2F2-02

囚人のジレンマの戦略における進化，学習，発生の相互作用

Interactions among Evolution, Learning and Development
in the Strategies for the Prisoner's Dilemma Game

小川行政　　有田隆也
Yukimasa Ogawa　Takaya Arita

名古屋大学　大学院情報科学研究科
Graduate School of Information Science, Nagoya University

Evolution, learning and development are typical adaptation mechanisms that allow organisms to adapt to an environment on different time scales. This paper investigates the relationship among these mechanisms using a computational model for the iterated prisoner's dilemma game, in which evolution, learning and development are expressed by a genetic algorithm, the Meta-Pavlov and a Turing machine, respectively.

1. はじめに

生物の代表的な適応機構として進化，学習，発生があり，それらが相互に作用しながら生命システムを作り上げてきた．その複雑な相互作用に対して，計算論的なモデルに基づく構成的アプローチが盛んになってきた．進化と学習の相互作用に関しては，Baldwin 効果 [Baldwin 96] が重要である．これは，ラマルク的な獲得形質の遺伝の仕組みがなくても，集団における個体の学習が集団全体の進化に方向性を与え，進化のスピードを促進するというものである．Baldwin 効果については，Hinton と Nowlan の先駆的実験 [Hinton 87] 以来多くの研究がなされてきた．特に，鈴木らは動的な環境においても Baldwin 効果が観察されることを確認した [Suzuki 04]．発生をも組み込んだ研究としては，Downing は Hinton らのモデルに発生過程を導入し，Baldwin 効果が生じる過程における，発生過程の進化と問題空間の大きさや解の特徴について論じている [Downing 04].

本研究は，最適解が固定していない動的な環境における進化，学習，発生の相互作用に関して知見を得ることを目的とする．動的な環境は大きく二つに分けることができる．一つは，集団が置かれた環境自体が世代を通して変化し，集団中の個体の適応度に影響を与える場合である．もう一つは，動的な要因を集団自体が内包しているような場合である．本研究では，後者の典型的な例として，繰り返し囚人のジレンマゲームを採用し，そこにおける進化，学習，発生の相互作用に関して検討する．

2. 繰り返し囚人のジレンマゲーム

繰り返し囚人のジレンマゲームは，2 人非ゼロ和ゲームの一種で，利己的集団における協調行動の創発に関して数多くの研究がなされている．ゲームは表 1 に代表される利得行列を用いて以下の手順で行われる．

二人のプレイヤーは協調，または裏切りのどちらかの手を同時に出し，出した手に応じて，利得行列 (表 1 参照) から両者は得点を得る．この対戦を繰り返し行い，合計得点を競う．

3. モデル

本研究では，進化を遺伝的アルゴリズム，学習をメタ・パブロフ学習，発生をチューリングマシンで表現したモデルを構築する．モデルの概要を図 1 に示す．発生過程では，遺伝子型から解読したチューリングマシンの遷移ルールとテープを用いて表現型を生成する [Downing 04]．そして生成された表現型を繰り返し囚人のジレンマゲームの戦略として定義して，メタ・パブロフ学習を用いて繰り返し囚人のジレンマゲームを行う．

表 1: 囚人のジレンマゲームの利得行列

自分の手＼相手の手	協調 (C)	裏切り (D)
協調 (C)	(3, 3)	(0, 5)
裏切り (D)	(5, 0)	(1, 1)

(自分の得点，相手の得点)

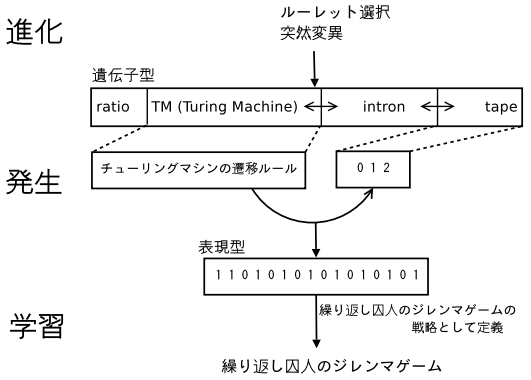

図 1: モデルの概要

3.1 遺伝子型

遺伝子型は，ratio，TM (Turing Machine)，intron，tape から構成される．

ratio は，TM，intron，tape の長さの比率が，それぞれ t ビットで格納されている．例えば TM，intron，tape の比率をそれぞれ $r1$，$r2$，$r3$ としたとき，$(r1, r2, r3) = (8, 5, 7)$ なら，TM は $N - 3k$ ビットの $8/(8+5+7) = 40\%$である (N は遺伝子のビット長)．

TM は，(s, x, s^*, x^*, a) の組からなる遷移ルールが格納されている．s は現在の TM のヘッドの状態，x は現在の tape のデータ，s^* は次の TM のヘッドの状態，x^* は次の tape のデータ，a はアクション (x^* を x に上書き，または x^* を x の右側

連絡先: 小川行政，名古屋大学大学院情報科学研究科，〒 464-8601 名古屋市千種区不老町，e-mail: ogawa@create.human.nagoya-u.ac.jp

に挿入) である．またヘッドの状態 s，s^* を p ビット，tape のデータ x，x^* を n ビットで表現するとアクションは 1 ビットで表現できるので，1 つのルールは $2p+2n+1$ ビットで表現することができる．ただしデータは 0〜2^n-1 の整数で表現されるが，等しい確率で 0，1，2 の数字が割り当てられる．

intron は，TM と tape を分離するために存在し，発生，学習には関係しない．

tape は，TM の初期の tape が，一つのデータ n ビットで格納されている．TM のデータと同様に，一つのデータには等しい確率で 0，1，2 の数字が割り当てられる．

3.2 発生

遺伝子型から解読した TM の遷移ルールと tape から表現型を生成する．ヘッドの動きは常に左から右に移動し，ヘッドの位置が tape の右端に達したら左端に戻るものとする．図 2 の手順をテープの長さが L に達するか，またはアクション回数が $max-devp-step$ に達するまで繰り返す．

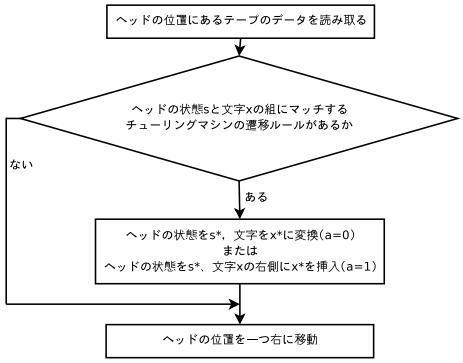

図 2: 発生過程

3.3 表現型

発生によって生成された表現型を，Lindgren のモデル [Lindgren 91] と同様に履歴に依存して次回の手を決定する戦略として定義する．記憶長 m の戦略は裏切りを 0，協調を 1 として以下のような 2 進数で表された履歴 h_m を持つ．

$$h_m = (a_{m-1}, \cdots, a_1, a_0)_2$$

a_0 は前回の相手の手，a_1 は前回の自分の手，a_2 は前々回の相手の手，$\cdots$ を表している．

ある履歴 k に対して，次回出すべき手を A_k とすると記憶長 m の表現型は，裏切りを 0，協調を 1，可塑性を持つ表現型を 2 と表現することで以下のように表すことができる．

$$S = [A_0 A_1 \cdots A_n - 1] \quad (n = 2^m)$$

3.4 学習

可塑性を持つ表現型は，対戦中にその表現型を用いた結果に応じて変更される．ここで鈴木らのモデル [Suzuki 04] と同様なメタ・パブロフ学習行列を表 2 に定義する．この学習行列は，プレイヤーの得点が相対的に高ければ変更せず，逆に小さければ変更するという強化学習の原理に基づくものであり，学習規則としてシンプルかつ典型的なものである．メタ・パブロフ学習による手の決定を以下に示す．

表 2: メタ・パブロフ学習行列

自分の手＼相手の手	協調 (C)	裏切り (D)
協調 (C)	C	D
裏切り (D)	D	C

- 繰り返し対戦を行う前，発生過程によって生成された表現型が可塑性を持つ場合，可塑性を持つ表現型を 0 か 1 に置き換える．
- 表現型と履歴を参照し対戦を行い，用いた表現型が可塑性を持つ場合，その表現型を対戦結果に対応するメタ・パブロフ学習行列の値と置き換えたものを新たな表現型とする．
- 次回対戦以降，新たな表現型を参照し手を決定する．

3.5 繰り返し対戦

以上のような個体同士でノイズありの繰り返し対戦を行う．ノイズとは，繰り返し対戦において，各個体の出すべき手が一定の確率で反転してしまうことである．

本研究で用いられる戦略が手を決定するためには履歴が必要である．そこで，各繰り返し対戦の一番はじめは，繰り返し対戦ごとにランダムに対戦履歴を作成する．

繰り返しゲームを行う状況として,「十分に長い間繰り返されるが，実際何回繰り返して行われるかはプレイヤーには分からない.」という設定にするため，繰り返しの回数は固定せず，対戦ごとに一定の確率で次回の対戦が行われるものとする．この確率を未来係数と呼ぶ．

また戦略における可塑的な表現型の初期値は，各繰り返し対戦ごとにランダムに 0，または 1 を割り当てる．

3.6 進化

上記のような繰り返し対戦を集団全体において総当たりで行い，その合計得点を各戦略個体の適応度とする．各適応度に応じてルーレット選択を行い，次世代の集団を生成する．その際，一定の確率で遺伝子のビットが反転する一点突然変異を導入する．

なお，計算量を軽減するために，はじめて行うカードの場合は，繰り返し対戦を 20 回行った平均得点を用いるとともに保存し，すでに行ったことのある対戦カードでは保存した得点を利用するものとする．また，保存した得点は 500 世代ごとに消去し，新たに計算し直すものとする．

4. 実験

4.1 設定

記憶長 4 の集団 (初期集団はランダム) において，パラメータとして遺伝子長 300，突然変異率 1/5000，個体数 1000，ノイズ率 1/25，未来係数 0.995，世代数 10000 を用いて 20 試行，進化実験を行った．また遺伝子型，発生過程で用いるパラメータを $t=5$，$p=3$，$n=5$，$L=16$，$max-devp-step=100$ とする．

ここで発生への依存度で個体をブループリント (blueprints) とレシピ (recipes) の 2 種類にわける．ブループリントとは 4 回以下のアクションで表現型を生成できる個体，一方レシピとは 5 回以上のアクションで表現型を生成できる個体と定義する．

4.2 進化の全般的傾向

実験結果を最終的に集団を占めた戦略の種類から図 3 のように分類した．最終的に集団を，ブループリントが占めた試行とレシピが占めた試行の大きく 2 つに分類することができた．また最終的に集団をレシピが占めた試行は，可塑性の割合が高かった試行と低かった試行に分類でき，さらに低かった試行のなかで平均得点の違いから 3 つに分類することができた．実験結果の一例を上記の分類毎にそれぞれ図 4，図 5，図 6，図 7，図 8 に示す．

ここで図中の平均得点とは，各世代に行われたすべての対戦の得点を平均したものであり，協調の度合として捉えることができる．可塑性の割合とは，個体の表現型における 2 のビットの割合を示し，各個体の学習依存度として捉えることができる．また今後，可塑性を持つ表現型 2 を x として表現する．

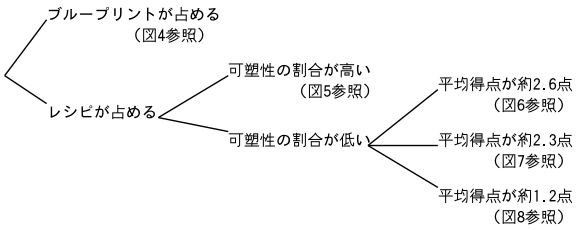

図 3: 試行の分類

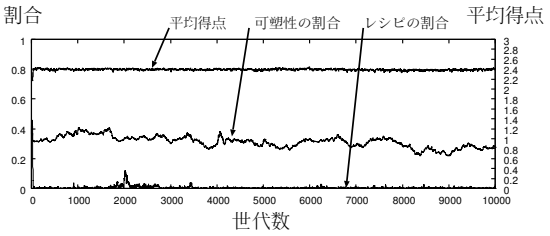

図 4: ブループリントが占めた試行

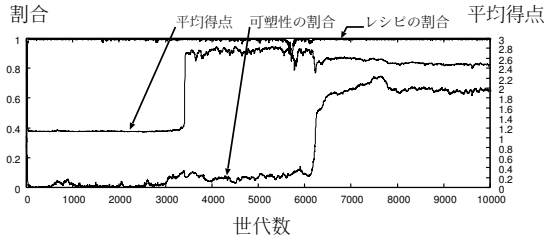

図 5: レシピが占めた試行 (可塑性の割合が高い)

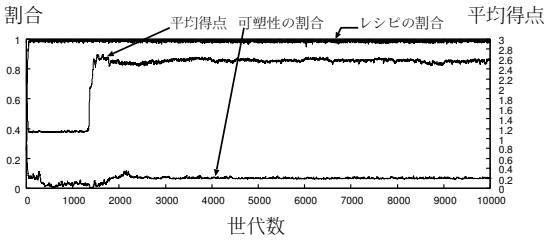

図 6: レシピが占めた試行 (平均得点が約 2.6 点)

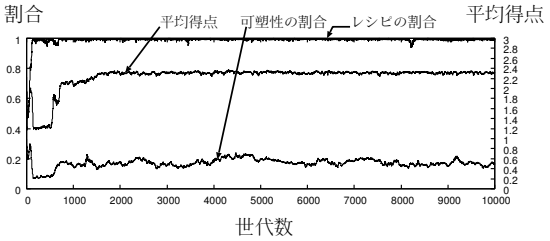

図 7: レシピが占めた試行 (平均得点が約 2.3 点)

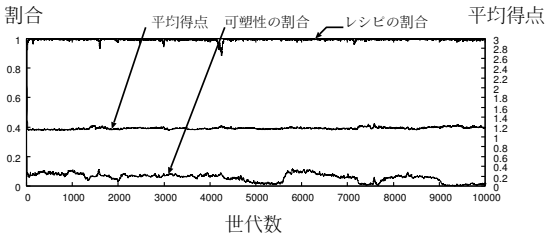

図 8: レシピが占めた試行 (平均得点が約 1.3 点)

レシピの割合とは，集団を占めるレシピの割合であり，集団の発生依存度として捉えることができる．

図 4 は最終的に，集団をブループリントが占めた試行である．はじめから平均得点が約 2.4 点，可塑性の割合が約 0.4 前後を推移し，集団をブループリントが占め，[x00x00x00x00001x] といった戦略を持つ個体が集団に多く存在した．全試行の中で最終的にブループリントが占めた試行は半数の 10 試行だった．図 4 のような進化を示したのが 5 試行，残りの 5 試行は一度 [0000000000001x01] といった 0 が連続する戦略を持つレシピに進化した後，ブループリントに進化する試行があった．なお進化と学習だけからなるモデルで実験を行った際，行った試行のすべてで図 4 のような進化の過程を示した．

図 5 は最終的に，集団をレシピが占め，可塑性の割合が高い集団に収束した試行である．全試行の中で 1 試行だけ確認した．はじめから約 2500 世代まで平均得点が約 1.1 点と低かったが，その後，急激に平均得点が約 2.8 点まで増加した．そして約 6300 世代から高い平均得点を維持しつつ，さらに可塑性の割合が急激に増加した．進化の概略を以下に示す (→ の左側は初期のテープ，右側は生成された表現型である)．

00101	→	0000000000000101
0x10001x0	→	011x101101011x01
100000x10	→	101101010101x101
100000x10	→	10xx0x0x0x0xx10x

平均得点が低かったとき，[0000000000000101] といった戦略を持つ個体が集団に多く存在した．これらの戦略を作成する TM の遷移ルール (現在のヘッドの状態，現在のデータ → 次のヘッドの状態，次のデータ，アクション) は $s_0, 0 \rightarrow s_0, 0, Insert$ だった．そして平均得点が増加するにつれて，[011x101101011x01] や [101101010101x101] といった戦略を持つ個体が集団に多く存在するようになった．これらの戦略を作成する TM 遷移ルールは $s_0, 0 \rightarrow s_0, 1, Insert$ であり，TM の遷移ルールの変化が急激な平均得点の増加の要因と考えられる．その後，可塑性の割合が高くなるにつれて，[10xx0x0x0x0xx10x] といった可塑的な戦略を持った個体が集団中に多く存在するようになった．これらの戦略を作成する TM の遷移ルールが $s_0, 0 \rightarrow s_0, x, Insert$ であり，TM の遷移ルールの変化が急激な可塑性の割合の増加の要因と考えられる．

図 6 は最終的に，集団をレシピが占め，可塑性の割合が低く平均得点が約 2.6 点といった集団に収束した試行である．全試行の中で 3 試行確認した．はじめから約 1300 世代まで平均得点が約 1.1 点と低かったが，その後急激に平均得点が約 2.6 点まで増加し収束した．また可塑性の割合は常に低い値を推移した．進化の概略を以下に示す．

00x	→	0000000100000000
00x	→	0101010101010101
01x	→	110x000000000001

平均得点が約 1.1 点と低かったとき，[0000000100000000] といった戦略を持つ個体が集団に多く存在した．その後，平均得

点が増加するにつれて [0101010101010101] といった戦略を持つ個体が集団に多く存在するようになった．そして最終的には [110x000000000001] といった戦略を持つ個体が集団を占め，安定した．図 6 の試行では tape はほとんど変わらず，TM の遷移ルールが複雑になることで進化した．最後に集団を占めた [110x000000000001] は，以下のようなチューリングマシンの遷移ルールから作成されていた．

$$
\begin{aligned}
s_0, 0 &\rightarrow s_1, 1, \mathit{Overwright} \\
s_0, x &\rightarrow s_2, 0, \mathit{Overwright} \\
s_1, x &\rightarrow s_0, 1, \mathit{Insert} \\
s_2, 0 &\rightarrow s_3, x, \mathit{Insert} \\
s_3, x &\rightarrow s_4, 0, \mathit{Overwright} \\
s_4, 1 &\rightarrow s_2, 0, \mathit{Overwright}
\end{aligned}
$$

図 7 は最終的に，集団をレシピが占め，可塑性の割合が低く平均得点が約 2.3 点といった集団に収束した試行である．全試行の中で 4 試行確認した．一度平均得点が上昇するが，約 1.2 点と低くなり約 700 世代から再び平均得点が増加し，約 1500 世代で約 2.3 点となり収束した．また可塑性の割合は常に低い値を推移した．進化の概略を以下に示す．

$$
\begin{aligned}
\text{x00x11100} &\rightarrow \text{000x000000011100} \\
\text{xx0x01101} &\rightarrow \text{xx0x010001000100} \\
\text{x101x0101} &\rightarrow \text{x100100x01000100}
\end{aligned}
$$

平均得点が約 1.2 点と低かったとき，[000x000000011100] といった戦略を持つ個体が集団に多く存在した．そして平均得点が増加するとともに [xx0x010001000100] といった戦略を持つ個体が集団に多く存在するようになった．最終的には [x100100x01000100] といった戦略を持つ個体が集団を占め，平均得点が約 2.3 点に収束し安定した．図 7 の試行では TM の遷移ルール，tape が変化することで進化した．最後に集団を占めた [x100100x01000100] は，以下のような TM の遷移ルールから作成されていた．

$$
\begin{aligned}
s_0, 0 &\rightarrow s_1, x, \mathit{Overwright} \\
s_1, 1 &\rightarrow s_1, 0, \mathit{Insert}
\end{aligned}
$$

図 8 は最終的に，集団をレシピが占め，可塑性の割合が低く平均得点が約 1.3 点といった集団に収束した試行である．全試行の中で 2 試行で確認した．はじめから最終世代まで常に平均得点と可塑性の割合は常に低い値を推移した．また集団に多く存在した戦略は [0000000000001110] であり．初期のテープは [01110]，TM の遷移ルールは，$s_0, 0 \rightarrow s_0, 0, \mathit{Insert}$ だった．

集団をレシピが占めたとき，必ず [0000000000001x01] といった 0 が連続する戦略を持つ個体へと進化した．これはチューリングマシンの遷移ルール $s_0, 0 \rightarrow s_0, 0, \mathit{Insert}$ があれば容易に生成できること，初期集団では裏切り的な集団が多く点数を搾取できることが原因だと考えられる．その後，図 5，図 6，図 7，図 8 に見られるように様々な戦略へと進化した．進化するとき，平均得点，または可塑性の割合が急激に増加する試行を確認した．これはレシピで突然変異が起こったとき，戦略が大きく変化することが原因と考えられる．そしてこのとき [011x101101011x01] といった協調的な戦略を持った個体が生成されやすい．その結果，[110x000000000001] といった戦略を持つ個体が集団を占めやすいと考えられる．

また最終的に集団を占めたブループリントとレシピの戦略を比べると、レシピはブループリントに比べ 1(協調) を持った戦略の方が多かった．一方、ブループリントはレシピに比べ 2(可塑性) を持った戦略が多かった。さらに最終的に集団中を占めた戦略の一つ一つの対戦に着目すると，レシピは 1（協調）という手を出すことで得点を獲得していた．一方，ブループリントは x（可塑性）という手を出すことで得点を獲得していた．

以上から，発生は一度裏切り的な戦略を持つ個体へと進化するので，進化の速度は遅くなるが，協調的な戦略を持つ個体を生成しやすく，協調的な集団への進化を促進させる傾向にあると考えられる．

4.3 ロバスト性

4.2 節の結果から，最終的に，集団を占めた戦略は平均得点が収束し安定しており，突然変異に対してロバスト性が高いと捉えることができる．しかし集団を占めた戦略がレシピの場合，戦略に 1(協調) を持つので，安定的な戦略とは考えにくい．そこで毎世代ランダムな遺伝子を持つ個体を挿入する実験を行い，集団のロバスト性を調べた．実験結果を表 3 に示す (r は挿入した個体の数)．表 3 は，最終世代において，どちらのタイプが過半数を占めるか，挿入する個体数 r ごとに 20 試行行った結果を示したものである．挿入するランダムな遺伝子を持つ個体数が増えるにつれて，最終世代にレシピが占める試行が減少し，ブループリントが占める試行が増加した．また $r = 10$ の場合，最終世代に集団をレシピが占めた試行で，$r = 0$ の場合に見られた [x100100x01000100] や [110x000000000001] といった戦略を持つ個体はほとんど見られなかった．以上の結果から，[110x000000000001] といった戦略を持つ個体は突然変異によって生成された個体に対するロバスト性は高いが，ランダムな遺伝子を持つ個体に対するロバスト性は低いと考えられる．

表 3: 最終世代に集団を占める個体毎の試行数

最終的に集団を占めた個体 ＼ 挿入数	0	1	5	10
ブループリント	10	12	13	17
レシピ	10	8	7	3

5. おわりに

本研究では，動的環境における進化，学習，発生の相互作用を解析するために，進化を遺伝的アルゴリズム，学習をメタパブロフ学習，発生をチューリングマシンで表現し，個体間の相互作用のみに依存した環境である囚人のジレンマの戦略進化モデルを構築し，適応機構間の相互作用を検討した．

実験の結果，行った試行の半数で発生の機構が有効に働くことがわかった．発生の機構が有効に働くとき，学習可能性は試行によって大きく異なるのに対し，あまり働かない場合は，学習が使われる可能性が高かった．また，発生は全般的に協調状態への進化を遅らせるが，協調的な戦略を持った個体による，突然変異に強い安定した協調を築く傾向にある．ただし，ランダムな戦略の侵入には弱いということが示された．

参考文献

[Baldwin 96] Baldwin, J. M.: A New Factor in Evolution, American Naturalist, Vol. 30, pp. 441-451, 1896.

[Hinton 87] Hinton, G. E. and Nowlan, S. J.: How Learning Can Guide Evolution, Complex Systems, Vol. 1, pp. 495-502, 1987.

[Suzuki 04] Suzuki, R. and Arita, T.: Interactions between Leaning and Evolution: The Outstanding Strategy Generated by the Baldwin Effect, Biosystems, Vol. 77, No. 1-3, pp. 57-71, 2004.

[Downing 04] Downing, K. L.: Development and the Baldwin Effect, Artificial Life, vol. 10, No. 1, pp. 39-63, 2004.

[Lindgren 91] Lindgren, K.: Evolutionary Phenomena in Simple Dynamics, Artificial Life II, pp. 295-311, 1991.