2009-ICS-154 (1)
2009 / 3 / 2

繰り返しゲームにおける限定合理的なプレイヤーの戦略に関する研究

Strategies of Bounded Rational Players in Repeated Games

舛井道晴
Michiharu Masui

東京工業大学大学院　社会理工学研究科　社会工学専攻
Tokyo Institute of Technology, Department of Social Engineering

1 背景

ゲーム理論の枠組みでは通常，すべての意思決定者は完全な推論能力や計算機，ときにはそれ以上の計算能力を備えたプレイヤーを仮定している。多項式時間では解くことができないような，計算量の大きいクラスの問題の意思決定も，合理的なプレイヤーは短時間でかつ正確に実行できることになっている。しかし近年，プレイヤーの合理性に関心がもたれるようになり，限定合理性という視点を取り入れた分析も盛んにおこなわれている（Neyman(1985), Binmore and Samuelson(1992), Cooper(1996)）。特に，計算機科学的制約を設けることにより，これまで多くの注目すべき理論的分析結果が得られており，特に囚人のジレンマゲームに関する分析は，多くの手法によって行われている。

限定合理性を考慮した分析を行うためには，どのように“限定”された合理性をモデル化するかが問題となる。そのために様々な手法が提案されているが，中でも有限オートマトンを用いたプレイヤーのモデル化手法がよく取り入れられている（Rubinstein(1986), Abrue and Rubinstein(1988)）。有限オートマトンにより戦略を表現し，その状態数を戦略の複雑さとする。そして，その複雑さの上限を設定されたプレイヤーを考える。

本研究では，計算機シミュレーションを用い，いくつかの繰り返し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。限定合理性という枠組みを与えることで，各プレイヤーの利得がどのように変化するかを分析することが目的である。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。例えば，アクセルロッドによって行われた囚人のジレンマゲームのコンピュータプログラム選手権は有名である（Axelrod(1984)）。この選手権では，A.Rapoport による Tit-For-Tat（仕返し戦略，TFT）の有効性を示すなど，以後のゲーム理論，広くは社会科学における計算機シミュレーションに大きな影響を与えた。だが一方で，この選手権に関する難点としてパラメータへの依存性などが指摘されており、それに関する研究もなされている。特に利得，繰り返し回数，対戦するプレイヤーの種類に結果が大きく依存する。利得を変えることにより，協調をとるプレイヤーが適応しやすい環境を整えることもできれば、裏切りをとるプレイヤーが適応しやすい環境を整えることもできる。繰り返し回数が少なければ裏切りをとるプレイヤーが有利となるケースも発生する。本研究では実験パラメータが結果に及ぼす影響も調べる。

2 モデル

ここで，本研究で扱うモデルを概説する。

2.1 ナッシュ均衡

ゲーム理論において重要な解概念であるナッシュ均衡を定義する。プレイヤーが n 人いる戦略形ゲーム G において，戦略の組 $(s_1^*, \ldots, s_n^*)$ がナッシュ均衡であるとは，すべてのプレイヤー $i(1, \ldots, n)$ に対して，戦略 s_i^* が他のプレイヤーの戦略の組 s_{-i}^* に対する最適反応であるときを言う。利得関数を f とすると，戦略の組 s^* がナッシュ均衡であるとは，すべてのプレイヤー i に対して，

$$f_i(s^*) \geq f_i(s_i, s_{-i}^*), \forall s_i \in S_i$$

ただし，上記は純粋戦略の範囲でのナッシュ均衡である。次に扱うゲームの構造を概説する。

2.2 囚人のジレンマゲーム

囚人のジレンマゲームとは，２人のプレイヤーが「協調 (Cooperation = C)」,「裏切り (Defection = D)」の２つの戦略を持ち，利得行列が以下のように与えられるゲームのことをいう。相手の行動に関わらず常に最適となる行動が存在する。

1 \ 2	C	D
C	2,2	-1,3
D	3,-1	0,0

表 1　囚人のジレンマ

2.3 両性の闘い

両性の闘いとは，関心・趣味は異なるが，何かしら２人で行動したいがゆえに利害構造を生んでしまう２人のプレイヤーの意思決定問題である。

1 \ 2	A	B
A	0,0	1,2
B	2,1	0,0

表 2　両性の闘い

2.4 チキンゲーム

チキンゲームは交渉を伴う問題においてとりあげられる。プレイヤーの少なくとも一方が譲歩しない限り、最悪な結末を生んでしまう構造を持っている。進化ゲームの文脈ではタカハトゲームの名で知られている。

1 \ 2	C	D
C	4,4	2,5
D	5,2	0,0

表 3　チキンゲーム

2.5 マシンゲーム

ここで，限定合理性なプレイヤーを表現する上で必要となるムーアマシンについて説明する。プレイヤー i のマシン M_i は次の 4 つの要素で表現される。

$$M_i = \{Q_i, q_i^0, f_i, \tau_i\}$$

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 は遷移関数を意味する。それぞれ、Q_i は有限集合，$q_i^0 \in Q_i$，$f_i : Q_i \rightarrow A_i(A_i$ はプレイヤー i の行動の集合)，$\tau_i : Q_i \times A_j \rightarrow Q_i(\mathrm{j} \neq i)$ である。

【例：Tit-For-Tat 戦略】

Tit-For-Tat 戦略とは，第 1 期は C をとり，それ以降は 1 期前に相手が選択した行動と同じ行動をとる戦略のことである。この戦略を実現するマシン $M_i = \{Q_i, q_i^0, f_i, \tau_i\}$ は次のように定義される。

- $Q_i = \{1, 2\}$
- $q_i^0 = 1$
- $f_i(1) = C, f_i(2) = D$
- $\tau_i(1, C) = 1, \tau_i(1, D) = 2, \tau_i(2, C) = 1, \tau_i(2, D) = 2$

このマシンを図式化すると，以下のようになる。このように，Tit-For-Tat 戦略は状態数 2 のマシンによって実現される。状態 1 における出力は C である。相手の出力が C である限り状態 1 にいるが，相手の出力が D であると状態 2 に遷移する。状態 2 における出力は D である。相手の出力が D である限り状態 2 にいるが，相手の出力が C であると同時に状態 1 に遷移する。一般に，マシンゲームとは，各プレイヤーの戦略がマシンに基づいているゲームのことをいう。

2.6 戦略の複雑さ

戦略をマシンにより表現するとき，戦略の複雑さはその戦略を実行するために必要な状態数と定義する。状態をプレイヤーの記憶ととらえると，表現するために必要な状態数が大きいオートマトンは，複雑であると考えられる。

ただし，ここでは最小化されたオートマトンのみを考えることにする。最小化されたオートマトンとは，同じ挙動を示すオートマトンの中で，最も状態数の少ないオートマトンのことを指す。たとえば，図 2 は最小化されたオートマトンではないが，図 3 は最小化されたオートマトンである。図 2 には同じ値を出力する，より状態数の少ないオートマトンが存在するが，図 3 のそれは存在しない。このような図 2 のケースは対象外とする

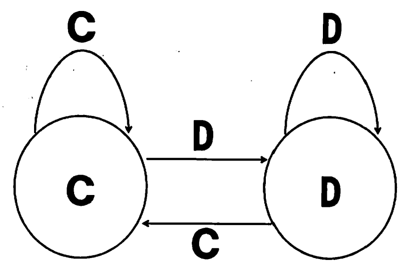

図 1　Tit-For-Tat 戦略

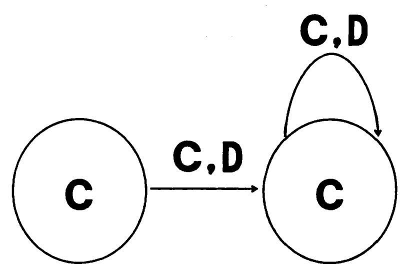

図 2　最小化されていないオートマトン

限定合理的なプレイヤーとは，計算能力に限界を持ったプレイヤーのことを指す。つまり，戦略の複雑さという意味において実行できる戦略が限定されているプレイヤーのことである。また，戦略をマシンにより実現するならば，戦略の表現に用いることのできるオートマトンの状態数が限定されたプレイヤーのことである。

3 設定

3.1 パラメータ

本研究の主なパラメータとして，利得，繰り返し回数などがある。また，導入する戦略（オートマトン）も重要な要素である。

【利得】

戦略の組に対する利得は，各プレイヤーが獲得できる利得に大きな影響を与える。前述した表における利得は代表的な数値であり，一意には定まらない。そこで利得のみを変えるという設定の下，シミュレーションを行い，利得の影響を調べる。

【繰り返し回数】

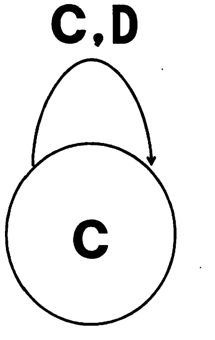

図 3　最小化されたオートマトン

繰り返し回数が少ない場合，その繰り返し回数は各プレイヤーが獲得できる利得に影響を与える。例えば，お互いが協調という結果を実現するために，あえて裏切りを許すという戦略も，繰り返し回数が少なければ，単に裏切られただけで終わってしまう可能性がある。そこで，安定した結果が得られるように，各プレイヤーの対戦の繰り返し回数を 100 回とする。

【導入する戦略】

対戦するプレイヤーの戦略は各プレイヤーが獲得できる利得に最も大きな影響を与える。例えば，自分から裏切ることのない戦略が良い結果を得るためには，やはり自分から裏切ることのないプレイヤーとの対戦が必要になる。そこで導入するオートマトンは，囚人のジレンマゲームにおいてよく知られた，すべて協調型（ALL-C），すべて裏切り型 (ALL-D)，Tit-For-Tat，トリガーなどのほかに，ノイズに強いされる Pavlov 型や Tit-For-Tat を変えた Tat-For-Tit を導入する。

3.2　対戦形式

3.2.1　総当たり型

参加するすべてのプレイヤーが総当たりをするゲームを考える。各プレイヤーは自身とも対戦する。さらにここでは，利得を変えることにより，その変化による結果への影響を分析する。総当たり型の目的は，総当たりという単純な枠組みの中でどのような戦略が有効であるかを分析することにある。設定を単純にすることにより，結果に対する明確かつ正確な分析が可能となる。

繰り返し回数	100
世代数	1

表 4　総当たり型におけるパラメータ設定

3.2.2　進化型（Axelrod 型）

Axelrod によって行われたプログラム選手権を応用する。以下のような流れで行う。第 0 世代として総当たり型の結果を用いる。なお，N をプレイヤーの集合，R_i をプレイヤーの存在比率，$P_{i,j}$ をプレイヤー j とのゲームにおける i の利得，EV_i をプレイヤー i の期待利得とする。

1. 各プレイヤーの存在比率を計算する。
   - 各プレイヤーの存在比率は 1 期前のゲームにおけるプレイヤー全体の総利得に対する，そのプレイヤーの総利得の比率によって決定。つまり，

$$R_i = \frac{\sum_{j \in N} P_{i,j}}{\sum_{i \in N} \sum_{j \in N} P_{i,j}}$$

2. 各プレイヤーの期待利得を求める。
   - 各プレイヤーの期待利得，各プレイヤーの存在比率に基づく。

$$EV_i = \sum_{j \in N} (P_{i,j} \times R_j)$$

3. 1,2 を第 1 世代として，指定世代数を繰り返す。

Axelrod 型の目的は進化の過程における戦略の有効性を分析することにある。総当たり型においてよい成績を残した戦略が，戦略の取捨選択の過程においても高い利得を獲得できるかを分析する。

繰り返し回数	100
世代数	1,2,3,4,5,10

表 5　Axelrod 型のパラメータ設定

3.2.3　進化型（戦略更新型）

Axelrod 型とは異なった進化の過程を考える。以下の流れで行う。

1. 繰り返しゲームをプレイ。
2. 平均利得を比較し，戦略を更新する。
   - 今回のゲームを含めた各プレイヤーの平均利得を計算し，それを比較する。その結果に基づき，平均利得の低いプレイヤーは確率 p で平均利得の高いプレイヤーへと乗り換える。このとき，平均利得の高いプレイヤーは戦略を更新しない。
3. 1,2 を第 1 世代として，指定世代数を繰り返す。

繰り返し回数	100
世代数	戦略が1種類になるまで
確率p	1/2

表6　戦略更新型のパラメータ設定

3.3　ノイズ

意思決定者が何かしらのきっかけで決めた行動を誤ってしまうことを考慮する。これを「ノイズ」と呼び，本研究では3つのノイズパターンに分類する。

【認識の誤り】

相手の行動を間違えて認識してしまう。例えば，相手が協調行動をしてきたにも関わらず，裏切りを選択してきたと認識してしまう誤りである。

【実行の誤り】

自分が間違えて行動してしまう。例えば，現在の状態における出力の値が C であるにも関わらず，D を出力してしまう。

【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。例えば，現在2番目の状態にあるにも関わらず，3番目の状態であると間違えてしまう。

繰り返し回数	100
世代数	1
ノイズ確率	0.01,0.03,0.05,0.10

表7　ノイズ型におけるパラメータ設定

以上のようなノイズを導入する目的は，戦略のノイズへの耐性を分析することにある。これはAxcelrodのノイズに対する分析との比較にも有用である。

4　まとめ

本研究では以上のような設定の元で様々なシュミレーション実験を行った。一例として興味深い発見を紹介する。以下は繰り返し囚人のジレンマゲームにおいて，ナッシュ均衡を導出するために用いられるTit-For-Tat戦略，トリガー戦略のほかに，すべて協調（ALL-C），すべて裏切り（ALL-D）の4つによる対戦結果である。表は，

戦略	平均利得
全体	2.09
ALL-C	1.38
ALL-D	2.83
Tit-For-Tat	2.21
Trigger	2.87

表8　Axelrod型，世代数10

Axelrod型の世代交代実験，ノイズを導入した実験の結果である。いずれの場合においても，「Tit-For-Tat戦略の平均利得 < ALL-D戦略の平均利得 ≃ トリガー戦略の平均利得」となっていることがわかる。Tit-For-Tatは個別の対戦においての優劣はつきにくいが，平均として好成績をあげることが知られている。だが，上の結果は

戦略	平均利得
全体	2.25
ALL-C	1.65
ALL-D	2.85
Tit-For-Tat	2.23
Trigger	2.83

表9　ノイズ型，実行の誤り，ノイズ率:0.10

それとは異なる結果を示している。このように，ゲームおよびプレイヤーに制限を加えることで，Tit-For-Tatよりも平均利得の比較という点で優れた戦略が観察されることは，非常に興味深い。

参考文献

1. R.Axelrod, “The Evolution of Cooperation.” Basic Book, 1984.
2. A.Rubinstein, “Finite Automata Play the Repeated Prisoner’s Dilemma.” *Journal of Economic Theory*,39:83-96,1986.
3. D.Abrue and A.Rubinstein,“The Structure of Nash Equilibrium in Repeated Games with Finite Automata.” *Econometrica*,59:1259-1281,1988.