

多体学習のダイナミクス

佐藤 譲

理研・脳科学総合研究センター

1 Introduction

複数の主体が相互作用をしつつ行動様式を獲得していく状況を考察する。主体は人間でも生物でも機械でもよいが、内部構造はわからない。主体は利己的に行動を決定し、その行動に対して環境あるいは他の主体を介してフィードバックがおこる。このフィードバックに基づいて主体は行動様式を変化させていく¹。行動するたびにその主体に対する環境からの応答があり、主体が外界から得る情報はこの応答だけである。この情報に基づいて各主体は行動様式を修正し、再び行動を選択するという過程を繰り返す。

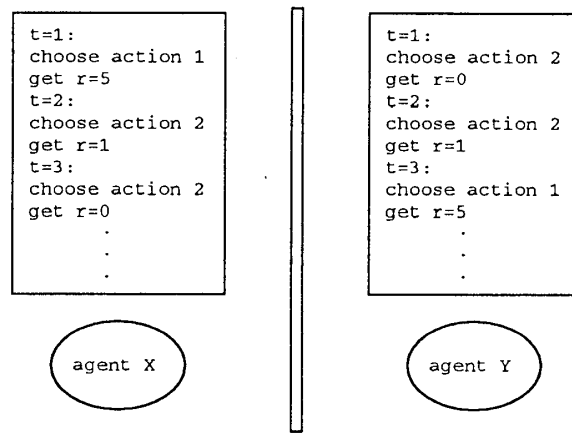


図 1: 相互適応: 相手の行動様式と環境全体の構造は未知。自分の行動選択に関する環境(相手)からの応答だけが、学習主体に与えられる情報である。

Fig. 1 では、ある行動に対するある時点での環境からの応答が与えられている。外

¹相互の意思決定という枠組みではゲーム論 [10] 全体が、この問題を考えているといってもいいし、また対話という文脈では [25] があるが、これらの研究では適応については陽に考察されない。相互適応ダイナミクスの先行研究としては例えば [15], [1], [14], [6], [22] 等を参照。

部からこれを眺める観測者は

$$\begin{pmatrix} (3,3) & (0,5) \\ (5,0) & (1,1) \end{pmatrix}$$

という有限双行列ゲームの構造をある時点において見いだすかもしれない²。しかし、主体自身は自分がおかれている環境や相手との関係についての知識をもたず、環境からの応答のみに基づいて全く利己的に行動様式を変化させている。本論では学習方程式として、ゲームダイナミクスが導かれるが、これは系全体の外部からの記述にあらわれるものであり、主体自身はゲーム論におけるプレーヤーのように、自分のプレイしているかもしれないゲームについての知識、あるいは環境の鳥瞰図は持っていない。(その意味でここでいう主体は限定合理的プレーヤー以下の存在である。)ここではこの状況の確率的な記述、つまり主体の行動確率のダイナミクスを考える。まずこのような視点の単純化に関して議論をしておく。

学習について基本的な問題のひとつは主体の内部構造の変化と外部的な行動の変化の関係を考えるということである。しかしこのモデルではこの問題を直接考察するのは難しい。ここでは理想化された多体学習系の振る舞いを調べることになるが、この場合には認知科学における行動主義に対するものと同様の批判が生じることになるだろう。行動主義は広く取れば「外部から測定可能な量だけに基づいて主体の認識構造を考察する」という立場である³。したがってここでの議論を例えば主体の内部構造のダイナミクスを含む形にまで拡張したとしても(例えば行動だけ切り取るのではなく、脳のダイナミクスを含めて考えたとしても)、この論点は依然残ることは強調されなくてはならない。さらにここで考察することは行動を介して相互作用する多体学習系の全体的な振舞いであって、これに基づいて主体の認識構造を直接考えようというものではないので、普通の意味での行動主義とは異なる。

次に確率的記述、とくに行動の選択確率の起源についてであるが、この問題には主観確率の問題に踏み込む以前にいくつかの解釈が可能である。ひとつは行動選択確率を行動の頻度と考えるというものである。この場合は行動系列が十分長い間観測されなければならない、さらに結果として安定な極限分布が存在しなくてはならない⁴。また行動様式といっても細かな文脈は全てつぶされてしまい、観測結果としてのヒストグラムについての分析しかできない。さらにフィードバックのプロセスを考えると、相互作用のダイナミクスに対して学習のダイナミクスが非常に遅いという特殊な状況設定が必要になる。もうひとつは多くの主体について調べた場合に関して、もっとも確からしい構造という意味でのアンサンブル平均に関する記述と考えるというものであ

²次のステップで2や-1といった表に無い応答が生じる可能性があるにもかかわらず、観測者がゲームの構造を断定しているということに注意。

³これは否定しがたい立場のようにも思えるが、この立場に対する反論としては例えば Chomsky の論 [2] がある。

⁴より強い仮定として、内部構造が決定論力学系であらわされると考えると、この力学系はエルゴード性でかつ強い混合性を持たなくてはならない。

るが⁵, 非対称な多体相互作用の時間発展を考える場合にはこの論理は破綻すると思われる. また生物集団の進化とは異なり, 学習は個体の適応という側面が強いので, 頻度としての解釈の方が自然であろう. 実時間でおこる速い学習についてはこの仮定は不当なものになるかもしれないが, 運動学習や生活習慣の獲得といった長期増強的な学習過程に関してはよい近似になる可能性がある. さらにこのような視点からのモデル研究により学習における主体間の情報の流れや学習の階層性を分析する枠組みが与えられると期待される.

以上の仮定をもとに議論を進める.

2 多体学習のダイナミクス

2.1 学習方程式の導出

まず離散時間の学習モデルを与える. 簡単のためにまず 2 体の主体 X, Y の相互作用について考える. 主体 X は N 個の行動が選択可能で, 主体 Y は M 個の行動が選択可能とする. 主体 X が行動 i ($i = 1, 2, \dots, N$) を選ぶ確率を $x_i(n)$ とする. n はタイムステップ. 状態ベクトルは $\mathbf{x}(n) = (x_1(n), x_2(n), \dots, x_N(n))$. $\sum_k^N x_k(n) = 1$. Y についても同様に $\mathbf{y}(n) = (y_1(n), y_2(n), \dots, y_M(n))$, $\sum_k^M y_k(n) = 1$ とする. この \mathbf{x}, \mathbf{y} の時間発展として学習過程を記述する.

$r_{ij}^X(n)$ を時刻 n で X, Y が行動 (i, j) を選んだときの環境からの X への報酬とし, $Q_i^X(n)$ を行動 i についての時刻 n での X の積算型記憶とする. 同様に $r_{ji}^Y(n), Q_j^Y(n)$ を与える.

$$\begin{aligned} Q_i^X(n+1) - Q_i^X(n) &= \frac{1}{T} \left[\delta_{ij}(n) r_{ij}^X(n) - \alpha_X Q_i^X(n) \right], \quad (i = 1, 2, \dots, N), \\ Q_j^Y(n+1) - Q_j^Y(n) &= \frac{1}{T} \left[\delta_{ij}(n) r_{ji}^Y(n) - \alpha_Y Q_j^Y(n) \right], \quad (j = 1, 2, \dots, M), \\ \delta_{ij}(n) &= \begin{cases} 1 & (X \text{ takes } i \text{ and } Y \text{ takes } j \text{ at step } n) \\ 0 & (\text{Otherwise}) \end{cases} \end{aligned} \quad (1)$$

ここで $\alpha_X, \alpha_Y \in [0, 1)$ は記憶の減衰率である. T は定数. 主体はこの記憶 Q に基づいて次のステップでの行動確率を変化させる. [15], [24].

$$\begin{aligned} x_i(n) &= \frac{e^{\beta Q_i^X(n)}}{\sum_k^N e^{\beta Q_k^X(n)}} \quad (i = 1, 2, \dots, N), \\ y_j(n) &= \frac{e^{\beta Q_j^Y(n)}}{\sum_k^M e^{\beta Q_k^Y(n)}} \quad (j = 1, 2, \dots, M), \end{aligned} \quad (2)$$

⁵“Behavior parameter” としての Nash の解釈 [9] も参照. ワンショットゲームにおける単プレイヤーの混合戦略 (!) をそのプレイヤーの選択可能な戦略木の空間的密度として捉える.

$\beta_X, \beta_Y \in [0, \infty]$ は学習率で、意思決定の鋭敏性をあらわす。 β が大きいとわずかな Q_i^X の違いに対して x_i の大きな差が割り当てられる。これは事象の発生確率がフィードバックにより修正されつつ進行する確率過程のモデルであり、基本的に強化学習のスキームに従う [4], [15], [24].

ここで学習過程が相互作用に比べて遅く、主体は行動パターンを変更するまでに、たくさんの試行錯誤を繰り返す (Fig. 2) として連続時間モデルを考える⁶.

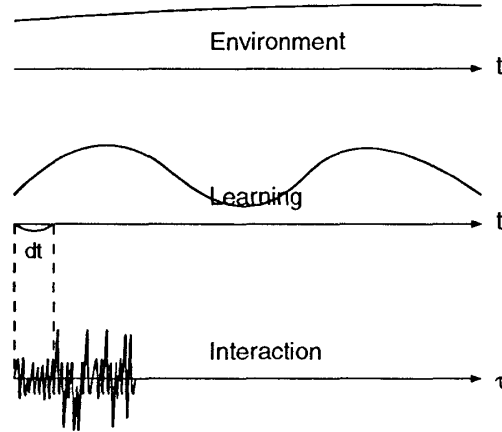


図 2: 学習のダイナミクスと行動のダイナミクス: 学習過程は相互作用に比べて遅い.

この場合相互作用のダイナミクスに対して、 \mathbf{x}, \mathbf{y} はほとんど一定と考えてよく、時刻 t での記憶 Q の時間発展は以下で与えられる.

$$\begin{aligned}\dot{Q}_i^X &= R_i^X - \alpha_X Q_i^X \quad (i = 1, 2, \dots, N), \\ \dot{Q}_i^Y &= R_i^Y - \alpha_Y Q_i^Y \quad (i = 1, 2, \dots, M).\end{aligned}\quad (3)$$

ここで R_i^X は X が行動 i を選んだときの Y の行動に対する dt の間に得る平均報酬. R_i^Y も同様. $\alpha_X, \alpha_Y \in [0, 1)$ は X, Y の記憶の減衰率. $Q(t)$ に基づく $x_i(t), y_i(t)$ の変化は以下で与えられる.

$$\begin{aligned}x_i(t) &= \frac{e^{\beta_X Q_i^X(t)}}{\sum_j^N e^{\beta_X Q_j^X(t)}} \quad (i = 1, 2, \dots, N), \\ y_i(t) &= \frac{e^{\beta_Y Q_i^Y(t)}}{\sum_j^M e^{\beta_Y Q_j^Y(t)}} \quad (i = 1, 2, \dots, M),\end{aligned}\quad (4)$$

$\beta_X, \beta_Y \in [0, \infty]$ は X, Y の学習率. Eq. (4) の微分から

$$\begin{aligned}\dot{x}_i &= \beta_X x_i (\dot{Q}_i^X - \sum_j^N \dot{Q}_j^X x_j) \quad (i = 1, 2, \dots, N), \\ \dot{y}_i &= \beta_Y y_i (\dot{Q}_i^Y - \sum_j^M \dot{Q}_j^Y y_j) \quad (i = 1, 2, \dots, M).\end{aligned}\quad (5)$$

⁶ 確率過程の連続時間モデルとそのダイナミクスについては [5] も参照. 競合的な確率過程に関しては [3] を参照.

を得るので, Eqs. (3), (4), と (5) から以下の学習方程式を得る.

$$\begin{aligned}\frac{\dot{x}_i}{x_i} &= \beta_X \Delta R_i^X + \alpha_X \Delta I_i^X, \\ \frac{\dot{y}_i}{y_i} &= \beta_Y \Delta R_i^Y + \alpha_Y \Delta I_i^Y,\end{aligned}\tag{6}$$

ここで

$$\Delta R_i^X = R_i^X - R^X, \quad \Delta I_i^X = I_i^X - I^X,\tag{7}$$

$$R^X = \sum_j x_j R_j^X, \quad I_i^X = -\log x_i, \quad I^X = \sum_j x_j I_j^X\tag{8}$$

であり, Y についても同様. Eq. (6) において,

- 第一項 (ΔR_i): 平均報酬 R と行動 i に対する報酬との差. β は学習率.
- 第二項 (ΔI_i): 行動選択の不確定性 I (Shannon エントロピー) と行動 i の情報量との差. α は記憶の減衰率.

である. $\Delta R_i^X > 0$ の場合 x_i の増加率は上昇し, 逆なら減少する. また主体のある時刻 t での行動 i の情報量と行動の Shannon エントロピー [21] の差 $-\Delta I_i^X$ に比例して x_i の増加率は減少し, 逆なら増加する. この項は記憶の減衰によって導入されるので, $\alpha_X > 0$ なら常に行動の不確定性を増大させる方向に作用する. とくに $x_1 = x_2 = \dots = x_N = 1/N$ (ランダム選択) の場合は行動 i を選んだという効果は消失し, 第二項は 0 になる. 以下では第一項を相互作用項, 第二項を記憶項とよぶ.

ここで環境変化が学習に比べて遅い場合を想定して, (i, j) のペアが決まったときの X, Y への報酬を $r_{ij}^X(x, y, t) = a_{ij}$ and $r_{ij}^Y(x, y, t) = b_{ji}$, ($i = 1, \dots, N, j = 1, \dots, M$) とおく. さらに各瞬間で主体の行動確率 \mathbf{x} と \mathbf{y} は独立であるとする, 行動 i に関する報酬の時間平均は

$$R_i^X = \sum_j^N a_{ij} y_j, \quad R_i^Y = \sum_j^M b_{ij} x_j.\tag{9}$$

この場合 Eq. (6) は

$$\begin{aligned}\frac{\dot{x}_i}{x_i} &= \beta_X [(A\mathbf{y})_i - \mathbf{x} \cdot A\mathbf{y}] + \alpha_X [I_i^X - I^X], \\ \frac{\dot{y}_i}{y_i} &= \beta_Y [(B\mathbf{x})_i - \mathbf{y} \cdot B\mathbf{x}] + \alpha_Y [I_i^Y - I^Y],\end{aligned}\tag{10}$$

に reduce される⁷. ここで $(A)_{ij} = a_{ij}$, $(B)_{ij} = b_{ij}$.

この場合 (A, B) を線形双行列ゲームのペイオフ・マトリクス, x, y をそれぞれ X, Y の混合戦略とみなすことができる [10]. 特に $\alpha = 0$ の場合 Nash 解の一つは (もし単体内部に存在すれば) Eqs. (10) の x -nullcline, y -nullcline の交点で与えられる. この不

⁷Eqs. (10) はとくに主体が完全な記憶を持つ ($\alpha_X = \alpha_Y = 0$) 場合, Multi-population replicator equation [23][12] になる.

動点は漸近安定ではないが、軌道の時間平均という意味を持つ [20]. 1 節で議論したように、主体は自分のプレイしているゲーム (A, B) の構造を知らない. ただ利己的に報酬を得ようとしているだけである. にもかかわらず, Eq. (9) の関係によってゲームダイナミクスが記述にあらわれることになる.

Eqs. (10) は一般に $\alpha_X = \alpha_Y = 0$ の場合保存系, さらに (A, B) がゼロサムゲームの場合には Hamilton 系となり [13], (A, B) が単体内部に Nash 解を持つ場合, 運動の定数は以下であたえられる.

$$\begin{aligned} H &= \frac{1}{\beta_X} D(\mathbf{x}^* || \mathbf{x}) + \frac{1}{\beta_Y} D(\mathbf{y}^* || \mathbf{y}) \\ &= \frac{1}{\beta_X} \sum_j^N x_j^* \log \frac{x_j^*}{x_j} + \frac{1}{\beta_Y} \sum_j^M y_j^* \log \frac{y_j^*}{y_j}, \end{aligned} \quad (11)$$

ここで $(\mathbf{x}^*, \mathbf{y}^*)$ は (A, B) の Nash 解, $D(p||q)$ は Relative entropy (Kullback-Leibler divergence) である. (A, B) がゼロサムの場合は H はダイナミクスのハミルトニアンになる. $D(p||q)$ は確率分布 p と q の間の情報理論的な距離である. H が運動の定数であるということは各主体の行動確率と Nash 解との情報理論的な距離の線形和が一定, ということの意味する. つまり $\alpha_X = \alpha_Y = 0$ で β_X, β_Y が finite positive の場合, Shannon の不等式より, 初期値が Nash 解で無い限り, ダイナミクスが Nash 解に到達することはない.

Eqs. (10) は $\alpha_X, \alpha_Y > 0$ の場合, 散逸率 $-[(N-1)\alpha_X + (M-1)\alpha_Y]$ の散逸系になる. Eqs. (10) において, (A, B) がじゃんけんゲームの場合に, ハミルトンカオス, ヘテロクリニックサイクル等多様なダイナミクスが出現することがわかっている [16], [17].

Eqs. (10) を 3 体系に拡張してみる.

$$\begin{aligned} \frac{\dot{x}_i}{x_i} &= \beta_X [(Ayz)_i - \mathbf{x} \cdot Ayz] + \alpha_X [I_i^X - I^X] \quad (i = 1, 2, \dots, N) \\ \frac{\dot{y}_j}{y_j} &= \beta_Y [(Bzx)_j - \mathbf{y} \cdot Bzx] + \alpha_Y [I_j^Y - I^Y] \quad (j = 1, 2, \dots, M) \\ \frac{\dot{z}_k}{z_k} &= \beta_Z [(Cxy)_k - \mathbf{z} \cdot Cxy] + \alpha_Z [I_k^Z - I^Z] \quad (k = 1, 2, \dots, L), \end{aligned} \quad (12)$$

ここで A, B, C はテンソル, $(A)_{ijk} = a_{ijk}$, $(B)_{ijk} = b_{ijk}$, $(C)_{ijk} = c_{ijk}$. 一般に K 体の学習系で $\alpha = 0$ かつ (A, B, C, \dots) が interior-Nash 解をもつ場合には軌道 $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots)$ の時間平均は Nash 解 $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \dots)$ に一致することがわかっている. 以下に 3 体のじゃんけんゲームにおける学習ダイナミクスのシミュレーション例を与える.

2.2 情報空間でのダイナミクス

ここでは多体学習における情報量のダイナミクスに関して考察する.

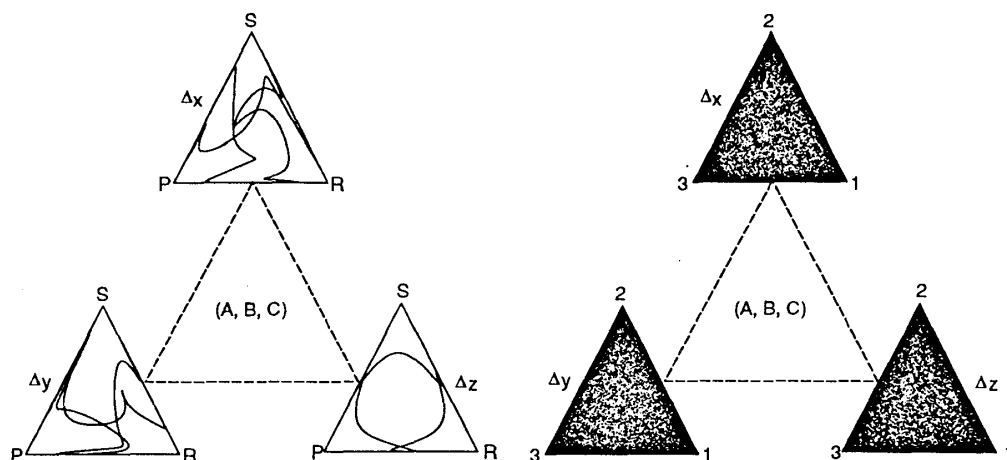


図 3: 3体学習ダイナミクスの例: (A, B, C) はじゃんけんゲームで相空間は R(Rock), S(Sissors), P(Paper) で張られる 3つの 2次元単体 $\Delta_X \times \Delta_Y \times \Delta_Z$. $a_{ijk} = \{ 2 (1人勝ち), 1 (勝者2人), -1 (敗者2人), -2 (1人負け), \epsilon_X (引き分け) \}$ で B, C についても同様. ゲームの Nash 解は単体の中心点で $\mathbf{x}^* = \mathbf{y}^* = \mathbf{z}^* = (1/3, 1/3, 1/3)$. パラメータは $\beta_X = \beta_Y = \beta_Z = 1$, $\alpha_X = \alpha_Y = \alpha_Z = 0.01$, $\epsilon_X = 0.5, \epsilon_Y = -0.365, \epsilon_Z = 0.8$ (左図), $\epsilon_X = 0.5, \epsilon_Y = -0.365, \epsilon_Z = 0.9$ (右図).

$\xi_i = -\log x_i, \eta_i = -\log y_i, \xi = (\xi_1, \dots, \xi_N), \eta = (\eta_1, \dots, \eta_M)$, として, 情報空間 (ξ, η) を導入すると, Eqs. (10) から

$$\begin{aligned} \dot{\xi}_i &= \beta_X [(Ae^\eta)_i - e^{\xi_i} \cdot Ae^\eta] - \alpha_X [\xi_i - e^{\xi_i} \cdot \xi] \quad (i = 1, 2, \dots, N), \\ \dot{\eta}_i &= \beta_Y [(Be^\xi)_i - e^{\eta_i} \cdot Be^\xi] - \alpha_Y [\eta_i - e^{\eta_i} \cdot \eta] \quad (i = 1, 2, \dots, M), \end{aligned} \quad (13)$$

がえられる. Eq.(13) は各行動の情報量の発展方程式を与える [18].

ここでの論点は多体学習系になったことによって情報の流れが双方向的になり, 単体系のように一方向的で固定した経路をもたなくなるということである. 各主体は環境から情報を得ようとするが, 他の主体も同様である. 情報の流れが不安定化する必要条件は主体間の関係が非推移的, すなわちじゃんけんのように R に対しては P が好ましく, P に対しては S が好ましく, S に対しては R が好ましい, といった構造をもっていることである⁸. おおざっぱにいうと相互学習がダイナミクスの不安定方向, 記憶

⁸これはダブルバインド的構造であるといってもいいかもしれない. Bateson のダブルバインド理論 [1] に固有の implicit/explicit の階層構造は, ここでは確率的記述によってつぶれてしまっているが, 対立構造の関係性は保たれている. いま $R > S > P > R > S > P > \dots$ という構造を考えると, X が S であったとすると, Y は R をとらなくてはならない. すると X は P をとらなくてはならない. すると Y は S をとらなくてはならない... という具合に行動に関する意思決定が無限退行する. これは「この指令に従うな」という指令を受けた主体が, この指令に従う → この指令に従わない → この指令に従う → ... という論理的な振動状態に陥る過程と, 非推移的構造が時間的に展開される点において同等である. またここでのバインディング構造は単純な 2 項対立から任意のネットワークの形をとる関係性に拡張されている. ゲームにおけるプレイヤーの論理構造についてのより明示的な分析は [7], [8] を参照.

の減衰が安定方向, ゲームの構造が中立安定な方向に対応する. したがってゲームが非推移的な場合には, 有界閉領域に閉じ込められた stretching-folding が実現され, 学習ダイナミクスがカオス的になる.

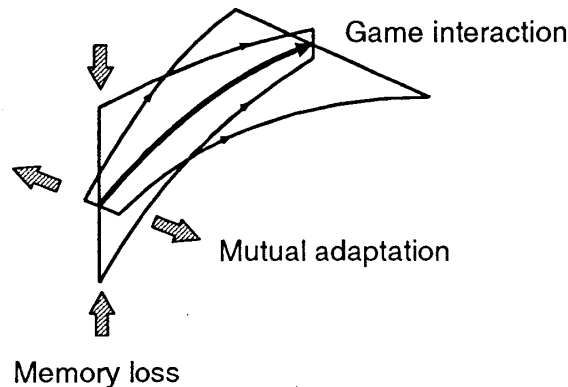


図 4: 情報空間での学習ダイナミクスの概念図: ゲームの非推移性によってダイナミクスが閉領域に閉じ込められ, 相互適応と記憶の減衰によって stretching and folding が実現される.

この場合の行動パターンの変化の仕方の不確定性 (学習のエントロピー) と (ξ, η) であらわされる行動の不確定性 (行動のエントロピー) の関係を考えるのは面白い問題である⁹.

2.3 行動獲得とゲームのダイナミクス

前節までは固定された関係性の下での学習系の相互作用を考えてきた. この場合 Fig. 1 で外部観測者は安定に同一のゲーム構造 (A, B) をみいだすことになる. しかし一般には外部観測者が切り出す線形双行列ゲーム自体が $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow G_4 \rightarrow \dots$ と時間的に構造化し, 遷移していくこともあるだろう. このような現象のメカニズムとしては学習過程におけるカオス的遍歴のような力学構造が考えられる. 各アトラクタ痕跡ごとに局所的なゲームダイナミクスが導出され, そこでの軌道平均はそのゲームの Nash 解になっている, という描像はもっともらしい. 例えば環境が変動する, 主体数が増減して系の自由度が変動する, 関係が並列多重といった場合には R が非線形になるのでこういった現象が生じると予想される.

また行動獲得において, 実際には各主体は無限に多くの行動を選択可能であり, どの行動を選択していくかというプロセスを含めて学習が進行する. 例えば初期分布が一樣であったとして, 相互作用によって行動パターンの分布が形成発展していく, (つまりある行動はよく選ばれ, 別の行動は無視されていく, というようにバイアスがかかっ

⁹学習過程にカオスが無い場合は「学習の不確定性」という概念自体が意味を持たない.

ていく。) という状況が考えられる. 上述のモデルの連続系への拡張を考えるためには行動の間になんらかのトポロジーを導入しなくてはならない. Fig. 5は主体の行動空間に一次元のリング状のトポロジーを導入した例である. このリングに沿って行動の記憶が拡散するとする. この系で $N, M \rightarrow \infty$ とすると Eqs. (13) は行動の確率密度

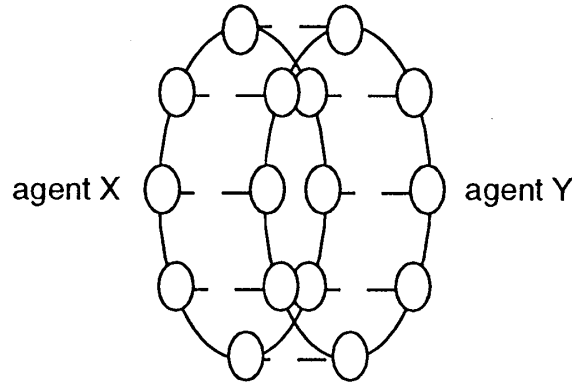


図 5: 一次元リングモデル: 2体学習の場合の例

$u(x, t), v(y, t)$ に対して以下のように拡張される

$$\begin{aligned} \frac{\partial u}{\partial t} &= \beta_X G^X(u, v) - \alpha_X \frac{\partial^2}{\partial x^2} (\log u), \\ \frac{\partial v}{\partial t} &= \beta_Y G^Y(u, v) - \alpha_Y \frac{\partial^2}{\partial y^2} (\log v), \end{aligned} \quad (14)$$

ここで G^X, G^Y が相互作用項, $\frac{\partial^2}{\partial x^2} (\log u), \frac{\partial^2}{\partial y^2} (\log v)$ が記憶項である. このモデルを用いて, 行動パターンの時空間的組織化に関する議論ができる [19]. 特に Eqs.(14) のダイナミクスの一部が Eqs. (10) に reduce できるとき, 「局所的な線形ゲームが出現した」と考えることができる. このような特殊な状況下では, 外部観測者は「 X と Y はゲーム (A, B) を play している」と解釈することになる. 一般にはこのゲーム構造は背後の学習ダイナミクスの影響を受けて複雑な時間発展を示すことになるので, 線形双行列の形での切出しは困難である.

3 Summary

理想化された多体学習系の集団的振る舞いを考えた. ここで提示されたモデルは多体相互適応をマクロに reduce した抽象モデルであり, 学習ダイナミクスの外部記述になっている. ここでの抽象は一つの例に過ぎないが, 情報理論的, ゲーム論的な分析についての方向性が明確化されている. 現段階では学習方程式の提示とその振る舞いし

か分析できていないが、このような観点から主体間の情報の流れや学習の階層性に関して力学系的視点から議論を展開できる。

参考文献

- [1] G. Bateson, "Steps to an ecology of mind : collected essays in anthropology, psychiatry, evolution, and epistemology," Northvale, N.J. : Aronson (1987).
- [2] N. Chomsky, "Cartesian Linguistics," Harper & Row, (1966).
- [3] J. Filar, K. Vrieze, "Competitive Markov Decision Processes," Springer-Verlag, New York, (1996).
- [4] D. O. Hebb, "The Organization of Behavior: Neuropsychological Theory," John Wiley & Sons, (1949).
- [5] Y. Ito, "Random Collision Models in Oriented Graphs," *J. Appl. Prob.*, **16**, p36, (1979).
- [6] K. Kaneko and J. Suzuki, "Imitation Games," *Physica*, **D 75**, p328-342, (1993).
- [7] M. Kaneko, T. Nagashima, "Game logic and its applications I," *Studia Logica*, **57**, p325-354, (1996).
- [8] 金子 守, 「ゲーム理論と蒟蒻問答」, 日本評論社, (2003).
- [9] J. Nash, "Non-cooperative games," *Annals of Mathematics*, **54**, p286, (1951).
- [10] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, (1944).
- [11] P. D. Taylor, "Evolutionary stable strategies with two types of players," *J. Appl. Probability*, **16**, p76-83, (1978).
- [12] T. Borgers and R. Sarin, "Learning through reinforcement and replicator dynamics," *J. Econ. Th.*, **77**, p1-14, (1997).
- [13] J. Hofbauer, "Evolutionary dynamics for bimatrix games: A hamiltonian system?," *J. Math. Biol.*, **34**, p675-688, (1996).
- [14] O. E. Rossler, "Chaos in Coupled Optimizers," *Ann. NY Acad. Sci.*, **504**, p229, (1987).

- [15] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal on Research and Development*, **11**, p601, (1967).
- [16] Y. Sato, E. Akiyama, and J. D. Farmer, "Chaos in a simple two-person game," *Proc. Natl. Acad. Sci. USA*, **99**, p4748, (2002).
- [17] Y. Sato and J. P. Crutchfield, "Coupled Replicator Equations for the Dynamics of Learning in Multiagent Systems," *Phys. Rev.*, **E67**, p015206(R), (2003).
- [18] Y. Sato, E. Akiyama and J. P. Crutchfield, in preparation.
- [19] Y. Sato, in preparation.
- [20] P. Schuster, K. Sigmund, J. Hofbauer and R. Wolff, "Selfregulation of behavior in animal societies. Part II: Games between two populations without self-interaction," *Biol. Cybern.*, **40**, p9-15, (1981).
- [21] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," The University of Illinois Press, (1949).
- [22] M. Taiji and T. Ikegami, "Dynamics of internal models in game players," *Physica*, **D 134**, p253, (1999).
- [23] P. D. Taylor, "Evolutionary stable strategies with two type of players," *J. Appl. Prob.*, **16**, p76, (1979).
- [24] Alan M. Turing, "Intelligent Machinery," National Physical Laboratory Report, (1948), in B. Meltzer and D. Michie (eds.), *Machine Intelligence*, Vol 5, p3-23, Edinburgh University Press, Edinburgh (1969).
- [25] Alan M. Turing, "Computing Machinery and Intelligence," *Mind*, Vol LIX, p236, (1950).