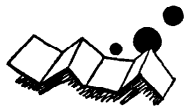


## 解説

### 5. 脳・神経系の数理モデル



#### 5.1 神経情報処理の数理†

甘利 俊一††

##### 1. はじめに——数理モデルと構成的方法

脳は10億年の生物の進化が作りあげた巨大にして緻密なシステムである。情報処理の観点から脳を見ると、1)高度の並列情報処理を実行している、2)柔軟な記憶・学習・自己組織の能力を持つ、この二点に著しい特徴がある。脳は神経細胞からなる回路網であり、生理学・解剖学は脳に関する多くの事実を明らかにしてきた。しかし、脳を総体として理解するのは容易なことではない。微視的に見るならば、神経細胞の動作を支配する細胞膜の分子過程を知ることが必要であり、これには分子科学、生化学、薬理学、遺伝学など多くの分野が関係している。また、巨視的に見るならば、脳の情報処理の基本単位である神経回路網とこれを階層的に組み上げたシステムの動作原理を理解しなければならない。神経細胞の動作から、複雑な非線形システムである神経回路網の動作原理を探ることは容易ではない。より巨視的に脳を理解しようとする、脳の入出力動作の結果である行動を論ずることになり、心理学、精神科学、認知科学が関係してくる。とくに問題解決や推論など、知的情報処理の原理を探ろうとすれば、人工知能論との対比も必要であり、トップダウン的な接近法を行わなければならない。

神経細胞の動作をもとに神経回路網の性質を調べ、さらにここから知的情報処理の原理を探るのがボトムアップ的接近である。しかし、現実の神経回路網はきわめて複雑で、その具体的な回路構成を調べることは現状ではほとんど不可能に近い。このため、現実の神経回路網そのものを調べるのではなく、ありそうな仮定をもとに神経回路網のモデルを作り、このモデルの動作を詳しく調べることにより脳の情報処理様式に迫ろうという研究方法がある。モデルの回路網で興味あ

る情報処理が実現できれば、逆に本物の脳の回路網もこのモデルと類似の構成を持つものと考えて、これを実証的に探ろうということになる。これは構成的方法またはモデルによる方法と呼ばれ、脳研究の一つの有力な方法である。

モデルの場合でも、興味ある動作を本格的に追求していくと、その構成はどんどん複雑になっていく。計算機シミュレーションによってその動作を調べるにしても、パラメータの値の設定が難しくなる。さらに、複雑なモデルのもとでは、どのような要因がどの機能にどう関係しているのかがわかりにくくなり、一つのモデルで得られた知見を他のモデルに応用することが難しくなる。この難点をさけるには、個々のモデルに含まれている本質的な機構を抽出し、可能なかぎりこれを普遍的な数学的理論として体系化したい。これができれば、逆にこの知識を利用していくつかの機構を組み合わせ、さらに現実的な肉付けをして新しい有効なモデルを構成することができる。これは個別のモデルに対する指針となりその理論の役割を果たす。こうした理論を建設するには、個々の神経回路網モデルの本質的な構造だけを抽出したできるだけ簡単なモデルを作り、その動作を数学的理論を用いて解析し、何がどういう動作に対して本質的であるかを明らかにしなければならない。このとき使われるモデルが数理モデルである。数理モデルにより、神経回路網の持つ高度の情報処理の可能性が明らかにされれば、今度はこれに種々の現実的肉付けをして現実的モデルを構成し、計算機シミュレーションでその動作を調べるとともに、現実の脳でそのような可能性が実現しているかどうかを探ることになる。

本稿では、神経回路網の数理モデルと数理的的手法について、とくに神経回路網が外界の情報に合わせて自己の構造を作りあげていく自己組織の機能に重点をおいて説明する。神経回路網の具体的なモデルについては次の福島氏の稿を参照されたい。なお、数理的方法の

† Mathematical Theory on Neural Models of Information Processing by Shun-ichi AMARI (Faculty of Engineering, University of Tokyo).

†† 東京大学工学部計数工学科

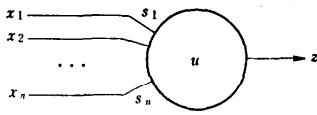


図-1 神経細胞のモデル

体系化を試みたものとしてたとえば文献1), 自己組織回路のモデルとして文献2), またこの方面の最近の動向として文献3), 4)などを挙げておこう。

2. 神経場のパターン力学

1個の神経細胞の動作は比較的単純である。これを簡単なモデルで表わしておこう。神経細胞は多数の入力信号(多いもので数万本)を受け取り、その荷重和を計算し、これを非線形に変換して出力信号とする。いま  $n$  個の入力信号の値が  $x_1, \dots, x_n$  だったとしよう。 $x_i$  は各入力線維のパルス頻度を表わすアナログ量で、時間  $t$  の関数である。細胞からの出力パルス頻度  $z$  を表わすのに、細胞の(平均)膜電位  $u$  を考えると都合がよい。 $u$  ははっておくと定数  $\tau$  で静止電位  $-h$  に近づくが、入力があると、その荷重和  $\sum s_i x_i$  に比例して増大する。すなわち、 $u(t)$  の動作は微分方程式

$$\tau \dot{u}(t) = -u(t) + \sum_{i=1}^n s_i x_i - h \quad (1)$$

で示される。 $\dot{\phantom{x}}$  は時間微分  $d/dt$  を表わし、入力  $x_i$  にかかる重み  $s_i$  を入力  $x_i$  のシナプス効率という。出力パルス頻度  $z(t)$  は  $u(t)$  の非線形関数で

$$z(t) = f\{u(t)\} \quad (2)$$

と表わせる。とくに、 $x_i$  が時間によらず一定であれば、(1)式の  $u$  はすぐに一定値に収束する。その時の出力  $z$  は

$$z = f(\sum s_i x_i - h) \quad (3)$$

である。 $f$  は一般に単調増大飽和形の関数である。とくに議論を簡単にしたときは、 $f$  を階段関数

$$f(u) = \begin{cases} 1, & u > 0 \\ 0, & u \leq 0 \end{cases}$$

で近似することがある。このときは、神経細胞は興奮 ( $u > 0, z = 1$ )、静止 ( $u \leq 0, z = 0$ ) の二状態をとる。

神経細胞は脳の中で複雑な回路網をつくる。とくに大脳では皮質の構造、すなわち2次元的な拡がりの中に神経細胞を多数配置して相互に結合した比較的均質な場の構造の回路網をなしている。そこで、ここでは神経場における神経細胞の興奮パターンを方程式をたてて、その性質を調べてみよう<sup>5)-7)</sup>。いま、神経場に

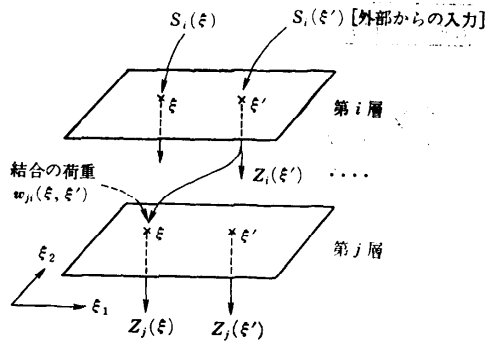


図-2 神経場

空間座標  $\xi = (\xi_1, \xi_2)$  を導入する。同じ場所  $\xi$  でも厚み方向の位置が違えば違った種類の細胞が並んでいるので、これを種類ごとに別の層に分けて考えると、皮質は多層の神経場をなす。場所  $\xi$  の付近にある第  $i$  番目の層の神経細胞について  $\xi$  の膜電位を平均したものを  $u_i(\xi, t)$  と考えよう。こうして連続化した場を考える。場の内部で神経細胞は相互に結合している。場所  $\xi'$  の第  $j$  番目の層にある神経細胞から出る出力線維が、場所  $\xi$  の第  $i$  番目の層にある神経細胞の入力端子結合するときのシナプス効率を  $w_{ij}(\xi, \xi')$  と書こう。また、場所  $\xi$  の第  $i$  層の神経細胞へ回路の外部から直接入力する外部入力すべて加算し荷重和をとったものを  $S_i(\xi)$  としよう。このとき、場の膜電位  $u_i(\xi, t)$  を支配する方程式は

$$\tau \frac{\partial u_i(\xi, t)}{\partial t} = -u_i(\xi, t) + S_i(\xi) - h + \sum_{j=1}^m \int w_{ij}(\xi, \xi') f\{u_j(\xi', t)\} d\xi' \quad (4)$$

である。

外部から刺激入力パターン  $S_i(\xi)$  が入ると、これに応じて神経場の上に興奮のパターンが生ずる。脳における思考過程は、神経場における興奮パターンのダイナミクスに基礎をおく並列情報処理であると考えられるので、神経場のパターン力学を調べておくことが重要である。もちろん場の性質は内部の結合を表わす  $w_{ij}(\xi, \xi')$  の形に強く依存している。また、思考過程では一つの場は記憶を含む他の多くの場と相互作用をしているから、事態はそう単純ではない。

簡単のため、興奮性および抑制性の二つの層からなる神経場を取り上げよう。興奮性の層(層1)は、この神経細胞が興奮するとこれに結合している他の神

神経細胞を興奮させる働きを持つ。すなわち、ここから出る出力のシナプス効率 $w_{i1}(\xi, \xi') \geq 0$ である。一方、抑制性の層(層2)では、この神経細胞の興奮は他の神経細胞の興奮を抑えるように作用するから、ここから出る出力のシナプス効率はいつも負、 $w_{i2}(\xi, \xi') \leq 0$ である。特に、興奮性の結合は狭い範囲に強い効果を及ぼすものとし、 $|\xi - \xi'|$ が小さいときに $w_{i1}(\xi, \xi')$ は大きい値を示すが、 $|\xi - \xi'|$ が大きいときは0であるとしよう。一方抑制性の結合はそう強くはないがこれよりずっと広い範囲に効果を及ぼすものとしよう。このような結合の場では、場の1カ所に興奮が起こると興奮性結合の効果でその興奮を自分のまわりに拡げてもっと強くしようとする協調的作用がある一方、もう少し離れた場所には抑制的に作用して他の場所の興奮を抑える競合的作用がある。この種の結合は側抑制または相互抑制と呼ばれ、実際の神経系にもよく見られる結合の様式である。

側抑制の場の力学過程は、ある程度数学的に解析することができる。いま、外部から第1層へ $S_i(\xi)$ の強さの刺激分布の入力が加わったとしよう。これを受けて場は興奮を始めるが、その静的な応答(定常状態の解)は次の性質を持つ<sup>7)</sup>。

1) 空間的に分布した入力 $S_i(\xi)$ で、強さのピークのある場所の神経細胞が局所的に興奮する(これを局在興奮という)。局在興奮の個数は $(S_i(\xi))$ の平均的な強さにもよるが一定個数以下に制御される。

2) 入力 $S_i(\xi)$ が引き起こした局在興奮は入力 $S_i(\xi)$ が消えた後にも、その場所に保持できる。また、入力 $S_i(\xi)$ の形が変われば、これらの局在興奮は新しい $S_i$ のピークのある方向へ移動するとともに、種々の相互作用を起こす。

こうした知見を利用すると、神経場の情報処理のモデルを作ることができる。たとえば、蛙はテクタムという神経場を持ち、ここへの入力 $S(\xi)$ は $\xi$ に対応する場所にどのくらいうまそうな(虫らしい)餌があるかを示す視覚情報の分布である。神経場は興奮の相互作用によって $S(\xi)$ の最大値のある場所を検出し、最大値のある場所だけに局在興奮を残す。蛙はこの場所に飛びつく。この局在興奮は虫の動き( $S$ の変化)に追従できる。また、最大値を一つ選ぶのではなく、最大値の候補を $k$ 個(大きい順にピークの位置を $k$ 個)選び出すこともできる。このようにして神経場に一定個数 $k$ 個以下の場所を局在興奮として記憶させることができる。これは短期記憶の原始的機構のモデルに使え

る可能性がある。

神経場の競合—協調作用を用いたもう少し程度の高いモデルに、両眼からの視差を融合して立体視を完成するモデルがある。これは、空間方向と奥行方向に対応する次元を持つ神経場で、奥行方向には強い競合的結合(抑制的結合)があり空間方向には協調的結合(興奮性結合)のあるモデルである。立体視の認識過程は場のダイナミクスによる並列情報処理として実現できる。これについても、類似のモデルがほぼ同時期に提案され<sup>9)-10)</sup>、文献10)ではその数学的解析が行われている(立体視のモデルについては次の福島氏の解説を参照)。

側抑制の神経場では、動的なパターンも発生する<sup>5)</sup>。神経場のある場所に入力があると、ここを中心に振発現象が起こり、これが進行波となってまわりへ拡散していく<sup>5), 7)</sup>。この途中で波形を修飾したり進行速度を変えたりする相互作用が起こる。神経場では最近話題になっているカオス的興奮波の発生などまだ多くの興味ある現象が起こるが、これらについてはまだあまり解析されていない。

### 3. 神経系の学習と自己組織——その数理

神経細胞は学習により自己の特性を変える能力を持っている。神経細胞の動作は(1)式または(3)式で表わされるから、入出力特性を変えるにはシナプス効率 $s_i$ を変えるのが最も簡単である。最近になって、シナプス効率 $s_i$ が変わることが生理学的にも実証されつつある。 $s_i$ の変化の法則はまだよくわかっていないので、ここでは次の仮定をしよう： $s_i$ はゆっくりと減衰するとともに、ある条件が満たされたときに入力信号 $x_i$ の強さに比例して増加する。これを数式で表わすと

$$\tau' \dot{s}_i = -s_i + c r x_i \quad (5)$$

と書ける。ここに $\tau'$ はシナプス効率の変化を表わす時定数で、神経興奮の時定数 $\tau$ よりもずっと大きい。また、 $r$ は学習を行うための条件を表わす学習信号で $r=0$ ならば学習は行われぬ。 $r$ は一般に入力信号 $x=(x_1, \dots, x_n)$ や出力信号 $z$ の関数として作られるが、神経細胞に外部から特別の教師信号 $y$ が到来するときは、 $y$ に依存している。 $c$ は定数である。

神経細胞の学習については、これまでに多くのモデルが提出された。たとえば、パーセプトロンは入力信号を2種類のカテゴリに類別するための神経回路モデルであり、教師信号 $y$ と神経細胞の出力 $z$ とが違って

いるときにこれを是正するように学習が進行する<sup>11)</sup>。これは学習信号として  $r=y-z$  を用いることに対応する。アダリン (adaptive linear neuron) と呼ぶモデルも同様である。また、連想記憶と関連して提案されている相関学習<sup>12)</sup>、教師信号  $y$  自身を学習信号として選ぶ  $r=y$  とするものである。その他、相関学習よりもパターンの分離のよい直交学習がある<sup>13)</sup>。以上は、いずれも教師信号  $y$  が別の経路から到着する場合である。一方、Hebb は特別の教師信号  $y$  は到来せず、自分の出力  $z$  を教師信号として用いる  $r=z$  という学習方式を神経系の自己組織の基本として考えている。この種のモデルも多く用いられている。ここでは教師信号  $y$  を用いる場合を学習、特別の教師信号がない場合を自己組織と呼ぶが、この区別は絶対的なものではない。いずれにせよ、ほとんどすべての学習・自己組織のモデルが(5)式で統一的に扱えるので<sup>13)</sup>、これを数学的に解析することが重要である。ところで、本当の神経細胞はどのような  $r$  を用いて学習しているのだろうか。神経細胞には多数の種類があるから、種類ごとにいろいろな  $r$  を用いて総合的な効果を出していると考えられる。

方程式(5)に従って学習が進んでいくと、神経細胞のシナプス効率  $s_i$ 、またはこれをベクトルの形にまとめた  $\mathbf{s}=(s_1, \dots, s_n)$  は、入力信号  $x_i$  (および教師信号  $y$ ) の過去の時間的な経過  $x_i(t), y(t)$  のすべてに依存して決まることになる。しかし、過去の履歴をすべておってこのような方程式をいちいち解くわけにはいかない。よく考えてみると、一つの神経細胞に入る入力信号  $x(t)$  は、その細胞にとっての環境 (これは外界の場合もあれば神経系内部の回路の場合もある) から発生する信号であって、出やすい信号と出にくい信号およびそれぞれに対応する教師信号の対からなる。そこで、神経細胞の環境情報源  $X$  を考える。これは入力信号  $\mathbf{x}$  のつくる空間に各信号の発生頻度を確率  $p(\mathbf{x})$  を確率として付加した確率的 (エルゴード的) な情報源と考える。神経細胞は、環境  $X$  の構造に適合するよう

に (教師信号のある場合にはこの指示に従って) 自己の特性を変えることになる。このとき、入力信号の時系列  $\mathbf{x}(t)$  は、長い時間をとれば  $X$  の信号の頻度分布  $p(\mathbf{x})$  に適合して各信号をこの割合で含む (エルゴード性)。この結果、確率過程である  $\mathbf{x}(t)$  を受ける学習方程式の解  $\mathbf{s}(t)$  は、右辺の第二項を  $\mathbf{x}$  の頻度分布で平均した平均学習方程式

$$t' s_i = -s_i + c \langle r x_i \rangle \tag{6}$$

で近似できることがわかる。ここに  $\langle \rangle$  は信号空間  $X$  上でのアンサンブル平均で、 $\mathbf{x}$  と  $y$  との同時確率分布を用いて

$$\langle r x_i \rangle = \int r(\mathbf{x}, y, \mathbf{s}) x_i p(\mathbf{x}, y) d\mathbf{x} dy$$

と書ける。

学習方程式は確率微分方程式で、その解はマルコフ連鎖をなす。しかし、エルゴード性により、これは環境情報源の確率構造だけで決まる平均学習方程式で十分に近似できるのである。事実、脳の学習の結果は信号を提示する細かい順序によらずに、より大きな環境の構造に応じて決まるであろう。平均学習方程式は、ある種の学習ポテンシャルの極小値へ収束する。学習ポテンシャルの性質についていろいろ調べられている<sup>11)</sup>。

話を少し具体的にしよう。パーセプトロン、アダリン、連想記憶などのモデルでは、一つの神経細胞の入出力関係を教師信号の指示に従って改善することが中心である。たとえば、簡単のため  $x_1, x_2, \dots, x_k$  の  $k$  個の信号だけをそれぞれ等確率で発生するという単純な情報源  $X$  を考えよう。入力信号が  $x_i$  であるときは  $y_i$  を出力すべきことを教師信号が指示するものとし、 $k$  個のすべてが教師のいうとおりに学習できれば、一つの神経細胞は  $k$  組の入出力関係  $(x_i, y_i) (i=1, \dots, k)$  を完全に憶えたことになる。記憶の場所は  $n$  個のシナプス効率  $\mathbf{s}=(s_1, \dots, s_n)$  であり、ここに  $k$  組の入出力関係が多重に分散して記憶される。ここで、 $r=y$  を学習信号として用いる相関学習を考えてみると、シナプス効率ベクトル  $\mathbf{s}$  は学習方程式に従って変化して、 $c_0=c/k$  として、

$$\mathbf{s} = c_0 \sum_{i=1}^k y_i \mathbf{x}_i$$

に収束することがすぐわかる。ところで、入力  $x_i$  に対する出力はこのとき  $z = f(\mathbf{s} \cdot \mathbf{x}_i - h)$  となることから(3)式からわかるから、もし

$$|\mathbf{x}_i|^2 = 1, \mathbf{x}_i \cdot \mathbf{x}_j = 0 \quad (i \neq j)$$

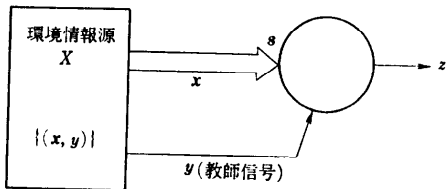


図-3 神経細胞の学習における環境情報源

が成立するように入力信号の組が規格化されていれば

$$\mathbf{s} \cdot \mathbf{x}_i = c_0 y_i \quad (i=1, \dots, k)$$

が成立する。これより  $k$  組の入出力関係が容易に実現できる。

この原理を用いた連想記憶モデルが中野のアソシエトロンをはじめ同時期に多数提案された<sup>14)~17)</sup>。とくに、文献17)では連想記憶の雑音に対する安定性、時系列の連想想起などが数学的に解析され、文献18)では連想記憶の容量が解析されている。一般の場合、入力信号が互いに直交し規格化されているとは考えにくい。このときには、信号  $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  の双対系  $\{\mathbf{x}_1^*, \dots, \mathbf{x}_k^*\}$  を

$$\mathbf{x}_i^* \cdot \mathbf{x}_i = 1, \quad \mathbf{x}_i^* \cdot \mathbf{x}_j = 0 \quad (i \neq j)$$

で定義して、

$$\mathbf{s} = c_0 \sum_{i=1}^k y_i \mathbf{x}_i^*$$

となる学習方法を見つければよい。これは  $\mathbf{r} = \mathbf{y} - \mathbf{s} \cdot \mathbf{x}$  という学習信号を用いることで実現できる<sup>18)</sup>。連想記憶は現在ではもっと精密化したモデルが提案されている<sup>19), 20)</sup>。

教師信号  $y$  のある学習は一般に学習ポテンシャルが単峰性の関数になるのに対し、教師信号のない自己組織の場合は多峰性になり、収束点が多数できて学習の結果が一意に決まらない。しかし、これはこれで並列情報処理にとってはかえって都合の良いことなのである。再び  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$  の  $k$  個の信号のみを含む環境情報源を考え、この信号を受ける神経細胞が自己の出力を学習信号とする  $\mathbf{r} = \mathbf{z}$  の Hebb 流の方式で自己組織するとしよう。学習方程式を解析すると、このままではいろいろな信号に同調してシナプス効率が上昇する一方で、興味ある結果が得られない。自分自身を先生とした自己学習はあまりうまくいかなそうに見える。しかし、入力信号  $\mathbf{x}$  の他に、抑制性の信号  $\mathbf{x}_0$  が一つ神経細胞に入力し、そのシナプス効率  $-s_0$  がまた同じ方式

$$\tau' \dot{s}_0 = -s_0 + c' z x_0 \quad (7)$$

に従って学習するとしよう。こうすると、神経細胞が興奮してそのときの入力に対するシナプス効率が増すと、同時に抑制性の効果も上がるので、ただやたらに興奮しやすくなるということにはならない。このモデルを数学的に解析すると、 $X$  に含まれる  $k$  個の信号のどれか一つに対して神経細胞が同調し、この信号が入力したときだけ興奮し他の信号に対しては興奮しないように、すなわちこの信号の表現細胞になるように自

己の特性を変えていくことがわかる。もっとよく調べると  $c' x_0^2 / c$  が分解能の役割を果たしていて、これより近いこの二つの入力信号は同一のもののみならず反応することになる。

どれか一つというのでは困ると思われるかもしれない。しかし、同じ環境情報源  $X$  から入力を受ける神経細胞が多数あれば、これらは  $X$  に属する信号を互いに分担して担当し、それぞれの細胞が各々一つずつの信号に反応しこれを表現するように機能分化するから全体としてうまく並列処理ができる。こうした特質を利用すると、自分の置かれた情報環境に適応し、そこに現われる典型的な信号を脳の内部に表現しこれを処理する機構を自分の内部に自動的に作りあげる自己組織系のモデルが作れる<sup>21), 22)</sup>。

#### 4. 神経場の自己組織とトポロジの学習

神経細胞は一つ一つが孤立して学習するわけではなく、回路網全体としての動作の中で自己組織を行う。したがって、回路網の動作と自己組織の動作とを別にして論ずるのではなく、互に関連したものとして連立して扱わなければならない。ここでは神経場の自己組織を例にとってその数理的構造を調べてみよう。話を簡単にするため、神経場  $F$  は一層で、その興奮パターン力学の方程式は(4)式を簡略にした

$$\tau \frac{\partial u(\xi, t)}{\partial t} = -u + W_0 f(u) + S(\xi) \quad (8)$$

とする。ただし  $W_0$  は積分演算子で

$$W_0 f(u) = \int w(\xi, \xi') f\{u(\xi', t)\} d\xi'$$

であり、 $-h$  の項は  $S$  に含めるものとする。また、神経場内部の結合のシナプス効率を表わす関数  $w(\xi, \xi')$  は側抑制形で、 $|\xi - \xi'|$  が小さいところで正、大きなところで負の正負両方の値をとるものとする。

神経場の各点の神経細胞は環境情報源  $X = \{\mathbf{x}\}$  から入力信号  $\mathbf{x}$  を共通に受けるものとし、場所  $\xi$  にある細胞はシナプス効率ベクトル  $\mathbf{s}(\xi)$  でこの信号を受け取るとする。また、抑制性の入力  $\mathbf{x}_0$  も考えて、場所  $\xi$  にある細胞はこれをシナプス効率  $-s_0(\xi)$  で受け取るとする。このとき、入力  $\mathbf{x}$  に対して場所  $\xi$  の受け取る刺激の総和は

$$S(\xi) = \mathbf{s}(\xi) \cdot \mathbf{x} - s_0(\xi) x_0 \quad (9)$$

である。信号の空間  $X$  は信号の頻度分布  $p(\mathbf{x})$  を確率構造として持つエルゴードの情報源であるとする。いま、場の内部の結合  $w(\xi, \xi')$  は固定されているとし

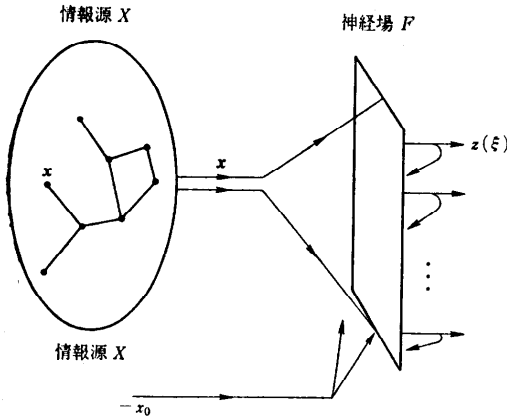


図-4 神経場の自己組織

て、 $X$  の情報を受けながら  $r=z$  という Hebb 流の方式で場の各点の神経細胞がその入力シナプス効率  $s(\xi)$  と  $s_0(\xi)$  とを自己組織により変えると、神経場の構造はどのように変わるであろうか。学習の方程式は

$$c' \frac{\partial s(\xi)}{\partial t} = -s(\xi) + c x f\{u(\xi, t)\} \quad (10)$$

$$c' \frac{\partial s_0(\xi)}{\partial t} = -s_0(\xi) + c' x_0 f\{u(\xi, t)\} \quad (11)$$

である。

入力信号  $x$  が一場に到着すると、場にはパターン力学の方程式(8)に従って興奮パターンが生じ、これがすぐに平衡状態に達する。入力  $x$  を受けているときの場所  $\xi$  の興奮状態を表わす平衡状態  $U(\xi, x)$  は

$$U(\xi, x) = W \circ f(U) + s(\xi) \cdot x - s_0(\xi) x_0 \quad (12)$$

を解いて求める。これは結合のシナプス効率  $s, s_0$  の全体に依存している。  $U(\xi, x)$  は入力  $x$  に対する神経場の応答特性を表わしている。場の応答特性をみるには、場所  $\xi$  を興奮させるのはどんな入力信号が来たときかを考えればよい。これは

$$R(\xi) = \{x | U(\xi, x) > 0\}$$

で示される  $X$  の部分集合である。すなわち場の1点  $\xi$  は  $R(\xi)$  に属する信号  $x$  が来たときに興奮する。これを  $\xi$  の受容野と呼ぶ。では、一つの信号  $x$  はどんな場所を興奮させるだろうか。これは神経場  $F$  の領域

$$E(x) = \{\xi | U(\xi, x) > 0\}$$

で表わせる。すなわち、 $E(x)$  は入力  $x$  が来たときに興奮する場所である。これを  $x$  の投射野と呼ぶ。信号空間  $X$  と場  $F$  との直積集合  $F \times X$  の中に  $U(\xi, x) > 0$  という領域を考えれば、これが信号  $x$  とそれを受容して興奮する場所  $\xi$  との関係を表わ

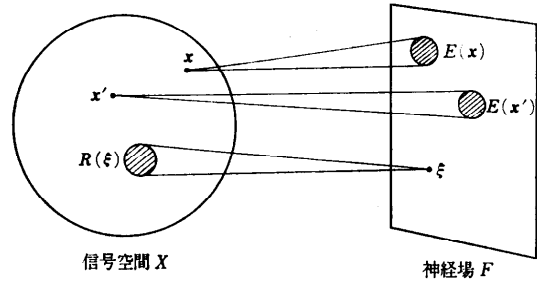


図-5 神経場の受容野と投射野

している。こうして、 $U(\xi, x)$  は信号  $x$  が場のどこで受容され表現されるかを示す。

信号空間  $X$  は信号  $x$  をすべて一様の確率で含むのではなく、ある構造を持って信号を発生する。たとえば、 $X$  が言語や文字などを表現する情報源ならば、何種類かの基準信号をうまく相互に区別できるように分布している。もちろん基準信号がいつもそのまま正確に入力するわけではないから、 $p(x)$  はこのまわりに分布している。基準信号は互いに似た部分と違った部分があって、複雑な親近関係で結ばれている。互いに近いものを線で結べば、多次元の空間  $X$  の中で複雑な関係を持つグラフになる。  $X$  の信号  $x$  は、自己組織によって場  $F$  の場所  $E(x)$  に写像される。  $k$  個の基準信号  $x_1, \dots, x_k$  に対して、これを表現する細胞  $E(x_1), \dots, E(x_k)$  は神経場のどこに配置されるのか、これは興味ある問題である。  $x_1, \dots, x_k$  の相互の関係の表わす多次元のトポロジは2次元の神経場  $F$  の中でどう表現されているのであろうか。もっと簡単な例は、入力信号  $x$  が別の神経場の局所的な興奮からくる場合である。このとき、 $X$  は入力信号の神経場の構造にしばられた2次元なトポロジを持つ。この場合は二つの神経場がともに2次元であるから都合がよい。一般に神経場  $F$  は自己組織によって信号空間  $X$  の複雑なトポロジを自己の内部に表現しなければならない。場はシナプス効率  $s(\xi), s_0(\xi)$  を変えることにより  $X$  と  $F$  との対応関係  $U(\xi, x)$  を変えて  $X$  の表現を内部に作っていく。

数学的には、対応関係  $U(\xi, x)$  が自己組織によってどう変化していくか、その方程式を導く方が直接的でわかりやすい。くわしい解析は省くが、環境情報源  $X$  のもとでの自己組織によって  $U(\xi, x)$  がどう変化していくか、その変化を支配する偏微積分方程式を導くことができる<sup>23)-25)</sup>。この方程式の定常状態の解を  $\bar{U}(\xi, x)$  とすると、これは

$$\bar{U}(\xi, \mathbf{x}) = W \circ f(\bar{U}) + K * f(\bar{U}) \quad (13)$$

を満たす。ただし、 $K$ は $X$ の確率構造に関係した演算子で

$$K * f(U) = \int (c \mathbf{x} \cdot \mathbf{x}' - c_0 x_0^2) \times f\{U(\xi, \mathbf{x}')\} p(\mathbf{x}') d\mathbf{x}'$$

である。

話を少し具体的にしよう。二つの神経場  $F_1$  と  $F_2$  とを結合する場合に、 $F_1$  の神経細胞の出力線維は  $F_1$  でのトポロジ (近さの関係) を保ちつつ  $F_2$  の神経細胞と結合し、 $F_1$  と  $F_2$  の間の連続写像ができあがる場合が多い。これは一見当り前のことのように見えるが、たとえば目の網膜と大脳後頭部の視覚領のように二つの神経場が離れている場合に、 $F_1$  の神経細胞から出る神経線維がどのようにして  $F_2$  での自分の行くべき位置を知るのかという問題が生ずる。 $F_1$  から出る神経線維は束になっているが、これを一度ばらばらにまぜて  $F_1$  でのトポロジを破壊しても、各神経はついに正しい行先を見出す。 $F_1$  や  $F_2$  の一部分を破壊しても、残りの部分どうしがつめ合って全体としてきれいな対応の結合ができるのである。この場合、 $F_1$  では互いに近い距離にある二つの神経細胞は同時に興奮する確率が高く、はなれた細胞は独立に興奮する傾向にある。したがって、 $F_2$  の環境入力情報源  $X$  は  $F_1$  に起こる興奮パターンの全体であると考えた場合、 $X$  は興奮確率の相関に起因する2次元的なトポロジを持ち、これに導かれて  $F_1$  から  $F_2$  への連続写像ができるのである。このような連続写像の形成がコンピュータシミュレーションで実際に確かめうる<sup>24)</sup>。また数学的な解析が行われている<sup>24)</sup>、<sup>25)</sup>。

面白いことに、ある状況のもとでは、二つの神経場の連続的な対応を示す解が不安定になる<sup>26)</sup>。このときは、場はブロック構造に分解され、ブロックどうしが1対1に対応することになる。実際の神経場でも、大脳視覚領は超コラム、コラムというブロック構造に分割されていて、各ブロック内で情報のトポロジカルな表現や配置が行われている (外山氏の稿参照)。多次元のトポロジを2次元の神経場で表現するために、神経系は一度場を離散化してブロックごとのトポロジ構造を定めるとともに、ブロック内での微視小世界で再び情報のトポロジ関連をつける方策を見出したと考えることができる。数学的解析はこのような場合にも有効である。その他、Kohonen は、自己組織によってフィンランド語の音韻が2次元神経場にどのように配

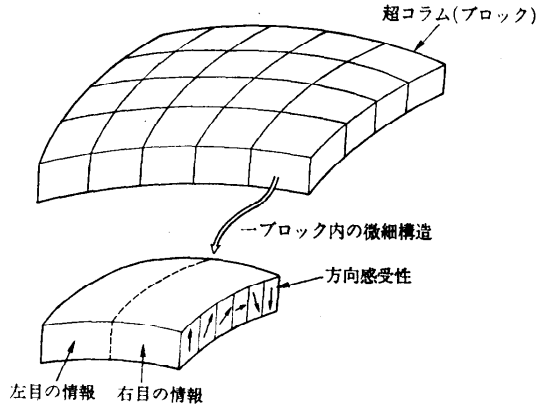


図-6 大脳視覚領。超コラムは網膜上の位置に

置されるか、その様子をシミュレーションで調べている<sup>27)</sup>。

神経場の興奮力学と自己組織の解析はまだ手がつきはじめたばかりであり、これからの興味ある課題である。

## 5. おわりに

神経情報処理の数理的構造、とくに自己組織の原理を神経回路網の数理モデルによるボトムアップ的手法で調べてきた。この結果、複雑なトポロジを持つ外界の情報の脳内表現の問題に行きつく。脳内情報の表現、変換、再生は、記憶や思考と関連してきわめて重要かつ興味ある問題である。今後はこの成果を生かして、認知科学や人工知能の成果も取り入れてトップダウンの方法も併用して研究を進めていく必要があると思う。この方向についてはここではふれられなかった。

## 参考文献

- 1) 甘利俊一：神経回路網の数理，産業図書，東京 (1978)。
- 2) 福島邦彦：神経回路と自己組織化，共立出版，東京 (1979)。
- 3) Amari, S. and Arbib, M.A.: Competition and Cooperation in Neural Nets, Springer Lecture Notes in Biomathematics, Vol. 45, Springer (1982)。
- 4) IEEE Transactions on Systems, Man and Cybernetics, Special Issue on Neural and Sensory Information Processing, Vol. SMC 13, No. 5 (1983)。
- 5) Wilson, H.R. and Cowan, J.D.: A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Tissue, Kybernetik,

- Vol. 13, pp. 55-80 (1973).
- 6) Ellias, S. A. and Grossberg, S.: Pattern Formation, Contrast Control, and Oscillations in the Short Term Memory of Shunting On-Center Off-Surround Networks, *Biol. Cybern.*, Vol. 20, pp. 69-98 (1975).
  - 7) Amari, S.: Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields, *Biol. Cybern.*, Vol. 27, pp. 77-87 (1977).
  - 8) Marr, D. and Poggio, T.: Cooperative Computation of Stereo Disparity, *Science*, Vol. 194, pp. 283-287 (1976).
  - 9) Sugie, N. and Sawa, M.: A Scheme for Binocular Depth Perception Suggested by Neurophysiological Evidence, *Biol. Cybern.*, Vol. 26, pp. 1-15 (1977).
  - 10) Amari, S. and Arbib, M. A.: Competition and Cooperation in Neural Nets, in *Systems Neuroscience* (ed. Metzler, J.), pp. 119-165, Academic Press, New York (1977).
  - 11) Minsky, M. and Papert, S.: *Perceptron—An Essay in Computational Geometry*, MIT Press (1971).
  - 12) Kohonen, T.: *Associative Memory*, Springer (1977).
  - 13) Amari, S.: Neural Theory of Association and Concept Formation, *Biol. Cybern.*, Vol. 26, pp. 175-185 (1977).
  - 14) Nakano, K.: Associatron—A Model of Associative Memory, *IEEE Trans.*, Vol. SMC 2, pp. 381-388 (1972).
  - 15) Anderson, J. A.: A Simple Neural Network Generating Interactive Memory, *Math. Biosci.*, Vol. 14, pp. 197-220 (1972).
  - 16) Kohonen, T.: Correlation Matrix Memory, *IEEE Trans.*, Vol. C 21, pp. 353-359 (1972).
  - 17) Amari, S.: Learning Patterns and Pattern Sequences by Self-Organizing Nets of Threshold Elements, *IEEE Trans.*, Vol. C 21, pp. 1197-1206 (1972).
  - 18) 上坂吉則, 尾関和彦: 連想形記憶の二, 三の性質, *信学論*, Vol. 55 D, pp. 323-330 (1972).
  - 19) Hirai, Y.: A Model of Human Associative Processor (HASP), *IEEE Trans.*, Vol. SMC 13, pp. 851-857 (1983).
  - 20) 福島邦彦: 階層構造をもつ連想記憶の神経回路モデル, *信学論*, Vol. 66 A, pp. 1130-1137 (1983).
  - 21) Fukushima, K.: Cognitron: A Self-Organizing Multilayered Neural Network, *Biol. Cybern.*, Vol. 20, pp. 121-136 (1975).
  - 22) Fukushima, K. and Miyake, S.: Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position, *Pattern Recogn.*, Vol. 15, pp. 455-469 (1982).
  - 23) Amari, S.: Topographic Organization of Nerve Fields, *Bull. Math. Biol.*, Vol. 42, pp. 339-364 (1980).
  - 24) Amari, S.: Field Theory of Self-Organizing Neural Nets, *IEEE Trans.*, Vol. SMC 13, pp. 741-748 (1983).
  - 25) Takeuchi, A. and Amari, S.: Formation of Topographic Maps and Columnar Microstructures, *Biol. Cybern.*, Vol. 35, pp. 63-72 (1979).
  - 26) Willshaw, D. J. and von der Malsburg, C.: How Patterned Neural Connections Can Be Set up by Self-Organization, *Proc. Roy. Soc.*, Vol. B 194, pp. 431-445 (1976).
  - 27) Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps, *Biol. Cybern.*, Vol. 43, pp. 59-69 (1982).

(昭和 59 年 8 月 20 日受付)