記憶と想起の物理学

人間文化研究科 複合現象科学専攻 上江洌 達也

概要:

何かを覚えたり (記憶)，思いだす (連想) ことができるのは脳の働きによることは誰でも知っていますが，その仕組みはどうなっているのでしょうか? 人の脳は，大きさが 0.1mm 程度の神経細胞 (ニューロン) が約１３０億個集まってできています。これらのニューロンどうしはつながりあってネットワークを形成しています。このようなニューロンのネットワークを簡単化したモデルを用いて，コンピュータによる文字の連想などの模擬実験を行ない，記憶の埋め込みや連想などの仕組みについてわかりやすく解説します。

I. 脳 (神経系) についての概要

脳は，約１３０億個の神経細胞と約１０００億個のグリア細胞からなっています。図１。

神経細胞 (ニューロン)：情報のにない手，信号伝達，信号処理。

グリア細胞：ニューロンの補佐役。 栄養の供給等。

神経細胞は，お互いに結合してネットワークを形成しています。

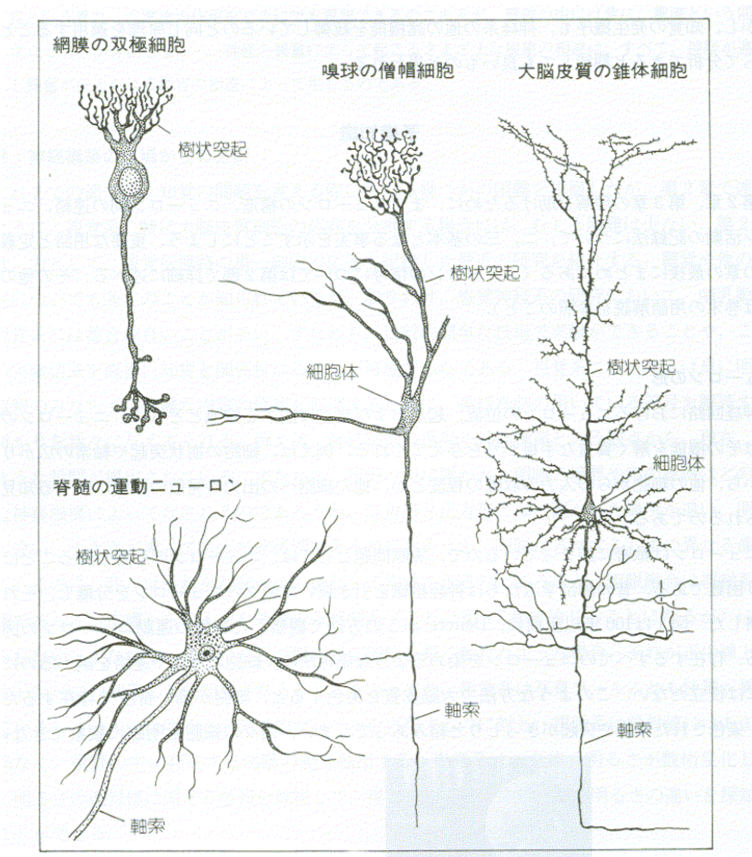


図 1: いろいろな神経細胞 (「ニューロンから脳へ」，クフラー他，廣川書店 [1])

図 2: 神経細胞の写真。理研 脳科学総合研究センター 保賀 宣子博士 提供

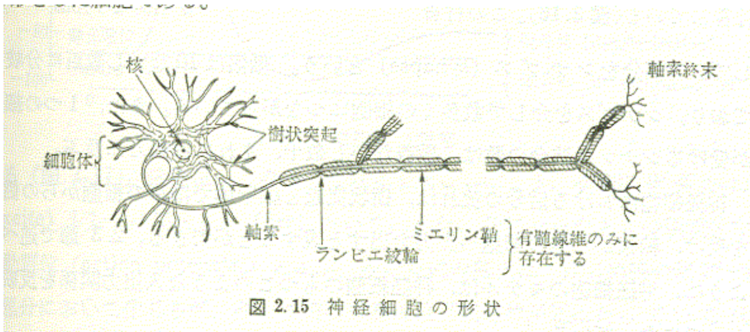


図 3: 一個の神経細胞の模式図 (「ニューロンコンピュータの基礎」，中野他，コロナ社 [2])

a. 1 個の神経細胞

神経細胞は，細胞体，樹上突起，軸索からなり，軸索の末尾の軸索終末は，ボタン状にふくらんで，次のニューロンの細胞体ないし樹上突起と接触して シナプスという構造を形成しています。図２，３。

細胞体: 数ミクロン ~ 数 10 ミクロン $(1 \text{ ミクロン} = \frac{1}{1000}\text{mm})$

樹上突起: 数 10 ミクロン ~ 1mm

軸索: 長さはニューロンによってまちまち。ヒトの脊髄運動ニューロンには 1m に達するものもあります。10 ~ 数 100 に分岐し，シナプスを介して数多くの細胞につながっています。

１つの細胞は，シナプスを介して数百 ~ 数万個の細胞と結合しています。

b. 信号伝達の仕組み

イオンチャンネル: 細胞膜には，K^+,Na^+,Ca^{2+}, Cl^- などについて，そのイオンのみを通すことができる孔 (チャンネル) がたくさん形成されており，細胞はこれらのチャンネルを必要に応じて開閉しています。これらの孔は，イオンチャンネルと呼ばれ，イオンの種類に応じて，K-チャンネルとか Na-チャンネルと呼ばれています。

膜電位: 神経細胞の細胞膜の内部と外部でのイオンの濃度が違うため，通常は外部 (電位 0 とする) より内部の電位が低く (約 -70mV) なります。この電位差を膜電位といいます。

脳における信号伝達は以下のプロセスで起こります。

1. 活動電位の発生

   膜電位がある値 (閾値，約 -50mV) を越えると，電位が約 20mV まで上昇し，活動電位と呼ばれる一過性のスパイク様の電位が生じます。これを，興奮 または，発火といいます。活動電位は，軸索を減衰せずに，速さ 0.5m/s ~ 100m/s で伝導します。

2. 活動電位が軸索終末に到達すると，Ca-チャンネルが開き，Ca^{2+} が流入します。すると終末内に存在する小さな袋 (シナプス小包) がシナプス末端部まで移動します。シナプス小包には，神経伝達物質が含まれており，シナプス小包の膜は，シナプス末端の膜と融合して破れ，神経伝達物質が細胞外に出ます。

3. シナプス前細胞とシナプス後細胞の間の非常に小さいすきま，シナプス間隙を 神経伝達物質が拡散します。

4. 神経伝達物質は，シナプス後細胞の樹上突起や細胞体にある レセプター (受容体) に作用して，膜電位に変動を起こします。

5. 多数のシナプスからの入力の総和が，シナプス後細胞の活動を決定します。膜電位が閾値を越えると，発火して先へ信号が伝わります。

シナプスの種類

興奮性シナプス : 膜電位を上昇 (脱分極) させる神経伝達物質を放出。

抑制性シナプス : 膜電位を減少 (過化脱分極) させる神経伝達物質を放出。

II. 脳 のモデル – ニューラルネットワークモデル

前節の内容をまとめると，脳の情報伝達機構は次のように考えられます。脳は，極めて多数のニューロンがシナプスを介して結合してネットワークを形成しており，各ニューロンの膜電位が閾値を越えると，発火して活動電位が生じ，その電位の作用によって，シナプスを介して接続しているまわりのニューロンの膜電位も変化します。このようにして，ニューロンの発火が次々に伝わっていきます。

a. 1個の神経細胞のモデル化

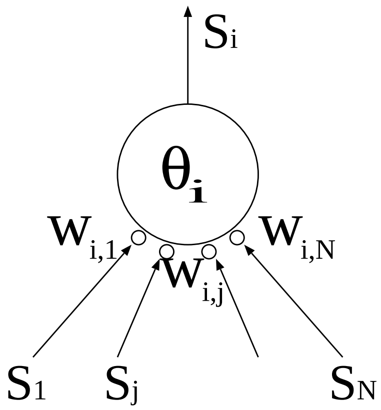


図 4: MaCulloch-Pitts モデル

1943年，マッカロとピッツは，脳の情報処理機能を数理的に扱うために，神経細胞の細かな構造は無視して情報伝搬のプロセスを単純化した次のようなニューロンのモデルを提案しました。(MaCulloch-Pitts モデル [4])。図4。

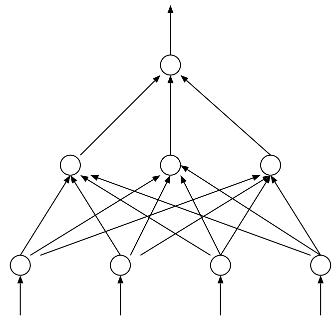
このモデルでは，ニューロンの状態を変数 S で表し，発火しているときは $S=1$，発火していないときは，$S=0$ とします。j 番目のニューロンから i 番目のニューロンへのシナプス結合の強度（シナプス結合荷重）は w_{ij} で表わします。また，i 番目のニューロンの閾値を θ_i とします。時刻 t での全ニューロンの状態を $(S_1(t), S_2(t), \cdots, S_N(t))$ とすると、次の時刻 $t+1$ に i 番目のニューロンが発火するか否かは、次の規則にしたがって決定されます。

$$\Sigma_{j=1}^{N} w_{ij} S_j(t) \geq \theta_i \text{ なら } S_i(t+1) = 1(\text{発火}),$$
$$\Sigma_{j=1}^{N} w_{ij} S_j(t) < \theta_i \text{ なら } S_i(t+1) = 0(\text{発火せず}). \quad (1)$$

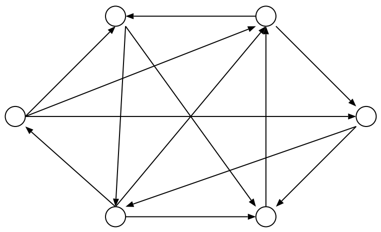
b. 多数の神経細胞のネットワーク

マッカロ-ピッツ ニューロンを多数結合させてネットワークが構成されますが，その構造として，フィードバックのないフィードフォワード型と，フィードバックのある自己結合型が考えられます。図 5 。

これらのモデルで，物理学の立場から，想起過程や記憶容量の計算，あるいは学習の問題などが，統計力学的手法を用いて解析されています。この講座では，特に記憶の埋め込みや想起のモデル，連想記憶モデルについて解説します。



(a)



(b)

図 5: (a) フィードフォワード型, (b) 自己結合型

c. 学習

さて，このようにネットワークを形成している脳において，記憶，学習はどこでどのようにして行われるのでしょうか? ある特定の細胞が，ある特定の記憶を担っているという考え方は，「おばあさん細胞」説と呼ばれています。この説だと，その細胞が壊れてしまうと，おばあさんを思いだせなくなってしまいます。別の考え方は，多数のニューロンが記憶に関与しているというものです。その場合には，一部の細胞が壊されただけで記憶が失われるということはありません。おばあさん細胞に相当する現象もあると考えられていますが，ここでは，多数のニューロンによる，記憶，学習のメカニズムについて考えます。

それには，まず，学習とは何かということに答えなければならないでしょう。極めて一般的に考えると，外からの刺激に対して，脳の応答が変化することを学習と呼んだらいい

と思われます。このように考えると，学習することによって，脳の中の何かが変化するはずです。上記の脳の構造から，もっとも自然に変化を考えることができる部位は，シナプスです。シナプス結合の強度が変化することによって学習が行われるという仮説は，シナプス可塑性説とよばれています。具体的には，ヘブによってとなえられたヘブ則に従って変化が起こると考えられています。

ヘブ則

１９４９年に Hebb によって提案された仮説 [3]。シナプスを介して結合している２つの神経細胞が同時に発火するという現象が繰り返し起こると，そのシナプス結合の強度が増大する。

III. 連想記憶モデル

本節では，記憶の埋め込みや記憶の想起が，ニューラルネットワークモデルでどのようにして実現できるかを解説します。

便宜上，i 番目のニューロンが発火している状態を $S_i = 1$，発火していない状態を $S_i = -1$ で表します。N 個のニューロンがお互いにシナプス結合によってネットワークを形作っている自己結合型のモデルを考えます。j 番目のニューロンの軸索終末と，i 番目のニューロンの細胞体または樹上突起とのシナプス結合の強さを w_{ij} とします。このネットワークに学習を行なわせると，w_{ij} が変化するはずですが，その結果，w_{ij} はどのような値をとるのでしょうか?

もっとも自然な仮定は，w_{ij} の値が，脳研究で明らかになったヘブ則を満たすように変化すると考えることです．

ここでは，例として， 50×50 の 2,500 個のニューロンを考え，それらが正方形の格子の上に並んでいるとします。このネットワークにいくつかの文字を学習させます。例えば，「想」という文字を学習させたいとします。簡単のため，「想」という文字を思いだしているとき，図６ (a) のように，発火パターンが「想」の文字になるとします。つまり，図の赤い部分を $S_i = 1$ とします。このとき，「想」の文字は N 個の 1 と -1 を並べた数の列 (N 次元ベクトル) で表わされます。これをパターンベクトルといい，例えば$\boldsymbol{\xi}^1$ で表わします。

$$\boldsymbol{\xi}^1 = (\xi_1^1 = -1, \xi_2^1 = 1, \cdots, \xi_N^1 = 1)$$

ヘブ則では，発火しているニューロンどうしは強めあうので，いま，「想」という文字のパターンベクトル $\boldsymbol{\xi}^1$ において，i 番目と j 番目のニューロンの値，ξ_i^1 と ξ_j^1，が発火しているときに強めあうようなシナプス結合の値を考えると，$w_{ij} = \xi_i^1 \xi_j^1$ とすればよいことがわかります。実際，共に発火しているときは，$\xi_i^1 = 1, \xi_j^1 = 1$ で，$w_{ij} = 1$ となり，いずれ

か一方のみが発火しているときは，$w_{ij} = -1$ となります。従って，前者の方が後者よりもシナプス結合荷重の値は大きくなっており，ヘブ則が満たされています。今の場合は，共に発火しない場合も $w_{ij} = 1$ となりますが，これは，拡張されたヘブ則とみなすことができます。

上の例は，1 個のパターンベクトル $\boldsymbol{\xi}^1$ を記憶する場合でしたが，一般に，p 個のパターンベクトル，$\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \cdots, \boldsymbol{\xi}^p$ を記憶する場合には，記憶が重ね合わされると考えて，

$$w_{ij} = \frac{1}{N}(\xi_i^1 \xi_j^1 + \xi_i^2 \xi_j^2 + \cdots + \xi_i^p \xi_j^p)$$

ととります。

このモデルは，ホップフィールドモデル [5] と呼ばれています。このモデルで，実際にパターンを埋め込むことが可能かどうかを見てみましょう。

まず，1 個のパターン「想」のみをヘブ則で埋め込んだ場合 (図 6 (a))，「想」という文字が崩れた状態 (図 6(b)) からはじめて，「想」という文字を思いだすことができるかを，コンピュータで調べてみると，図 6 (c)，(d) に示したように，ちゃんと思いだせることがわかります。これは，手計算によっても簡単に示せます。では，記憶させるパターンの数 p を増やしていくとどうなるのでしょうか? パターンを 1 つ増やして，「生」とうパターンも埋め込んでみましょう。すると，この場合も，問題なく思いだせることがわかります。では，いくらでも記憶させることができるのでしょうか? あるいは，ある決まった値までしか記憶できないのでしょうか? パターンの数 p の値を増やしたときに，「想」の文字から出発して，その文字のまま留まるかどうかを調べると，ある個数 $p = p_c$ までしか記憶できないことがわかります。

ニューロンの数 N を変えてコンピュータを用いて調べると，記憶できる限界の値 p_c は，N に比例することがわかります。実は，$\frac{p_c}{N} \simeq 0.14$ となることがわかります。この比の値は，物理の理論を用いて理論的に計算することができ，数値実験の結果と同じく，約 0.14 となることを示すことができます。

VI. まとめ

これまでみてきたように，ニューロンの特性を抽出したニューラルネットワークモデル，ホップフィールドモデルによって，記憶の埋め込みや想起が実現できることがわかりました。ホップフィールドモデルは，現実の脳に比べて簡単化されているとはいえ，理論的な解析が可能で，物理学者による脳研究の理論的なアプローチへの道を開いたという意味においても，極めて重要なモデルです。その後の研究で，物理学で発展させられてきた洗練された手法が，ニューラルネットワークの理論的解析において重要な役割を果たすこと

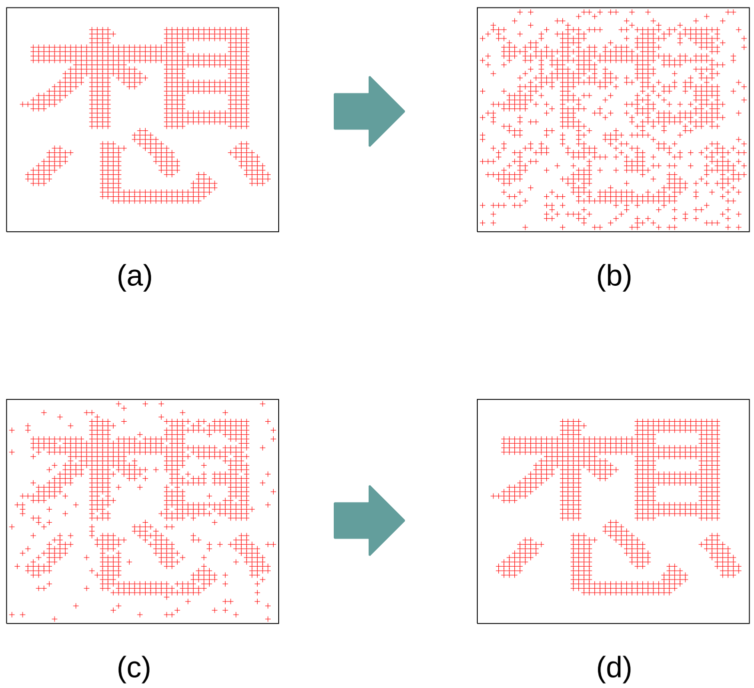


図 6: 50×50 個のニューロンによる 文字「想」の記憶と，想起過程。(a) 埋め込んだパターン，(b) 乱れたパターン，(c) 時刻 $t = 1$，(d) 時刻 $t = 9$。

が明らかになっています。脳は，複雑系の典型例とされますが，脳にとどまらず，複雑系や，生命科学，情報科学への物理からのアプローチが今後ますます盛んに行なわれることと思われます。

最後に，一般向けの解説書 [6] を参考文献として追加します。

参考文献

[1] クフラー他，「ニューロンから脳へ」，廣川書店 (1998).

[2] 中野馨編著，ニューロコンピュータの基礎，コロナ社 (1990).

[3] D. O. Hebb, *The organization of behavior*, (Wiley, 1949).

[4] W. S. MaCulloch and W. H. Pitts, Bull. Math. Biophys., Vol.5,(1943)115.

[5] J. J. Hopfield, Proc. Natl. Acad. Sci., **79**(1982)2554.

[6] 池谷裕二，記憶力を強くする，ブルーバックス (2002)；松村道一「脳科学への招待」，サイエンス社 (2002).