

# 脳の科学 第7回

## ニューラルネットワーク

担当：浅川伸一

2009年11月27日

人間の感情と、他の生物のそれと、近代的な型の自動機械の反応との間に鋭い乗り越えられない区画線を引く心理学者は、私が私自身の主張に慎重でなければならないのと同様に、私の説を否定するのに慎重でなければならない

ウィーナー著「人間機械論」みすず書房, p.73

### 1 脳の構成論的研究

近年、脳はさまざまな方法で研究されている。fMRI に代表される機能的脳画像研究や、ネコやサルにマイクロ電極を差し込んで細胞の動作を測定する電気生理学的手法や、動物を用いた脳の破壊実験、脳波、薬理学的方法、神経心理学と呼ばれる障害を持った脳の観測、心理学実験などである。とりわけ、機能的脳画像研究の進歩が著しい。時間解像度、空間解像度とも精度が飛躍的に向上している機能的脳画像研究は、従来の心理学的研究手法を根本から変革する力すら持っているように思われる。

これらの方法に加えて脳の構成論的研究、すなわち、ニューラルネットワークと呼ばれる脳のモデルを作って、このモデルが実際の脳と同じ機能を果たしていると考えられるのならば、モデルの持っている機構が脳にも存在する可能性があるとする研究分野がある。情報論的必然性という言葉があって、入出力関係が複雑で巧妙であるほど、それを実現する情報処理の筋道はそういくつもあるはずはない、というものである [33]。入出力関係が脳とモデルとの間で一致したときメカニズムも一致している可能性は、この情報論的必然性に導かれてかなり高いのかも知れない。

脳の構成論的研究とは、このようなモデルを作ってコンピュータによるシミュレーションを通して脳の機能を類推してゆく研究である。脳の構成論的研究で重要なことは、生理学的に分かっていることはできるだけモデルに取

り入れ、分かっていないことについては大胆に仮定してモデル化を行なうということである。

## 2 ニューラルネットワークの定義

ニューラルネットワークモデルとは、脳の振る舞いを模倣するための表現のことである。必ずしもプログラムによって表現されている必要は無いが、数式を使って表現される場合が多い。神経回路網 (neural network) モデル、PDP(parallel distributed processing) モデル、あるいはコネクショニスト (connectionist) モデルと呼ばれることもある。論文などのタイトルに上記のいずれかの言葉を入れるとき、心理学者は PDP モデルやコネクショニストモデルを使う傾向がある。

ここでは、ニューラルネットワークモデルを「生体の中枢系で行なわれている情報処理の機能、性能、および特性を記述するための抽象化された表現」と定義しておく。また、ニューラルネットワークとは「脳の基本素子であるニューロン (神経細胞) とニューロンが結合したネットワーク (神経回路網) の構造や情報処理のメカニズムにヒントを得て脳の持つ情報処理能力のモデル化を目指す研究分野のこと」と言っても良いだろう。

ニューロンの動作を抽象化した場合、ニューロンと呼ばずにユニット、素子、あるいはプロセッシングエレメントなどと呼ぶことがある。この場合は一つ一つのニューロンを抽象的に表現したと言うよりニューロン集団の振る舞いを記述していると考えられることもできる。認知心理学でニューラルネットワークモデルという場合には脳の部位を一つのユニットとして扱うことが多い。

## 3 並列処理と直列処理

言語、意識、思考などと言った高次認知機能は直列的である。一方、視覚や聴覚といった知覚情報処理は並列計算が前提となっている。脳内では多数のニューロンが同時的、並列的に活動している。このことから意識的な活動は無意識的な並列処理が組み合わされた結果として生じるのではないかと言う仮説が成り立つだろう。

## 4 Marr の 3 つのレベルとの関連

Marr[22] は 3 つのレベルを区別した。

1. 計算論のレベル
2. 情報表現とアルゴリズムのレベル

### 3. ハードウェアのレベル

ニューラルネットワークモデルとは 1 と 2 とを含む研究だと言えるだろう。

どの説明水準においても説明には階層的なレベルがあることを認識すべきである。例えば、コーヒーを飲むと眠れなくなることがある。これはコーヒーの中にカフェインと呼ばれる塩基性有機化合物が含まれているためである、という説明は一つの説明レベルである。ところがカフェインのどのような作用によって興奮して眠れなくなるとかを説明するレベルでは、上述の説明では説明になっていない。

この授業では説明のレベルをニューロン以上に限定してきた。

## 5 何ができるか

今まで提案されてきたモデルの一部を列挙すると次のようになる。

- 単語の読み acquired dyslexia [26, 27, 30, 29]
- 書字 spelling [4]
- 発話 speech production [15, 10]
- 相貌認知 face recognition [5, 12, 13]
- 視覚的物体呼称 visual object naming [14, 30]
- 空間的注意 spatial attention [8, 18, 13]
- 学習と記憶 learning and memory [1, 2, 6, 24, 23]
- 意味記憶 semantic memory [11, 17, 13]
- 行動制御 control of action and responding [3, 9, 7]

## 6 簡単な歴史

1943 年 McCulloch and Pitts の形式ニューロン

1949 年 Hebb の学習則

1962 年 Rosenblatt パーセプトロン

1969 年 Minsky & Papert パーセプトロン, 中野 アソシアトロン

1972 年 Kohonen 連想記憶モデル, Anderson 連想記憶モデル

1970 年代 甘利らによる数理解析, Marr と Albus による小脳パーセプトロン説

1975年 福島 コグニトロン

1980年 福島 ネオコグニトロン

1982年 Hopfield

1983年 Farmann & Hinton ボルツマンマシン

1985年 Hopfield & Tank 巡回セールスマン問題

1986年 Rumelhart & Hinton バックプロパゲーション, Sejnowski & Rosenberg NETTalk

マッカロックとピッツ McCulloch and Pitts の形式ニューロン formal neuron は第二次世界対戦中であり、ヘップ Hebb の学習則は GHQ による日本占領中 (1952 年まで) だから意外に古いことが分かる。ローゼンブラッド Rosenblatt によるパーセプトロン perceptron の提案から、第一次ニューロブームと呼ばれる時代に入る。このブームはミンスキーとパパート Minsky and Papert による批判で一気に下火になる。第二次ニューロブームと呼ばれるのはホップフィールドネットとラメルハートらによるバックプロパゲーション法の再発見によって花開いた形である。この他にも要な研究が抜けているが、歴史の概略を説明する目的なので省略してある。重要なのは第一次ニューロブームと第二次ニューロブームの間に日本人研究者による優れた研究があったことであろう。

## 7 ニューラルネットワークの特徴

人間の行なうさまざまな行為はすべてニューロンの活動とニューロン間の結合の強度として表現可能であると考えるのがニューラルネットワークである。この意味において、ニューラルネットワークは「強い AI」[32] を主張している。ニューラルネットワークにおける特徴を挙げるとすれば、分散表現と統計的学習の漸進的学習、および相互作用の3点に要約できる [13, 28]。

分散表現: ニューラルネットワークにおいては、知識はそれぞれのユニット集団の活性化パターンとして表現される。例えば、ある単語の意味は別の単語の意味とは異なる活性化パターンとして表現されており、類似した概念は互いに類似した活性化パターンとして表現される。(ポピュレーションコーディングのこと)

統計的構造の漸進的学習: 長期的な知識はユニット間の関係、すなわちユニット間の結合強度としてネットワーク内に埋め込まれている。ユニット間の結合強度は学習によって徐々に変化する。すなわち、学習により

外界から提供される情報 (単語間の類似度や相互関係など) の統計的性質が徐々に獲得される。

学習にはシステムの望ましい出力を得るために外部の教師信号を利用するパーセプトロン型の教師あり学習と、明確な外部信号を仮定せず、ユニット間の結合を外界の環境に合わせて変化させるヘップ型の学習則に従う教師なし学習との 2 種類がある。非常に多くのニューラルネットワークモデルが今までに提案されてきているが極論すればこの 2 種類の学習則しか無いといってよい。

相互作用: ユニットの密接に連結されており、相互に影響しあう。すなわちユニット間の結合強度に応じて、互いに活性化パターンを強め合ったり、弱め合ったり、振動したりというような複雑な相互作用をする。

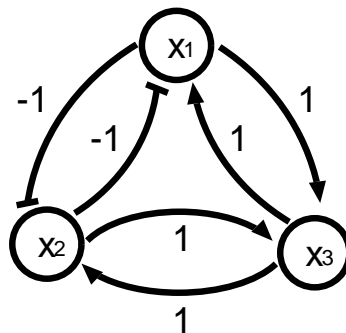


図 1: 3 つのユニットからなるネットワーク 8 類似の状態が定義できる

たとえば図 1 は 3 つのユニットからなる簡単なニューラルネットワークの例である。各ユニットが 1 か 0 かの 2 状態を取るとすれば、このネットワークには  $2^3 = 8$  種類の状態が定義できる。図中に結合強度が +1 か -1 かで書かれているが、次の時刻におけるユニットの活性値は直前の他の二つのユニットの状態によって変化することになる。

## 8 生理学的事実の抽象化

ここでは、ものごとを極端に単純化して捉えることから始める。この単純化に不満を覚える人もあるだろう。しかし、真実は単純さと複雑さの狭間にある。現在あらゆる理論は現実の人間のモデルとしては単純すぎる。逆に言えば人間の行為・活動は非常に複雑である。この複雑な現象は、そのままではとらえることができないため、ある程度の単純化が必要になる。この単純化のことをモデルと言う。モデルが妥当性を持つためには現実に則して、扱っている現象を説明できる必要がある。しかし、脳の科学では現象が複雑

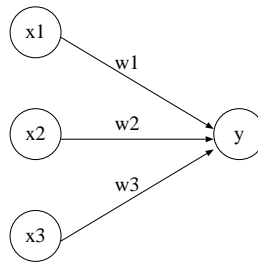


図 2: ニューロン  $y$  は  $x_1, x_2, x_3$  からの入力を受けて、入力ニューロンからの活性値を結合係数  $w$  によって重み付けられた値の関数として自身の出力値が定まる

すぎるため、現在どのようなモデルを用いた説明でも単純すぎるのである。それならば、モデルを作る場合には本質を損なわない限り極力簡単なモデルを用いる方がよい。定性的に同じ動作をする 2 つのモデルがあったとすれば、より簡単なモデルを選ぶべきである。

これをオッカムのカミソリ Occam's razor という。「同じような理論で、難しいのと簡単なのがあった場合、難しい方が有効である証拠が見つかるまで、簡単なほうを使うべきである。」あるいは、「不必要に物事を増やしてはならない。」などといわれる。

## 8.1 表記

ニューロンはシナプスを介して多数のニューロンから信号を受け、興奮するかもしれないかが決まる。この意味で多入力 1 出力の情報処理装置と見なすことができる。 $n$  個のニューロンから信号のを受け取っているニューロンで、各入力信号の強さ  $x_1, x_2, \dots, x_n$  をとする。このニューロンの膜電位の変化量を  $u$ 、出力信号を  $z$  とする。 $i$  番目の軸索に任意の単位の強さの信号が与えられたときに、この影響を受けて変化する膜電位の量を  $w_i$  と表す。 $w_i$  はシナプスの伝達効率を表す量でシナプス荷重 (または 結合荷重、結合強度) という。抑制性のシナプス結合については  $w_i < 0$ 、興奮性の結合については  $w_i > 0$  となる。

## 9 基本回路の設計

本節では、基本的な論理回路と簡単な記憶回路を神経回路網で構成する方法を考える。シリコンウェハ上に構成される論理回路をニューロンユニットでも実現できることを示す。

## 9.1 AND 回路

2入力1出力の回路において、2つの入力が共に真であるときのみ真を出力し、そうでなければ偽となる論理演算である論理積 (AND) を考える。論理積は引数を2つとる演算であり、出力を  $y$  とすれば  $y = f(x_1, x_2)$  のように書くことができる。 $x_1, x_2$  とともに1または0の値をとるものとすれば、 $y$  が1であるためには  $x_1 = 1$  かつ  $x_2 = 1$  でなければならぬ (図3)。

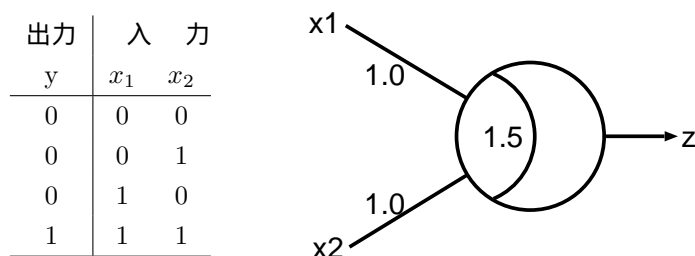


図3: 論理積 AND の真偽表 (左) と等価な演算を実行可能な形式ニューロン (右)。右図では  $x_1, x_2$  の2つの入力が共に1でなければ閾値 1.5 を越えることがない。従ってこのニューロンは2つの入力が無ければ発火できない。

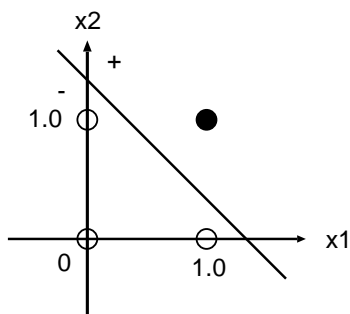


図4: 図3と等価な幾何学的表現  $y \geq x_1 + x_2 - 1.5$ 。2次元平面を2領域に分割し、一方を発火する領域 (+)、他方をしない領域 (-) に分割することに等しい。このような分割の仕方は無限に存在することに注意。

この素子がマッカロック・ピッツの形式ニューロンで表せるとすれば、2つの入力の総和が閾値を越えれば発火し、そうでなければ発火しない素子を考えていることと等価である。

## 9.2 NOT 回路

同様に図5は否定 (NOT) 演算である。入力があれば ( $x = 1$  のとき) かならず負の結合荷重 (-1.0) によって閾値 (0.5) を下回るので、このニューロン

は発火できなくなる。

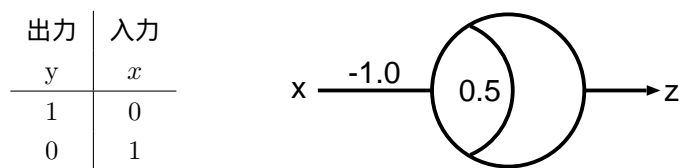


図 5: 否定 NOT 演算と等価なニューロン

### 9.3 簡単な記憶回路 — フリップフラップ回路 —

上記の AND 素子と NOT 素子とを繋いで簡単な記憶回路を作ることができる。図 6 で各素子は 1 か 0 かを値として取りうる形式ニューロンだとす

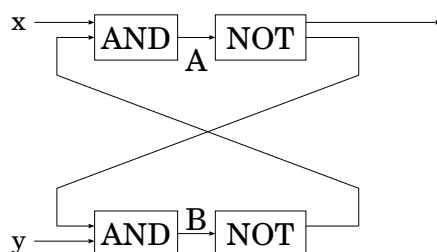


図 6: フリップフラップ回路の一例

る。今、入力  $x$  と入力  $y$  とが共に 1 であれば  $A = 1, B = 0$  あるいは  $A = 0, B = 1$  のときだけこの回路は安定である。ここで  $x = 0, y = 1$  とすると  $A = 0, B = 1$  の状態になり、 $x = 1, y = 0$  とすると  $A = 1, B = 0$  の状態になる。しかも、この状態は  $x = y = 1$  に入力を戻しても保存されている。これは 1 ビットの記憶回路でありフリップフラップ回路 (flip-flop circuit) と呼ばれている。このことは AND と NOT を実現できる神経回路素子があれば記憶回路を作ることができることを示している。しかも工学的に実現されている回路と完全に等価である。フリップフロップ回路を 16 個または 32 個まとめてレジスタ (register) と呼ぶ。市販されている PC の CPU の性能を指して 16 ビットマシンとか 32 ビットマシンとか呼ぶのは、このレジスタの大きさ (記憶装置への基本的な入出力単位の基本でもある) による。一般にコンピュータの速度はこのフリップフラップ回路が安定するまでの時間に大きく依存する。なぜなら、コンピュータの基本動作は原理的に、上述のフリップフラップ回路が安定するのを待って、次の命令をレジスタに読み込むことの繰り返しだからである。



## 10 学習について

ニューラルネットワークモデルでは、シナプスの結合係数を変更することをニューラルネットワークの学習と呼ぶ。学習時における結合係数の変化を記述したルールを学習則 (learning rule) という。ここではヘップ Hebb の学習則とバックプロパゲーション法を紹介する。

### 10.1 ヘップ則

Hebb の原文では以下のような記述である。

When an axon of cell A is near enough to excite cell B and repeatedly or consistently takes part in firing it, some growth process or metabolic changes takes place in one or both cells such that A's efficiency, as one of the cell B, is increased.

—Hebb,D.O.,(1949), The Organization of Behavior, p.62 —

「同時に発火したニューロン間のシナプス結合は強められる」ことを主張しているのがヘップ則 (Hebbian rule) である [16]。ヘップの学習則とも表記されるヘップ則は以下のように定式化できる。ニューロンの発火状態を 1, 休止状態を 0 と表現することにし, ニューロン  $y$  から ニューロン  $x$  へのシナプス結合係数を  $w$  とする。このときヘップの学習則は, シナプス結合係数の変化  $\Delta w$  として表現され、 $\Delta w = \lambda xy$  と書くことができる。ここで  $\lambda(\geq 0)$

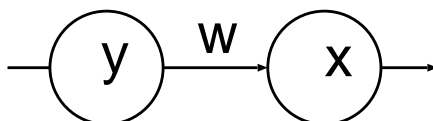


図 7: ヘップの学習則

を学習定数という。  $x$  と  $y$  は 1 と 0 の 2 とおりの状態にしかならないため, 可能な組み合わせは 4 通りになる。このうち  $\Delta w \neq 0$  となる組み合わせは、

		$x_j$	
		1	0
$y_i$	1	$\lambda$	0
	0	0	0

表 1: ヘップの学習則

( $x = 1$ ) かつ ( $y = 1$ ) の場合だけである。

すわなち結合係数は、入力ニューロンと出力ニューロンの同時発火頻度に比例して、入力ニューロン  $x$  の  $\lambda$  倍だけ増大することを意味する。ある時刻  $t$  で、入力  $x$  が  $y$  を発火させたとき、次回  $t + 1$  に同じ入力を与えられるとより強く発火させる効果を持つようになることもできる。逆に言えば同じ刺激を入力し続けると  $|w|$  が限りなく大きな値になってしまうことも意味している。このようにヘップ則には、学習回数を重ねても一定の値に収束しないという特徴を持っている。

ヘップ則の変形はさまざまに考えられていて、例えば同時に発火しなかった時にシナプス結合を弱めるアンチヘップ則、減衰項を加える方法、入力と出力の差 (の 2 乗) を利用する方法、などが考案されている。海馬の興奮性シナプス経路に高頻度 (100Hz 程度) の電気刺激 (これをテタヌス刺激 tetanus stimulus という) を 1 秒程度加えると長い間シナプスの伝達効率が変化する。この現象は長期増強 long term potentiation (LTP) と呼ばれる。LTP は Hebb の学習則の一例であると考えられている。

## 11 シナプス可塑性

シナプスの伝達効率の変化の強さと、強められるか弱められるかという変化の方向が、個々の入力スパイクとスパイク発火の時間的關係で決まる。

海馬や皮質の錐体細胞間のシナプスは、シナプス前ニューロンの入力スパイクがシナプス後ニューロンの発火に先立つ場合には増強され、この時間關係が逆転すると弱められるスパイク時間依存のシナプス可塑性 STDP。

強化されたものはさらに強まり、減弱されたものはさらに弱まるというポジティブフィードバックが働くため、大きな伝達効率を得る集団と、小さな伝達効率しか持てない集団に分かれていく。

### 11.1 パーセプトロン

パンデモニウムモデルを素直に実現したニューラルネットワークモデルとしてパーセプトロンが挙げられる。パーセプトロンは工学の分野でパターン認識との関連で長い研究の歴史がある。パーセプトロン perceptron とはローゼンブラット [31] によって提案された図 8 のような 3 層の階層型ネットワークモデルである。パーセプトロンはマッカロック・ピッツの形式ニューロンを用いて学習則にヘップ則を使ったモデルで、単純な認識能力を獲得することができる。

パーセプトロンは 3 層の階層型ネットワークでそれぞれ、 $S$ (sensory layer),  $A$ (associative layer),  $R$ (response layer) と呼ばれる層からなっている。 $S \rightarrow A \rightarrow R$  のうちパーセプトロンの本質的な部分は  $A \rightarrow R$  の間の学習にある。

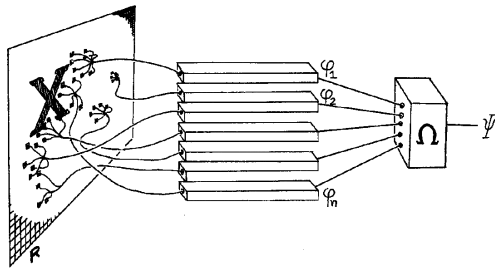


図 8: パーセプトロン。ミンスキー&パパート著, 中野, 阪口訳, 1993, パーソナルメディア

最下層の入力層は外界からの刺激入力表現している。中間層では入力情報の変換が行われ、最上位層である出力層で認識に到る。

パーセプトロンとは、入力パターンに現われる  $P^+$  と  $P^-$  という 2 つのパターンを識別する問題を考えたとき、 $P^+$  が入力されたとき 1 を、 $P^-$  のとき 0 を出力する機械である。あるいは、出力ユニットが 1 つしかない場合を考えれば、パーセプトロンは入力層に現われるパターンを 1 と 0 とに 2 分する機械であるということができる。

## 11.2 バックプロパゲーション

### 11.2.1 XOR 問題、線形分離不可能な問題

パーセプトロンでは絶対に解けない問題に排他的論理和 (XOR) 問題がある。排他的論理和とは、2 つの入力のうちいずれか一方のみが 1 のとき 1 を出力する問題である。図 9 左を見るとわかるとおり、XOR 問題は一本の判別直線で白マルと黒マルを分離できない、すなわち線形分離不可能な問題である。

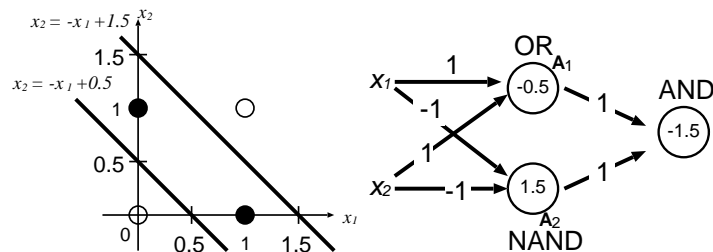


図 9: XOR 問題の幾何学的表現と XOR 問題を解くためのネットワーク

図 9 右は図 9 左の幾何学表現を対応するネットワークとして表現したもの

である。一番左が入力層、一番右が出力層、中間の 2 つが中間層である。ユニットの中に書かれた数値は各ユニットのしきい値を示している。中間層の 2 つのユニットのうち上の OR と書かれたユニットは、 $x_1 + x_2 - 0.5 > 0$  のとき発火する。この式を書き換えると、 $x_2 > -x_1 + 0.5$  となるので図 9 左の下斜線より上の領域に対応する。一方、中間層の下の NAND(not and) と書かれたユニットは、 $-x_1 - x_2 + 1.5 > 0$  のとき発火するので、移項して  $x_2 < -x_1 - 1.5$  とすれば、図 9 左の上斜線より下の領域に対応していることが分かる。さらに、AND と書かれた出力ユニットは、2 つの中間層ユニットからの情報の論理積 (AND) を計算するユニットになっている。そこで、2 つの中間層ユニットの両方が発火する場合だけ、出力ユニットも発火し、1 を出力する。これは、図 9 左では、「下の斜線より上の領域」でかつ「上の斜線より下の領域」に対応する。すなわち、図中の黒丸の領域だけが分離されることになる。このような 2 本の直線は図中にいくらかでも引けることから、XOR 問題の解も無数に存在することが分かる。

入力層		中間層		出力
$x_1$	$x_2$	$a_1$	$a_2$	r
0	0	0	1	0
0	1	1	1	1
1	0	1	1	1
1	1	1	0	0

表 2: 図 9 に対応する XOR 問題の真偽表

図 9 左にあるとおり、中間層のユニット 1 個は 1 つの線形判別関数に相当すると考えられる。中間層から出力層への結合では各出力の論理積 AND を計算していることに相当する。 $n$  個の中間層を用意すれば原理的には  $\frac{n^2+n+2}{2}$  個のカテゴリ分類が可能である。パーセプトロンが XOR 問題を解くことができない理由は入力層から中間層にいたる結合係数を変更する手段がないことなのである。

### 11.2.2 誤差逆伝播法 (一般化デルタルール)

XOR 問題でも見たように、パーセプトロンの問題点は学習が出力層と中間層の間だけで行われ、入力層と中間層の結合係数を更新することができない。換言すれば中間層ユニットの誤差を明示的に知る方法はない。このことは信用割り当て credit assignment 問題と呼ばれる。この信用割り当て問題を解決したのがバックプロパゲーション (誤差逆伝播法あるいは一般化デルタルール) である。

## 12 Linsker[21, 20, 19] の皮質マップの自己組織化

網膜と第一次視覚野 V1 の間には連続的な 1 対 1 対応が存在する retinotopy。鼓膜の周波数選択特性と第一次聴覚野 A1 との間にも対応関係が見られる tonotopy。同様に体表面の感覚と体制感覚野の間にも対応関係が見られる somatopy。すなわち感覚器官と第一次感覚野との間の神経結合は、類似した刺激に対して皮質上の同じような位置に対応する受容野を持つことが知られている。このような 2 つの神経場間の連続的な結合関係のことをトポグラフィックマッピング topographic mapping と言う。

視覚野のトポグラフィックマッピングについては、さらに細かいことが分っていて、任意の視覚位置に対して、眼優位性 ocular dominancy, 方位選択性 orientation selectivity, 色などの情報が処理されるように規則正しく配列されている。これをハイパーコラム hypercolumn 構造という。ハイパーコラムは、2次元しかない皮質上に、2次元の網膜位置、方位、視差情報(立体視)、色情報処理などの多次元情報をなるべく効率よく処理しようとする生体情報処理の機構を表していると言える。

このような構造は、大まかな構造は遺伝子によって決定される<sup>1</sup>が、細かい構造については神経回路の自己組織化 self organization によって達成されると考えられている。このような例は枚挙に暇がない<sup>2</sup>。一例として Linsker[21, 20, 19] の行なったシミュレーションを紹介する。

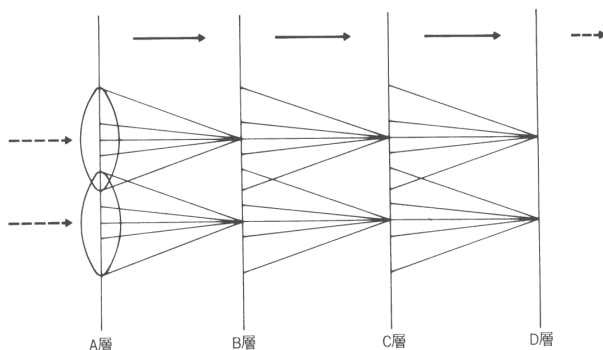


図 10: Linsker のネットワーク [21, 20, 19]

Linsker は図 10 のような数層の細胞で構成されるモデルを考えた。

<sup>1</sup>人間の脳の回路網には  $10^{15}$  個ほどのシナプス結合がある。一方、人間の遺伝子の配列は  $10^9$  程度である。つまり遺伝情報だけでは脳は作れない。そこではまず遺伝情報によって大まかな回路網が作られ、出生直後から様々な刺激を受けることで複雑な脳ができあがるのであろう。このことは、遺伝子の情報が全て解明されれば人間の性格あ心の働きが全て分かってしまうという言明が、誤解に過ぎないことを示している。繰り返すが、ニューロンが作る神経回路網のシナプスは遺伝子が表現している情報よりはるかに多い。

<sup>2</sup>外側膝状体の 6 層構造の形成やら、Blakemore and Cooper の縦縞しか見せなかった子猫の実験とか...

各細胞層では、多数の細胞が 2 次元的に広がっている。B 層の各細胞は A 層の細胞から入力を受ける。たとえば、A 層の円で囲まれた領域にある細胞 100 個がそれぞれ B 層の一個の細胞に結合している。同様にして、B 層から C 層、C 層から D 層へと続く結合も同様である。

結合強度の変化は Hebb の学習則と呼ばれる学習側の変形で入力と出力ともに活動度が高い場合には結合を強める他に、両方とも活動が低ければ結合の強さを弱めるように学習させた。

入力は A 層から入るが、A 層においてはランダムで相関がない活動が自発的に起きていると想定された。その結果、次のようなことが起きた。A から B への結合では、結合の強さは正の大きな値になるか、負の大きな値になるかのどちらかであった。結合が正の興奮性になった場合にどのようなことが起きたのであろうか？ そのとき細胞が観察しているのは、細胞層 A のある領域の入力の平均である。B 層では、隣り合う細胞が受け取る A 層の細胞の範囲は重なりあっている。したがって、この二つの細胞は A 層におけるパターンの似た部分を見ており、ある時刻に一方の細胞の活動度が高ければ、もう一方の細胞の活動度も高い可能性がたかい。つまり、B 層において、近くに位置する 2 つの細胞の活動度の間には相関が生じている。

次に B 層から C 層への結合の変化が図 11 に示されている。

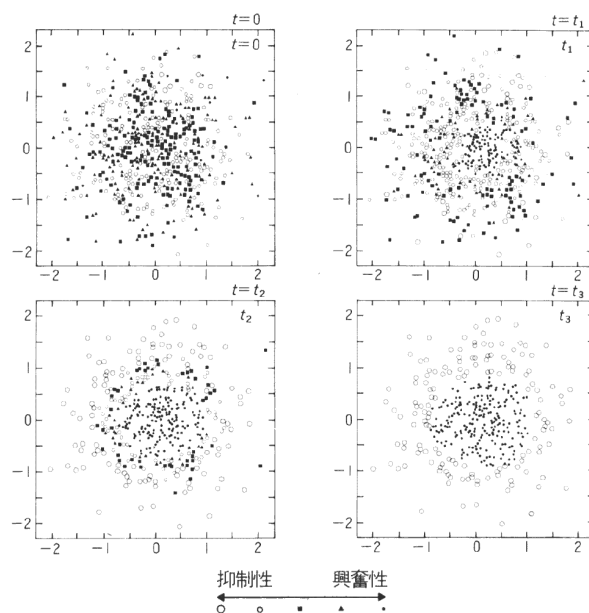


図 11: Linsker のシミュレーション結果 [21, 20, 19] より

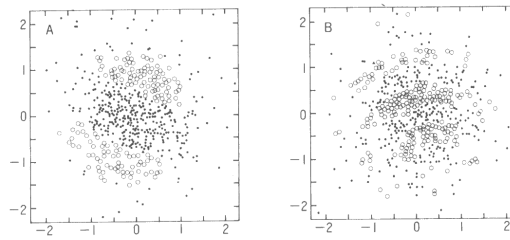


図 12: Linsker のシミュレーション結果 [21, 20, 19]。方位選択性の受容野が形成されている。すなわち知識が創発されたと考えてよいのではないか。

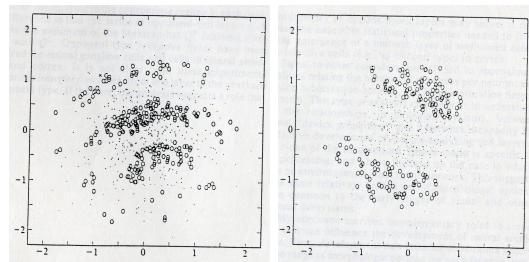


図 13: Linsker のシミュレーション結果 [21, 20, 19]

図 12 では、C 層の一つの細胞を取り上げ、そこから B 層をながめている。つまり C 層の一つの細胞へ入力する B 層の細胞を見ているわけである。シミュレーション開始時 ( $t=0$ ) にはすべての結合はランダムであるが、時間が経過するにつれて中央に点が見えてくる (興奮性の領域)。周辺には抑制性の領域が現れる (オン中心オフ周辺型細胞)。すなわち、ランダムな結合と Hebb 則だけを使って、on 中心型細胞が生じることが示されたことである。さらに先の層には方位選択性の細胞が現れる。

Linsker の提案したネットワークを追試するのは、非常に簡単なプログラムで可能である。にもかかわらず、第一次視覚野での細胞の特徴をよくとらえた優れたデモンストレーションと言える。Linsker のシミュレーションの意義を考えると、「神経系の動作原理は極めて単純であるが、その結果で上がった機構は複雑な特性を持つ。」といえるのではないかと、ということになる。すなわち我々の認識の基礎にある脳内過程は意外と単純原則に従って、組織化が起こっており、その結果として複雑な認識が成立するようになるのではないかと。

一人一人の顔が異なるように、脳もそれぞれ個性がある。これは遺伝情報と出生直後から積み重ねてきた経験によって、それぞれの脳は大きく異なる

ことを意味する。実際、脳細部を比較してみるとそれぞれの脳で一对一の対応が取れないことも多い。もしそうであるなら、脳の同じような場所に同じ程度の損傷を受けたとしても、同じ障害が起きるとは考えがたい。脳とその神経回路網はそれぞれ千差万別であると考えの方が自然である。

自己組織化の特徴を抽象化してとらえれば、多次元刺激をその刺激の持つ規則性に従って 2 次元の皮質上への対応問題ととらえることができる。入力層の空間多次元多様体から 2 次元部分空間への写像という。

### 13 ニューラルネットワークへの批判

ニューラルネットワーク批判には 2 つの点を挙げるができる。一つはモデルが特定の現象を解釈するための post hoc なモデルにすぎないのではないかという点である。Post hoc なモデルであっても、今まで信じて疑われなかった理論に対する alternative を提出し、問題の再考を迫っている点は評価されるべきであると考え。ある分野の科学的知識が深まるとは、絶えず alternative からの挑戦を受け、それに答える形で理論が洗練されて行くことなのだろう。

もう一つの批判はニューラルネットワークのプログラムの多くはおもちゃのように小さく、実際の脳と比べると著しく見劣りするという批判である。確かに、このスケーラビリティの問題は深刻であると言って良い。かつてミンスキー Minsky とパパート Papert [25] が批判したとおり、蟻の体型をそのまま拡大して象の大きさにしたのでは、おそらく巨大な蟻は自分の体重を支えきれずに動けないであろう。地球の重力に反して体を支えるためには象のような太さの足が必要なのである。すなわち現在のニューラルネットワークプログラムを単に拡張しただけでは解決できない問題が存在するに違いない。しかし、このことはニューラルネットワーク研究を全否定することにはならないだろう。スケーラビリティの問題を解決するためにはどのような手法が有効であるのかを議論してゆくのはこれからの課題である。

### 参考文献

- [1] J. A. Anderson. Cognitive and psychological computation with neural models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5):799–815, 1983.
- [2] Jean P. Banquet, Philippe Gaussier, Jose L. Contreras-Vidal, Angelika Gissler, Yves Burnod, and Debra L. Long. A neural network model of memory, amnesia, and cortico-hippocampal interactions. In R. W. Parks, D. S. Levine, and D. L. Long, editors, *Fundamentals of neural*



*network modeling: Neuropsychology and cognitive neuroscience*, pages 77–119. Mit Press, Cambridge, MA, 1998.

- [3] R. S. Bapi and D. S. Levine. Modeling the role of the frontal lobes in sequential performance. i. basic structure and primacy effects. *Neural Networks*, 7:1167–1180, 1994.
- [4] G. D. A. Brown and R. L. Loosemore. Computational approaches to normal and impaired spelling. In G. D. A. Brown and N. C. Ellis, editors, *Handbook of spelling: Theory, process and application*. John Wiley & Sons, Chichester, 1994.
- [5] M. A. Burton, A. W. Young, V. Bruce, R. A. Johnston, and A. W. Ellis. Understanding covert recognition. *Cognition*, 39(2), 1991.
- [6] G. Carpenter and S. Grossberg. Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16:131–137, 1993.
- [7] J. D. Cohen, K. Dunbar, and James L. McClelland. On the control of automatic processes: A parallel distributed processing model of the stroop effect. *Psychological Review*, 97(3):332–361, 1990.
- [8] J. D. Cohen, M. J. Farah, R. D. Romero, and D. Servan-Schreiber. Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience*, 6(4):377–387, 1994.
- [9] Jonathan D. Cohen and D. Servan-Schreiber D. Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99:45–77, 1992.
- [10] G. Dell, M. Schwartz, N. Martin, E. Saffran, and D. Gagnon. Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 1997.
- [11] Martha J. Farah and James L. McClelland. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120(4):339–357, 1991.
- [12] Martha J. Farah, Randall C. O’Reilly, and Shaun P. Vecera. Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review*, 100(4):571–588, 1993.

- [13] Matha J. Farah. Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, 17:43–104, 1994.
- [14] B. Gordon. Confrontation naming: Computational model and disconnection simulation. In M. A. Arbib, D. Caplan, and J. C. Marshall, editors, *Neural models of language processes*. Academic Press, New York, 1982.
- [15] T. A. Harley and S. B. G. MacAndrew. Modelling paraphasias in normal and aphasic speech. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 378–383, San Mateo, 1992. Morgan Kaufmann.
- [16] Donald O. Hebb. *Organization of behavior*. John Wiley & Sons, Inc., New York, USA, 1949.
- [17] D. Horn, E. Ruppin, M. Usher, and M. Herrmann. Neural network modeling of memory deterioration in alzheimer’s disease. *Neural Computation*, 5(5):736–749, 1993.
- [18] G. W. Humphreys, T. Freeman, and H. J. Muller. Lesioning a connectionist model of visual search: Selective effects on distractor grouping. *Canadian Journal of Psychology*, 46:417–460, 1992.
- [19] R. Linsker. From basic network principles to neural architecture: Emergence of orientation columns. *Proc. Natl. Acad. Sci. USA, Neurobiology*, 83:8779–8783, 1986.
- [20] R. Linsker. From basic network principles to neural architecture: Emergence of orientation-selective cells. *Proc. Natl. Acad. Sci. USA, Neurobiology*, 83:8390–8394, 1986.
- [21] Ralph Linsker. From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. USA, Neurobiology*, 83:7508–7512, 1986.
- [22] David Marr. *Vision*. W. H. Freeman and Company, San Francisco, 1982. ビジョン , 乾敏郎・安藤広志訳 , 1987 , 産業図書.
- [23] James L. McClelland, Bruce L. McNaughton, and Randall C. O’Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995.

- [24] James L. McClelland and David E. Rumelhart. Amnesia and distributed memory. In James L. McClelland and David E. Rumelhart, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*, volume 2, pages 503–527. MIT Press, Cambridge, MA, 1986.
- [25] Marvin Minsky and Seymour Papert. *Perceptrons, Expanded Edition*. MIT Press, Cambridge, MA, 2 edition, 1988. パーセプトロン, 中野馨, 坂口豊訳, パーソナルメディア, 1993.
- [26] M. C. Mozer and M. Behrmann. On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, 2(2):96–123, 1990.
- [27] K. E. Patterson, M. S. Seidenberg, and J. L. McClelland. Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris, editor, *Parallel distributed processing: Implications for psychology and neuroscience*, pages 131–181. Oxford University Press, 1990.
- [28] David C. Plaut. A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In Morten H. Christiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 8, pages 244–278. Ablex Publishing, Westport, CT, 2001.
- [29] David C. Plaut, James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103:56–115, 1996.
- [30] David C. Plaut and Tim Shallice. Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5(1):89–117, 1993.
- [31] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958. In J. A. Anderson and E. Rosenfeld (Eds.) *Neurocomputing* (1988), MIT Press.
- [32] John R. Searle. Minds, brains, and programs. In Douglas R. Hofstadter and Daniel C. Dennett, editors, *The Mind's I — Fantasies and Reflections on Self and Soul*, chapter 22. Basic Books, 1980. サール「心・脳・プログラム」. ホフスタッター, デネット編著. 坂本百大監訳「マインズ・

アイ (下) —コンピュータ時代の心と私—」. 第 22 章 p. 178-210. TBS  
ブリタニカ . 1992.

[33] 中野馨. 脳をつくる —ロボット作りから生命を考える. 共立出版, 1995.