型では，2種類のアルゴリズム，アクセルロッド型，戦略更新型を適用する．

アクセルロッド型のアルゴリズム

1，各プレイヤーの存在比率を計算する．
2，各プレイヤーの期待利得を求める．
1, 2を1世代とし，指定世代数回繰り返す．

戦略更新型のアルゴリズム

1，繰り返しゲームをプレイする．
2，平均利得を比較し，確率 p で戦略を更新する．
1, 2を1世代とし，指定世代数回繰り返す．

ノイズ型では，3種類のノイズ，認識の誤り，実行の誤り，記憶の誤りを想定する．

認識の誤り
相手の行動を間違えて認識してしまう．

実行の誤り
自分が間違えて行動してしまう．

記憶の誤り
自分の戦略を表現するオートマトンにおいて，現在の状態を間違えてしまう．

各モデルにおけるパラメータの設定は表1から表4の通りである．

4　結果

総当たり型の結果は図1の通りである．*Tit For Tat* 戦略の利得は *All-D* 戦略の利得より低く，全戦略の平均

表 1　総当たり型におけるパラメータ設定

繰り返し回数	100
世代数	1
最大状態数	3, 4

表 2　アクセルロッド型におけるパラメータ設定

繰り返し回数	100
世代数	1, 2, 3, 4, 5, 10
最大状態数	3

表 3　戦略更新型におけるパラメータ設定

繰り返し回数	100
世代数	戦略が 1 種類になるまで
最大状態数	3
確率 p	1/2

表 4　ノイズ型におけるパラメータ設定

繰り返し回数	100
世代数	1
最大状態数	3
ノイズ率	0.01, 0.02, 0.03, 0.04, 0.05, 0.10

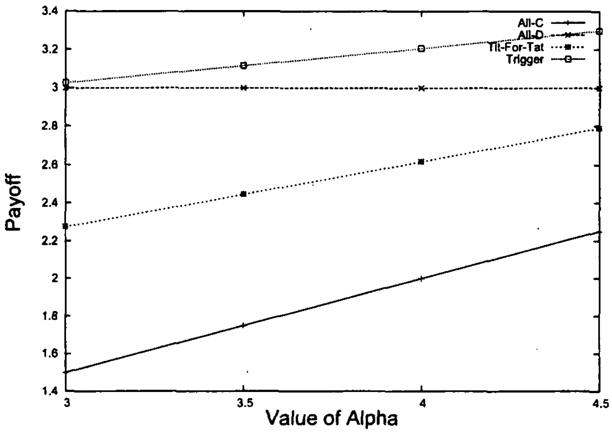

図 1　α の変化による各戦略の利得変化

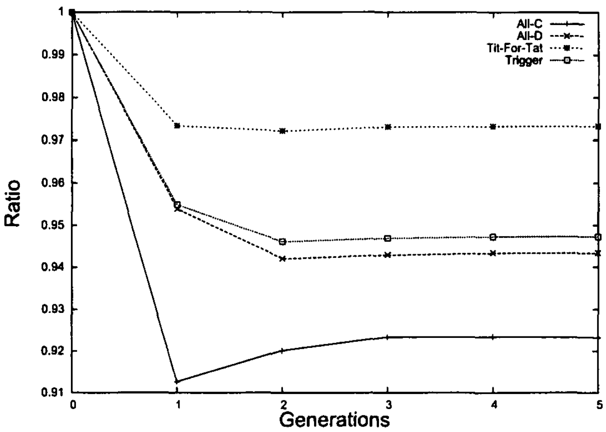

図 2　0 世代の利得を 1 としたときの各世代の利得

利得に近い．さらに，C を基調とするプレイヤーが適応しやすいように利得を変えても，*All-D* 戦略の利得より低いという結果は変わらない．

アクセルロッド型の結果は図 2 の通りである．*Tit For Tat* 戦略の数値は 0.97 と他の戦略に比べ非常に高い．進化過程において，*Tit For Tat* 戦略は安定した戦略といえる．しかし，絶対的な値をみると，例えば，*Tit For Tat* 戦略の利得が *All-D* 戦略の利得よりも低い．

また，戦略更新型では，シミュレーション開始から 10 時間後のデータをみると，トリガー戦略や *All-D* 戦略を用いるプレイヤーは存在するが，*Tit For Tat* 戦略を用いるプレイヤーは存在しない．2 つのアルゴリズムを適用した進化過程においても，*Tit For Tat* 戦略は優れた戦略であるとはいえない．

ノイズ型では，実行の誤り，記憶の誤りの下では，総当たり型において高い利得を獲得した他の戦略と同様，*All-D* 戦略は利得を下げる．一方，認識の誤りの下では，優れた戦略の多くが利得を下げるなか，*All-D* 戦略は利得を下げない．

5　おわりに

本研究では，マシンゲームの新しいシミュレーション環境を構築した．その下で，パラメータの結果への影響に注意し，限定合理性を考慮したゲームにおける戦略を分析した．

参考文献

[1] D. Abreu and A. Rubinstein. The Structure of Nash Equilibrium in Repeated Games with Finite Automata. *Econometrica*, 56:1259–1281, 1988.

[2] R. Axelrod. The Evolution of Cooperation. Basic Books, 1984.

[3] A. Rubinstein. Finite Automata Play the Repeated Prisoner's Dilemma. *Journal of Economic Theory*, 39:83–96, 1986.