法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-07

動的な内部状態を管理する2人ゲーム

HOSHI, Satoru / 星, 聡

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
9
(開始ページ / Start Page)
137
(終了ページ / End Page)
142
(発行年 / Year)
2014-03
(URL)
https://doi.org/10.15002/00010535

動的な内部状態を管理する2人ゲーム

Two-person's games with dynamic internal parameters

星 聡
Satoru Hoshi
法政大学大学院 情報科学研究科 情報科学専攻
E-mail: satoru.hoshi.7x@stu.hosei.ac.jp

Abstract

Recent video games have become so complex and highly qualified that non-player characters, called NPC, are required to behave more intelligently and naturally against their situations and environments. There are few theoretical ways to build NPSs as intelligent players, although most games apply ad-hoc rule-based methods to NPCs for their making decisions. Since such methods of NPCs play in a simple and typical manner, human players tend to get bored as games go. For these problems, this paper proposes a decision-making model using game theory involving player's internal parameters. In this model, the payoff functions and state transitions will be dynamically calculated with internal parameters of both players for maximizing internal parameters. Big difference from traditional game theory resides on such maximization method for dynamically changing. This paper demonstrates that two players' best-response for a game with two internal parameters and two strategies is decided into a linear space problem. It also discusses the mathematical properties of a game, the long term gaming behavior (i.e. divergence or convergence of parameters) and the manipulation of the equilibrium coming from repetition of games on the border of strategic change. Experiments are carried out under a detailed condition of the relationship between player's states and strategies using parameter matrices for cooperative or uncooperative features in a simple two-person repeated game.

1 序論

近年，ビデオゲームの高度化に伴い，より大規模かつ複雑な環境や振る舞いに適応できる知的ノンプレイヤーキャラクタ（NPC）の登場が期待されている [1]．この背景として，ルールベース表現（rule-based representation）をはじめとした，従来のアドホックな手法をによる知的 NPC の実現が困難になりつつあるという実情がある．この状況を受けて，現在，個々のビデオゲームの性質に依存しない，統一的かつ汎用的な知的 NPC の実現に向けた取り組みが行われている [2]．本研究はこれを動機として，知的 NPC の汎用的な意思決定手法をゲーム理論 [3] で実現するモデルを提案する．ゲーム理論は，戦略を検討する考え方として古くから利用されてきた領域である．その中でも，無限繰り返しゲームはプレイヤ間の長期的な関係性を説明するモデルであり，実世界における企業間や国家間の協調や対立関係を分析するために用いられてきた [4]．本研究では，知的 NPC の意思決定を複数の内部状態をもつプレイヤ同士の無限繰り返しゲームとみなし，各プレイヤが，連続的に変化する自己および他者の内部状態から動的に利得を導くことで意思決定を行うモデルについて，数理的に分析する．

プレイヤが状態の概念を持つゲーム理論を論じたものとして，秋山らの研究 [5, 6, 7] がある．この研究では，コモンズの悲劇（tragedy of the commons）を題材とし，プレイヤ個々の内部状態と時間変化を伴う共有リソースを定義し，プレイヤのとる戦略が動的に変化するきこりのジレンマ・ゲーム（lumberjack's dilemma game）を扱っている．また，進化ゲーム理論的アプローチを用いて意思決定を行っており，プイレヤの世代交代と学習を繰り返すことによって，周辺環境の変化に対してより最適解に近い戦略を選択する手法について研究している．

この他，利得が変化するゲーム理論の例としては，無限繰り返し囚人のジレンマ・ゲームにおいて協調解が均衡解であることを証明したフォーク定理（folk theorem）をはじめとして，有限繰り返し囚人のジレンマ・ゲームにおいてプレイヤ間の合理的な協調行動を導くためのモデルを提案した松原らの研究 [8] など，多数が存在する．

プレイヤ自身の内部状態がゲームに反映される点において，本研究と秋山らの研究は共通しているが，秋山らの研究における内部状態は利得（伐採できる木の量）の増減に関与する一要素であるのに対し，本研究における内部状態は利得と直結しており，プレイヤは最も内部状態の成長が期待できる戦略を選択するという点で差異がある．さらに，本研究における利得は相手プレイヤの内部状態にも影響を受け，双方の内部状態と戦略が互いに作用しあうという点に大きな特徴があり，このようなゲームのモデルについて論じたケースは筆者らの知る限り存在しない．

本研究では，秋山らの研究を参考にしつつ，はじめに 2 章において内部状態をもつゲームのモデルを定義し，3 章では戦略と内部状態のいくつかのパターンに関する数学的性質を明らかにする．それらを踏まえ，4 章および 5 章において，提案モデルを用いて実際にシミュレーションを行った結果を考察する．また，本モデルの性質やゲームの結果が，実世界でのどのような現象に相当するかも同様に議論する．

2 内部状態をもつゲームの定式化

本章では，状態をもつ戦略ゲームの基本モデルを定式化する [9]．以下では，とくに断りが無い限り自プレイヤを A，相手プレイヤを B とする．

状態をもつプレイヤの無限繰り返しゲームにおいて，プレイヤ A, B は，各時刻 t における自己の利得 P_A, P_B を最大化させるべく戦略を選択する．ここで，A の戦略数を m，B の戦略数を n とし，A が戦略 $i(i = 1, \ldots, m)$，B が戦略 $j(j = 1, \ldots, n)$ を選択したときの利得を，それぞれ P_A^{ij} とする．

- P_A^{ij}: A が戦略 i，B が戦略 j をとるときの A の利得
- P_B^{ij}: A が戦略 i，B が戦略 j をとるときの B の利得

これを，A の戦略 i，B の戦略 j に対する利得表にまとめたものを，表 1 に示す．

なお，次節以降ではプレイヤ A に注目して詳細を述べるが，プレイヤ B に関しても同様の手法を用いている．

2.1 内部状態と貢献度

一般的なゲーム理論の場合，P_A^{ij}，P_B^{ij} は定数で与えられるが，本研究では，P_A^{ij} が A の内部状態で変化するモデルを考える．内部状態とは，たとえば国家間の戦争ゲームである場合，各国の人口，軍事力，財力などに相当するものである．A の内

Supervisor: Prof. Satoru Fujita

表1 ゲームの利得表

	S_B^1	S_B^2	...	S_B^n
S_A^1	(P_A^{11}, P_B^{11})	(P_A^{12}, P_B^{12})	...	(P_A^{1n}, P_B^{1n})
S_A^2	(P_A^{21}, P_B^{21})	(P_A^{22}, P_B^{22})	...	(P_A^{2n}, P_B^{2n})
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
S_A^m	(P_A^{m1}, P_B^{m1})	(P_A^{m2}, P_B^{m2})	...	(P_A^{mn}, P_B^{mn})

部状態を $\mathbf{x}_A$ で表し，A の内部状態の数を a とすると，次のように表現できる．

$$\mathbf{x}_A = (x_{A1}, x_{A2}, \cdots, x_{Aa})^{\mathrm{T}}$$

利得 P_A は内部状態と連動するため，プレイヤ A の目標は自己の内部状態を最高の状態に導くことであると考えられる．ただし，プレイヤ A にとってすべての内部状態が等価値であるとは限らない．そこで，以下のように内部状態 $\mathbf{x}_A$ にそれぞれの利得への貢献度 $\mathbf{v}_A$ を掛け合わせた線形結合モデルを考える．

$$P_A = \mathbf{v}_A \cdot \mathbf{x}_A$$

2.2 時刻

内部状態 $\mathbf{x}_A$ は，A の戦略 i，B の戦略 j によって変化する．これを表現するために，時刻の概念を導入する．時刻 $t+1$ における内部状態を $\mathbf{x}_A(t+1)$ とすると，この値は，時刻 t のプレイヤ A と B の内部状態 $\mathbf{x}_A(t)$，$\mathbf{x}_B(t)$ とそのときに A，B がとった戦略 i，j に依存する．これを式に表すと次のようになる．

$$\mathbf{x}_A^{ij}(t+1) = f^{ij}(\mathbf{x}_A(t), \mathbf{x}_B(t)) \tag{1}$$

ここで，$f^{ij}(\mathbf{x}_A(t), \mathbf{x}_B(t))$ は，引数にプレイヤ A，B の時刻 t における内部状態をとり，それぞれの戦略 i，j のもとで，時刻 $t+1$ の状態を生み出す任意の関数である．図1は，現在の時刻 t におけるプレイヤ A と B の内部状態と戦略の組み合わせから，次の時刻 $t+1$ における $\mathbf{x}_A$ が連続的に計算されていく様子を示したものである．

$$P_A^{ij}(t) = \mathbf{v}_A \cdot \mathbf{x}_A^{ij}(t+1) \tag{2}$$

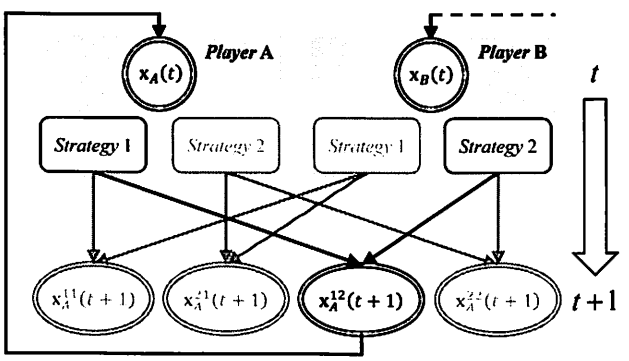

図1 内部状態の変化と戦略選択の関係

2.3 意思決定手法

ゲームにおいて，プレイヤ A は式 (1) の $\mathbf{x}_A^{ij}(t+1)$ をもとに，戦略 i，j を選択した場合の利得として式 (2) の $P_A^{ij}(t)$ を計算し，利得を最大化する戦略として，プレイヤ B に対する最適反応戦略（best-response strategy）を逐次採用する．本稿では，相手プレイヤの行動が完全に観測できない不完全観測（imperfect monitoring）のケース [10] については議論せず，完全観測が可能なケースに限定して論を進める．

3 数学的性質

2章で述べたモデルは，内部状態をもつゲームの一般形を定式化するためのものであった．本章では，このモデルに簡素化を施したいくつかのパターンについて，数学的性質を明らかにする．

3.1 2 × 2 の双行列ゲーム

3.1.1 モデル

本節では，各プレイヤの戦略数と内部状態数を，それぞれ2，1とした簡素なモデルについて数学的性質を述べる．なお，このモデルにおける利得関数 f^{ij} は，以下で示すような単純な線形結合とし，貢献度ベクトル v を省略することで利得 $P(t)$ と内部状態 $x^{ij}(t+1)$ を等価として扱う．

$$\begin{aligned} P_A^{ij}(t) &= x_A^{ij}(t+1) = c_A^{ij} x_A(t) + d_A^{ij} x_B(t) \\ P_B^{ij}(t) &= x_B^{ij}(t+1) = d_B^{ij} x_A(t) + c_B^{ij} x_B(t) \end{aligned} \tag{3}$$

式 (3) を戦略 i，j による状態変化と捉えて行列表現すると，式 (4) のように表すことができる．

$$\begin{pmatrix} x_A^{ij}(t+1) \\ x_B^{ij}(t+1) \end{pmatrix} = \begin{pmatrix} c_A^{ij} & d_A^{ij} \\ d_B^{ij} & c_B^{ij} \end{pmatrix} \begin{pmatrix} x_A(t) \\ x_B(t) \end{pmatrix} \tag{4}$$

式 (4) より，式 (3) の内部状態に対する係数部分は，内部状態を線形結合させる 2×2 の係数行列と読み替えられ，これを内部状態遷移行列と定義する．

次に，本モデルを戦略選択の観点から捉えると，プレイヤ A，B はそれぞれ2つずつ戦略を有しているため，内部状態遷移行列は $(1,1)$，$(1,2)$，$(2,1)$，$(2,2)$ の4要素をもつ行列となる．したがって，時刻 t における内部状態 x_A，x_B と戦略 i，j から導かれるすべての $x_A^{ij}(t+1)$ は式 (5) のように表され，$\mathbf{C}$ を自己反映行列，$\mathbf{D}$ を他者反映行列と定義する．

$$\begin{aligned} &\begin{pmatrix} x_A^{11}(t+1) & x_A^{12}(t+1) \\ x_A^{21}(t+1) & x_A^{22}(t+1) \end{pmatrix} \\ &= \begin{pmatrix} c_A^{11} & c_A^{12} \\ c_A^{21} & c_A^{22} \end{pmatrix} x_A(t) + \begin{pmatrix} d_A^{11} & d_A^{12} \\ d_A^{21} & d_A^{22} \end{pmatrix} x_B(t) \end{aligned} \tag{5}$$

以上より，初期状態 $x_A(0)$，$x_B(0)$ が決まり，全ての戦略対 i，j に対して内部状態遷移行列が決まれば，順次 $x_A(t+1)$，$x_B(t+1)$ を決定し，それに対応した $P_A^{ij}(t)$，$P_B^{ij}(t)$ が決まり，表1で示したような利得表から時刻 t の最適戦略 i_{opt}，j_{opt} を決定することができる．

3.1.2 最適反応戦略と内部状態の関係

本モデルにおけるプレイヤ A の最適反応戦略 i_{opt} は，プレイヤ B のある戦略 j に対して，それぞれの自己の戦略から得られる利得の差分 δ_A を計算し，その正負によって決定できる．すなわち，プレイヤ B が戦略1（$j=1$）をとるとき，プレイヤ A の戦略1（$i=1$）で得られる利得から戦略2（$i=2$）で得られる利得を引いた結果が0より大きければ，$j=1$ の場合のプレイヤ A の最適反応は戦略1となり，差分が0未満であれば戦略2が最適反応となる．同様に，プレイヤ B が戦略2（$j=2$）をとる場合についても差分を計算する．これら2つの計算式を表したものが式 (6) と式 (7) である．

$$\delta_A^{j=1} = \left(c_A^{11} - c_A^{21}\right) x_A(t) + \left(d_A^{11} - d_A^{21}\right) x_B(t) \tag{6}$$

$$\delta_A^{j=2} = \left(c_A^{12} - c_A^{22}\right) x_A(t) + \left(d_A^{12} - d_A^{22}\right) x_B(t) \tag{7}$$

また，$\delta = 0$ を計算することによって $x_A x_B$ 平面上における最適反応戦略の境界線を求めることができ，内部状態と最適反応戦略を結び付けて議論することが可能となる．相手の戦略が1の場合（実線），2の場合（破線）の分布図，それぞれを重ね合わせたプレイヤ A の最適反応戦略の分布図を図2に示す．

プレイヤ B に関しても同様の手順で最適反応戦略の分布図を作成することが可能である．いま，プレイヤ B から見た戦略の分布図を図3とすると，両者の分布図を重ね合わせることで，時刻 t におけるゲームは平面 $x_A x_B$ 上の点 $Q(x_A(t), x_B(t))$ と

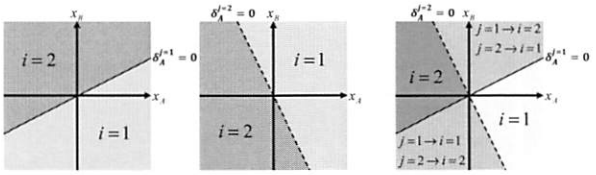

図 2　x_Ax_B 平面上におけるプレイヤ A の最適反応戦略

戦略変更境界線の領域問題として表現され，ゲームの性質を容易に求めることが可能となる（図 4）．

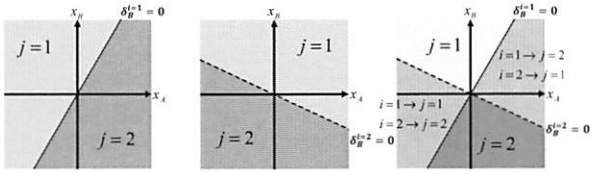

図 3　x_Ax_B 平面上におけるプレイヤ B の最適反応戦略

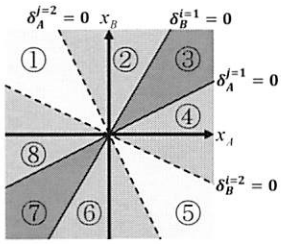

図 4　ゲームの分布図

図 4 のそれぞれの領域は，時刻 t におけるゲームの解のパターンを表しており，双方の戦略を (i,j) とすると，次の 4 つに分類することができる．

双方が支配戦略をもつ場合

プレイヤ A も B も相手の戦略によらず一意に最適反応決定できるため，ゲームは常にただ 1 つの解をもつ．

領域 1: $(2,1)$，領域 5: $(1,2)$

一方が支配戦略をもつ場合

プレイヤの一方は最適反応戦略を相手の戦略に依存しているが，相手プレイヤが支配戦略をもつため，相手プレイヤの支配戦略に対する最適反応が唯一の戦略となり，ゲームはただ 1 つの解をもつ．

領域 2: $(?,1)\to(2,1)$，領域 6: $(?,2)\to(2,2)$

チキンゲームに陥る場合

プレイヤ A も B も支配戦略を持たず，相手の戦略に対して 2 つの最適反応をもつ．ただし，どちらかが先に戦略を明示した場合，それに対する相手の最適反応戦略と明示した戦略が互いに最適反応になり，解が 2 つ存在する．

領域 3 : $i=1\to j=2\to i=1\to\cdots$

領域 3′: $i=2\to j=1\to i=2\to\cdots$

領域 7 : $i=1\to j=1\to i=1\to\cdots$

領域 7′: $i=2\to j=2\to i=2\to\cdots$

最適反応が完全に循環する場合

図 4 の例にはそのような領域は存在しないが，プレイヤ A も B も支配戦略を持たず，相手の最適反応戦略と自分の最適反応戦略が完全に循環してしまい，解を持たない場合がある．

例: $i=1\to j=1\to i=2\to j=2\to i=1\to\cdots$

本モデルは動的な利得によって決定される純戦略ゲームと定義することができるが，履歴情報や推論を用いず，時刻 t における利得のみで戦略を決定する場合，上記のように戦略を一意に決定できない場面が発生しうる．そのような場合，繰り返し囚人のジレンマ・ゲームでは意思決定を確率的に行う混合戦略ゲームに拡張することが一般的であり，ゲームにおけるナッシュ均衡（nash equilibrium）を求める問題に帰着する．

多数の戦略をもつ 2 人双行列ゲームにおいてナッシュ均衡を求める手法は，線形相補問題の解としてすでにアルゴリズムが提案されている [11]．しかし，ナッシュ均衡は必ずしもパレート最適（pareto efficient）でないことが証明されており [12]，n 人ゲームにおいて厳密にナッシュ均衡を求めることは非常に困難である．これはゲームの高次元化にしたがって計算が困難となることを意味し，内部状態数と戦略数の増加を考えたとき，厳密にナッシュ均衡解を求めることはリアルタイム性の観点から現実的ではない．したがって，最適反応戦略の循環やチキンゲームが発生した場合，本モデルでは各プレイヤは非ゼロ和ゲームにおけるマクシミン法（maximin）に従って戦略を決定するものとする．マクシミン法とは，相手プレイヤのある戦略に対して，自己が得られる利得の最小値が最大となる戦略を最適戦略として選択する手法である．

3.1.3　内部状態の収束と発散

3.1.2 節は，ある時刻 t のゲームに限定して数学的性質を考察したものであった．次に本節では，戦略 i，j について，ゲームが繰り返されたときに内部状態がどう変化していくかを考察する．

簡素化を施した本モデルにおける内部状態は，初項を $x_A(0)$，$x_B(0)$，公比を内部状態遷移行列とした等比数列である．したがって，$x_A^{ij}(t)$ を $x_A^{ij}(t-1)$ と $x_A^{ij}(t-2)$ から決まる隣接 3 項間の漸化式とみなし，特性方程式の 2 解をそれぞれ

$$\alpha_{ij}=\frac{\left(c_A^{ij}+d_B^{ij}\right)+\sqrt{\left(c_A^{ij}+d_B^{ij}\right)^2-4\left(c_A^{ij}d_B^{ij}-d_A^{ij}c_B^{ij}\right)}}{2}$$

$$\beta_{ij}=\frac{\left(c_A^{ij}+d_B^{ij}\right)-\sqrt{\left(c_A^{ij}+d_B^{ij}\right)^2-4\left(c_A^{ij}d_B^{ij}-d_A^{ij}c_B^{ij}\right)}}{2}$$

とおくと，戦略 i，j に対する x_A の一般項は

$$x_A^{ij}(t)=\frac{\alpha_{ij}^t}{\alpha_{ij}-\beta_{ij}}\left(x_A^{ij}(1)-\beta_{ij}x_A(0)\right)-\frac{\beta_{ij}^t}{\alpha_{ij}-\beta_{ij}}\left(x_A^{ij}(1)-\alpha_{ij}x_A(0)\right)$$

と求められる．ここで，$x_A^{ij}(1)$ は $x_A(0)$ と所与の c_A^{ij}，d_A^{ij} からただちに求められるため，初期状態から $\lim_{t\to\infty}x_A^{ij}(t)$ の収束や発散といった性質が容易に求められることがわかる．

例として，内部状態遷移行列が次に示すような任意の θ に対する回転行列や線型変換であるような場合について考える．

$$\begin{pmatrix} c_A^{ij} & d_A^{ij} \\ d_B^{ij} & c_B^{ij} \end{pmatrix}=\begin{pmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} c_A^{ij} & d_A^{ij} \\ d_B^{ij} & c_B^{ij} \end{pmatrix}=\begin{pmatrix} a & 0 \\ 1 & \tan\theta \end{pmatrix} \tag{9}$$

式 (4) について式 (8) が成り立つとき，戦略 i，j に対する x_A の一般項を求めると

$$x_A^{ij}(t)=\cos t\theta x_A(0)-\sin t\theta x_B(0)$$

となり，確かに原点を基準とした回転移動の座標変換式になっていることが確認できる．また，式 (9) についても同様にして x_A の一般項を求めると

$$x_A^{ij}(t)=a^tx_A(0)+\frac{a^t-1}{a-1}\tan\theta x_B(0)$$

が得られ，とくに $|a|<1$，$\tan\theta>0$ のとき

$$\lim_{t\to\infty} x_A(t) = \frac{\tan\theta}{1-a} x_B(0)$$

となり，ゲームを繰り返し行ったとき，プレイヤ A の内部状態は正の値に収束する．

$x_A(0)=5.0$，$x_B(0)=5.0$ として，式 (8) と式 (9) の条件を満たす任意の θ のもとでゲームを繰り返し行った結果を図 5 に示す．

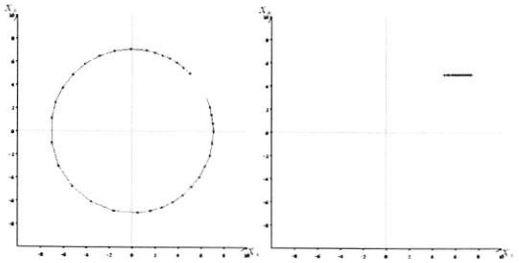

図 5　回転行列，線形変換行列を内部状態遷移行列とした場合の点 $Q(x_A(t), x_B(t))$ の軌跡

以上より，内部状態遷移行列はゲームの振る舞いを決定するうえで重要であり，かつ，よく制御された行列を用いれば，ゲームは発散せずに振動や収束という性質をもつことが示された．

3.2　高次元空間および非線形利得関数への応用

3.2.1　拡張モデル

3.1 節で述べた例は，提案モデルを最も単純な形式で用いた場合の数学的性質であった．本節では，3.1 節のモデルを，内部状態数の追加，利得関数の非線形化，バイアス項（定数項）の追加という 3 点について拡張を施した式 (10) のモデルについて数学的性質を考察する．

$$\begin{aligned} &\mathbf{x}_A = (x_{A1}, x_{A2})^{\mathrm{T}} \\ &P_A = \mathbf{v}_A \cdot \left(\mathbf{c}_A^{ij} \ln \mathbf{x}_A + \mathbf{d}_B^{ij} \ln x_B + \boldsymbol{\gamma}_A^{ij} \right) \end{aligned} \tag{10}$$

3.2.2　3 次元戦略分布図と内部状態

式 (10) における $\ln \mathbf{x}_A$，$\ln x_B$ は，内部状態と利得の関係に非線形性を持たせるための関数であり，国家間の戦争ゲームに当てはめるならば，自国の人口の増加が必ずしも国家全体の利得につながらないような状態を意味している．これを 3.1.2 節の戦略分布図の観点から考察すると，この拡張は戦略変更境界線 δ の非線形化に相当することがわかる．同様に，各内部状態に対するバイアス値 $\boldsymbol{\gamma}$ は，戦略分布図における境界線の切片に相当する．したがって，バイアス値 $\boldsymbol{\gamma}$ は，内部状態に対して各プレイヤが 2 つの戦略のうちどちらを選択しやすいかを表す事前情報であり，同時に，パラメータに対する下限値と解釈することができる．

以上から，内部状態数や戦略数が増加した場合についても 3.1 節のモデルとほぼ同様の性質を持ち，戦略の分布を求めることが可能である [13]．

4　実験

3 章で議論した 2×2 の双行列ゲームについて，作成した図 6 のシミュレータを用いて実験を行う．

4.1　実験モデルの定義

実験モデルでは，完全な自己成長が可能な条件下で，相手の影響がわずかにある場合を検証する．各プレイヤの他者反映行列について，図 7 に示すように戦略の組み合わせによって内部状態に正の影響を受ける場合と負の影響を受ける場合に場合分けする．すなわち，相手プレイヤは自己の成長を助ける戦略をとるか，成長を打ち消すような戦略をとることができる．ただし，ゲームは繰り返し行われるため，$\mathbf{C}\simeq 1$，$\mathbf{D}\simeq 0$ とし，シ

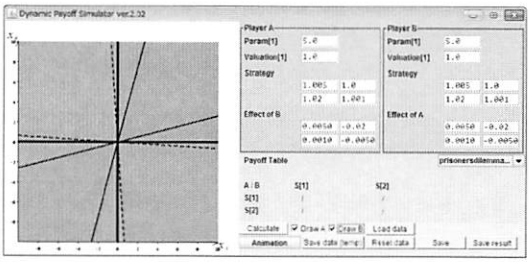

図 6　2×2 双行列ゲームシミュレータ

ミュレーション回数を 5000 回とすることで，擬似的な無限繰り返しゲームとする．

このようなモデルは，実世界における領地の獲得競争などに当てはめて考えられる．各プレイヤは，自己の領地の大きさに比例してより多くの領地を獲得することができ，相手の協力を得ることでより領地を獲得することができる．しかし，相手が領地を収奪する戦略をとった場合，自己が獲得した領地は相手によって相殺される．

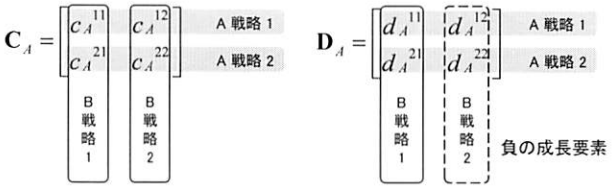

図 7　プレイヤ A から見た戦略の定義

この条件のもと，各プレイヤの内部状態遷移行列と初期状態を式 (11) のように定義した．

$$\begin{aligned} &x_A(0) = 6.0 \\ &x_B(0) = 5.0 \\ &\mathbf{C}_A = \begin{pmatrix} 1.002 & 1.005 \\ 1.007 & 1.004 \end{pmatrix}, \mathbf{D}_A = \begin{pmatrix} 0.004 & -0.005 \\ 0.006 & -0.010 \end{pmatrix} \\ &\mathbf{C}_B = \begin{pmatrix} 1.006 & 1.001 \\ 1.005 & 1.007 \end{pmatrix}, \mathbf{D}_B = \begin{pmatrix} 0.005 & -0.006 \\ 0.007 & -0.007 \end{pmatrix} \end{aligned} \tag{11}$$

図 8 は，式 (11) から求められたプレイヤ A，B の最適反応戦略の分布図と，ゲームの分布図である．

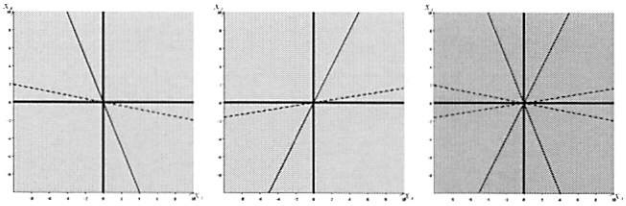

図 8　プレイヤ A，B の戦略分布図とゲームの分布図

4.2　ゲームの振る舞い

本モデルの最も重要な性質は，自己の内部状態と利得に相手の内部状態が反映されるという点にある．そこで，前節で定義した実験モデルをもとに，双方が相手に負の影響を与える戦略 2 の場合の他者反映行列（d_A^{22}，d_B^{22}）を変化させたとき，ゲームの振る舞いがどのように変化するかを検証する．図 9 は，d_A^{22} と d_B^{22} を -0.02 から 0 まで変化させたときのゲームの収束点を x_A と x_B の比率で表したものである．ただし，煩雑化を避けるため，図 9 に掲載する結果はゲームの振る舞いを考察するうえで重要な性質をもつ部分のみとした．なお，図中の不連続な点はプレイヤ A，B のどちらかが死滅し，ゲームが終了したことを意味している．

図 9 の結果から，相手プレイヤが戦略 2 を採用した場合の自己に対する負の成長要素が増大するにつれて，内部状態の収束

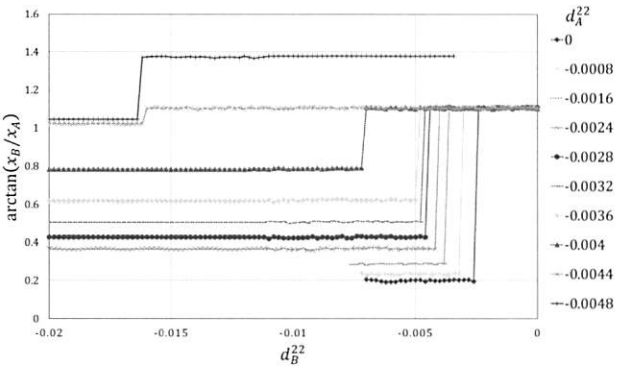

図 9　d_A^{22}，d_B^{22} を変化させたときのゲームの収束点

点が不利になることがわかる．さらに，収束点の変化は連続的ではなく，ある一定値を超えたときに急激に変化していることも見て取れる．

次に，実験結果からゲームの長期的な振る舞いのパターンを考えると，次の 6 つのパターンが発生していることがわかった．

① プレイヤ A が死滅する場合
② プレイヤ B が死滅する場合
③ 双方が戦略を変更せずに内部状態が発散する場合
④ 動的な均衡を保ちながら戦略変更境界線上を発散する場合
⑤ 動的な均衡を保ちながら戦略変更境界線上を縮退（双方の内部パラメータが 0 に向かって収束）する場合
⑥ 双方の内部状態が停留し，ゲームが不動状態に陥る場合

図 10 に，これらのパターンの例を掲載する．

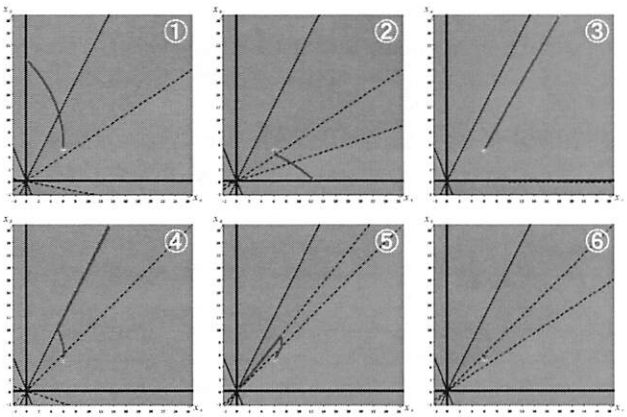

図 10　実験モデルを用いたゲームの振る舞いのパターン

図 10 を照らし合わせて図 9 の結果を分析すると，右上の不連続点はプレイヤ A が死滅するパターンであり，左下の不連続点はプレイヤ B が死滅するパターンであることがわかる．パラメータの変化によらず収束点が一定を保っている領域は，図 10 の ③，④，⑤ のいずれかのパターンであると考えられる．いま，戦略変更境界線上を推移している ④，⑤ のパターンに注目すると，内部状態が縮退する戦略と成長する戦略が無数に繰り返されていることが見て取れる．実験モデルにおける負の成長要素は相手プレイヤが戦略 2 を採用することによってに発生するため，内部状態が縮退（0 に向かって収束）するのは，一方あるいは両方のプレイヤが戦略 2 を採用した場合である．④ と ⑤ のケースでは，いずれも負の成長による縮退と協調による成長を繰り返しており，内部状態の遷移によって最適戦略の性質が交互に入れ替わることで境界線上を縫うようにゲームが展開される．このとき，各戦略による内部状態の成長が縮退を上回れば，内部状態は動的な均衡を保ちながら境界線を正の無限大へ発散していき，縮退がより大きければ 0 に向かって収束する．一方，③ のケースでは，成長と縮退の繰り返しは発生しておらず，ゲーム開始から終了まで成長のみを繰り返している．この場合も戦略の変更は発生しうるが，ゲームの結果に決定的な影響を与えるのは，最終的にどの領域で発散したかである．これは，内部状態が正の無限大に発散する状態遷移関数が複数存在するとき，関数によって x_A と x_B の発散速度は異なるため，繰り返し回数が有限である限りゲームの結果には必ず差異を生じることを意味している．ゲームの収束点を数学的に考察すると，本モデルを用いた無限繰り返しゲームの結果は，正の無限大に発散する場合，縮退する場合，プイレヤ A が死滅する場合，プレイヤ B が死滅する場合の 4 通りしか存在しない．しかし，実世界においてゲームが繰り返される回数は有限であり，内部状態遷移行列の違いによって，発散や収束速度にも大きな差異を生じる．したがって，極めて発散・収束速度の遅いケースは，有限回の繰り返しゲームのもとでは完全な不動状態であると考えることが可能である（⑥）．

最後に，パラメータごとの各戦略の割合を繰り返し回数で正規化したものを図 11，ゲームの振る舞いの分布を図 12 に示す．1 つの戦略がほぼ 100% を占めている発散領域は ③ のパターンを表し，複数の戦略の分布がある発散領域は ④ のパターンを表している．以上より，d_A^{22} と d_B^{22} の変化に対応したすべてのゲームの性質を把握することができる．

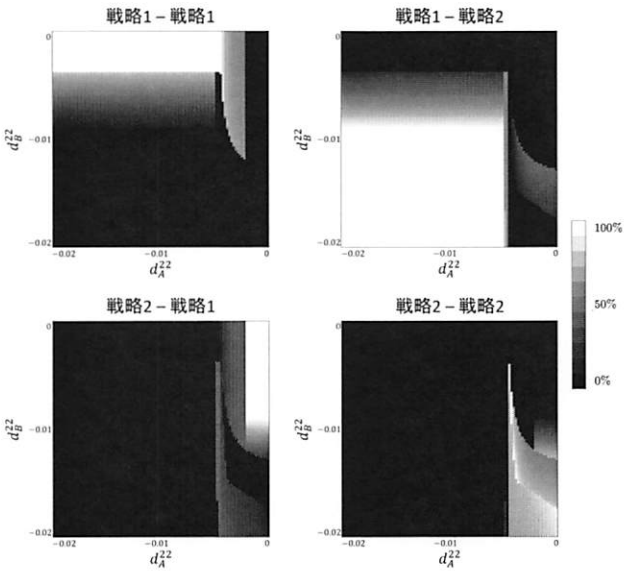

図 11　パラメータごとの各戦略の割合

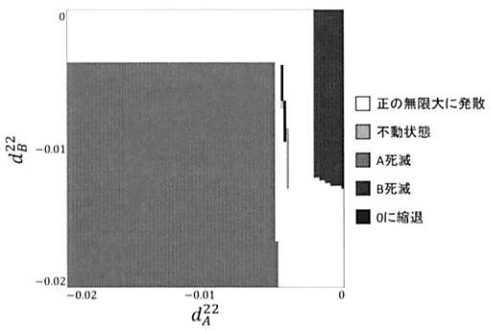

図 12　ゲームの振る舞いの分布

以上の実験結果より，内部状態遷移行列の違いによってゲームの振る舞いが異なることが明らかとなったが，著者らの先行研究 [9] により，初期状態の違いによってもゲームの振る舞いに差異が生じることがわかっている．また，本稿では言及しなかった自己反映行列に関しても，数値の組み合わせによってプレイヤの戦略や長期的なゲームの振る舞いに大きな影響を与えることもすでに明らかとなっている．

5 考察

実験結果より，プレイヤが内部状態をもつ意思決定モデルを導入することで，ゲームの利得やプレイヤの最適戦略が状況に応じて動的に変化し，状態の変化に応じた戦略をプレイヤが自動的に選択されていることが確認できた．また，2×2 の双行列ゲームについて，各プレイヤの内部状態と戦略の関係は x_Ax_B 平面の領域問題であり，時刻 t におけるゲームの解は x_A と x_B によって決定されるという数学的性質を確認し，意思決定のための条件のひとつを明確にした．

次に，図9の結果では，ある一定値を境としてゲームの収束点が遷移しており，プレイヤは最適反応戦略として2つの戦略を自由に選択できるが，一旦不利な状況に置かれたプレイヤは，自身の死滅を回避するため相手から負の影響を受ける状態でも相手に対して自己の内部状態が正の影響を与え，自身に対して不利な均衡がゲームにおける最良の内部状態となっていた．これを実世界に当てはめて考えた場合，一旦イニシアチブを握られたプレイヤが妥協や従属を余儀なくされ，結果的に両者の成長が維持されていたと解釈することができる．また，ゲームが縮退するケースでは，不利な状況におかれたプレイヤが妥協することによって，自身の成長は見込めないものの，共倒れを狙っている戦略と解釈することができる．実験によって発生したこれらの結果は，実世界で生じる駆け引きを本モデルがよく説明していると言え，協調を導くためにパラメータを組み込む手法を用いた松原らの研究 [8] を包含するものである．また，このようなゲームの発散や縮退に加えて，不動状態の存在や図5のような結果は，自己の利得と内部状態が相手プレイヤの内部状態と戦略に依存する不安定な場合であっても，内部状態遷移行列の調整によってゲームの均衡を作り出すことが可能であることを示している．これは，将来的にプレイヤにより有利な状況を作り出すため，プレイヤ自身による自律的な戦略の調整が行える可能性を示唆している．

一方で，同じ最適反応戦略を用いる場合でも，内部状態遷移行列と内部状態の違いによって，ゲームの結果に著しい差異が生じることが確認された．この結果は，ゲームの経過の違いによる解の領域の更なる考察と，戦略が一意に決定できない場合の履歴情報の利用や将来的利得の予想の必要性を示唆している．とくに，本稿では最適反応戦略とマクシミン戦略をある時刻 t のゲームに限定して決定していたが，今後は将来利得や長期的な割引率を考慮した繰り返しゲーム本来の最適戦略について検討する必要がある．

最後に，知的 NPC の意思決定手法への本モデルの適用を考えたとき，数値化されたパラメータによって適切な意思決定を行えることは，個々のビデオゲームの性質に依存せず，汎用的かつ統一的な意思決定手法を構築するという点において非常に有効であると考えられる．しかし，本来知的 NPC は本稿で論じた意思決定手法のほかに，知識表現と世界表現との連携が必要である [2]．したがって，より精度が高く，より人間らしい（例えば，失敗を犯し，失敗から学習するような）意思決定を行うためには，限定合理性，時間制約を持ったインクリメンタルな推論，結合周辺環境の組み込みが必要となる．特に，本研究において，ゲームは自己と同等の戦略判断能力を有する行動主体との二者間に限定していたが，相手を周辺環境としてとらえる手法や，追加情報として周辺環境を利得関数に組み込む手法について今後さらに議論する必要がある．また，利得関数の複雑化や状態数・戦略数の増加に伴って，数学的性質を明らかにすることが困難になることが予想され，高次元のゲームを制御可能とする手法についても同様に議論が必要である．

6 結論

本稿では，知的 NPC への応用を前提として，プレイヤの意思決定をゲーム理論によって実現するモデルを提案し，検証した．初めに，内部状態をもつプレイヤが行うゲームを定式化し，プレイヤ A とプレイヤ B がもつ戦略と状態からどのように利得が導き出されるか明らかにした．さらに，意思決定段階において，プレイヤの状態について漸化式を解くことによってプレイヤのもつ各戦略について，将来的利得とゲームの帰結も容易に計算できることを示した．以上をもとに，提案モデルを状態数1，戦略数2に限定し，利得の計算を簡素化した 2×2 の双行列ゲームについて数学的性質を考察した．この結果，プレイヤの最適反応戦略はプレイヤ A，B の内部状態 x_A，x_B と戦略の組み合わせによって決定される x_Ax_B 平面上の領域問題であることが明らかとなった．これらを踏まえて，2×2 の双行列ゲームについて，作成したシミュレータを用いて実験を行った．実験の結果，内部状態から合理的に戦略を決定し，状態に即した意思決定を行えることが示された．また，長期的なゲームの振る舞いについて，パラメータの調整によってゲームの結果を操作することが可能であり，協調戦略によって均衡がもたらされる場合や，囚人のジレンマ・ゲームに陥る場合など，一般的な非協力・非ゼロ和ゲームの特性が本手法にもよく表れていることが確認できた．

今後のこの研究は以下の点で拡張が望まれる．

- より多くの状態数・戦略数への拡張
- 履歴情報・推論の利用
- 周辺環境の組み込み
- 協調戦略を含む繰り返しゲームとしての戦略の最適化

参考文献

[1] 情報処理 特集 ゲーム情報学, 第53巻, pp. 100–152. 一般社団法人情報処理学会, 2012.

[2] 三宅陽一郎. 次世代ゲーム AI アーキテクチャ 2012, SQUARE ENIX オープンカンファレンス, 2012.

[3] John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior. *Bull. Amer. Math. Soc*, Vol. 51, pp. 498–504, 1945.

[4] 梅原嘉介. 対立と協調の経済学—「進化ゲーム理論」による「社会的ジレンマ問題」への処方箋. 工学社, 2009.

[5] Eizo Akiyama and Kunihiko Kaneko. Dynamical systems game theory and dynamics of games. *Physica D: Nonlinear Phenomena*, Vol. 147, No. 3, pp. 221–258, 2000.

[6] Eizo Akiyama and Kunihiko Kaneko. Dynamical systems game theory ii: A new approach to the problem of the social dilemma. *Physica D: Nonlinear Phenomena*, Vol. 167, No. 1-2, pp. 36 - 71, 2002.

[7] Eizo Akiyama. Dynamics of coupled players and the evolution of synchronous cooperation - dynamical systems games as general frame for systems interrelationship, 2007.

[8] 松原繁夫, 横尾真. 繰り返しゲームにおいて協調行動を生成する先読み型行動選択方法. 人工知能学会誌, Vol. 12, No. 6, pp. 881–890, nov 1997.

[9] 星聡, 藤田悟. 動的な状態管理を行う無限繰り返しゲーム. In *Joint Agent Workshops and Symposium 2013*, 2013.

[10] Olivier Compte. Communication in repeated games with imperfect private monitoring. *Econometrica*, Vol. 66, No. 3, pp. pp. 597–626, 1998.

[11] Carlton E Lemke and Joseph T Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial & Applied Mathematics*, Vol. 12, No. 2, pp. 413–423, 1964.

[12] 鈴木光男. 新装版 ゲーム理論入門. 共立出版, 2003.

[13] 星聡, 藤田悟. 複数の内部状態を管理する2人ゲームの数理モデル. 第76回情報処理学会全国大会, 2014.