

解説

ニューロンモデルを用いた文字認識

朱 小燕・岩瀬 泰章・神保 孝志・梅野 正義

名古屋工業大学工学部電気情報工学科 〒466 名古屋市昭和区御器所町

(1989年1月9日受理)

Pattern Recognition by Neural Model

Xiao-yan ZHU, Yasuaki IWASE, Takashi JIMBO and Masayoshi UMENO

Department of Electrical and Computer Engineering, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya 466

1. ま え が き

生物の脳は膨大な数の基本素子ニューロン(神経細胞)からなる情報処理装置である。ニューロンは複雑に結合し、ニューラル・ネットワーク(神経回路網)になり、高度な並列分散処理を行なっている。

近年、脳の構造や情報処理メカニズムにヒントを得、脳の優れた情報処理能力を実現しようとする研究が盛んになってきた¹⁻³⁾。脳や目などの情報処理手法をLSIチップ上に作り込む試みも始められた⁴⁻⁶⁾。

パターンを認識するために多くの方法が考案されているが^{3,7)}、いずれも、脳の構造をまねたニューロン結合構造と複雑な結合荷重関数を用いたニューラル・ネットワークからなる多層構造で、さらに、なんらかの学習アルゴリズムで荷重を変えてパターンを認識できるようなモデルである。したがって、モデルの柔軟性とハード化の容易さは両立しない。柔軟性をよくするとモデルが複雑になり、また学習機能をもたせると多数のニューロンの結合を可変にしなければならない。このようなモデルをハード化しようとする、配線の数が増えるため回路作成が困難になり⁸⁾、チップ内に入れられるニューロン数もかなり少なくなる。

われわれは、情報が脳に入る前に目から脳までの間でいろいろな処理がなされていることを考え、生体の視覚系と脳における生理学的な知見を基にして、単純化した工学的モデルを用いた文字パターン認識システムを提案した。このシステムは大きく分けて三つの部分からなる多層ネットワークである。まず、網膜層で特徴点を検出

し、特徴検出層で特徴点付近の状態を判断した後、認識部分で認識を行なう。各層は異なるタイプのユニットをもち、入力パターンが出力層へ伝播していくとき、各中間層に現われるユニットの活動パターンは独特な意味をもつ。認識 C^1 層と C^2 層のユニットの結合係数を変化させることによりシステムを学習させることができる。このモデルは簡単で、学習により変化する部分が少ないため、ハード化には有利である。また、パターンの特徴で認識するので、パターンの大きさや位置ずれなどに対して柔軟性がある。

2. 生体の視覚系と認識組織

人間はものを見る時、まず、網膜からパターン情報を受け取って、外側膝状体を介して脳に伝え、そこで情報を解釈、同定、認識する。

網膜は多層構造である。感覚器(目の網膜の光受容器)からの出力は複雑な細胞層に入り、そこで情報のかなりの部分が処理されている。その一連の網膜層の中の最後の細胞は神経節細胞と呼ばれ、特定のパターンにしか応答しない。一般に、一つの神経節細胞は網膜上の特定領域にある多数の細胞とつながり、その領域が刺激された場合にのみ応答する。神経節細胞の応答領域をその細胞の受容野(receptive field)という。受容野は細胞によって異なる。神経節細胞には多くの種類があり、そのうち、光がついた瞬間に活動が高まる on型を中心にもち、光が消えた瞬間に活動が高まる off型を周辺にもつオン・センター、オフ・サラウンド(on中心型)タイプの細胞がある。これと逆のタイプの細胞はオフ・センター、

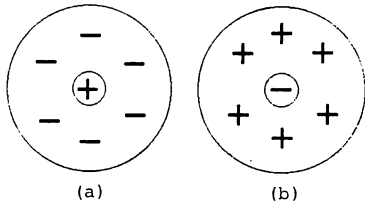


図1 受容野
(a) on type, (b) off type

オン・サラウンド (off 中心型) の細胞という。この両タイプの神経節細胞の受容野の図を図1に示す^{9,10)}。

網膜上に映った刺激はさまざまな神経節細胞で処理され、神経に沿って脳に伝わってくる。脳の中で情報処理を担っている基本素子は神経細胞である。一つの神経細胞はシナプスを介して $10^3 \sim 10^4$ もの他の神経細胞と結合し、神経系を構成している。神経細胞のそれぞれの結合の強さは年齢・経験・知能などによって違い、人間の成長につれて、網膜、脳の神経回路、つまりニューラル・ネットワークが自己形成される。これは、人間や生物と機械あるいはいままでのコンピュータとの基本的な違いである¹¹⁾。

3. システム構成

われわれは生体の網膜と脳にある神経細胞をまね、パターン認識システムを提案した。このシステムは網膜層、特徴検出層および認識部分からなる階層的なネットワーク構造になっている。網膜層は入力パターンの特徴点(交差点や端点など)を検出する。特徴検出層では検出された特徴点の特徴を判断し、バックプロパゲーション系の学習機能をもつ認識部分でパターンを識別する。システム構成の模式図を図2に示す。

各細胞層は、ユニットと呼ばれる多入力/一出力型の閾値特性をもつ飽和型非線形素子を n_k 個並べて構成される。これらのユニットは生体内でのニューロンに相当する動作をする。モデル化されたものを図3に示す。 k 番目の細胞層の i 番目のユニットの出力 O_i^k は

$$O_i^k = F^k(u_i^k) \quad (1)$$

$$u_i^k = \sum_j W_{ij}^{k-1} O_j^{k-1} - \theta^k \quad (2)$$

である。ここで、 θ^k は閾値、 W_{ij}^{k-1} は $k-1$ 層の j 番目のユニットと k 層の i 番目のユニットとの間の結合係数である。

各層におけるユニットは入力層から中間層、中間層から出力層という向きに結合しており、各層内での結合および出力層から入力層に向かう結合は存在しない。 k 番

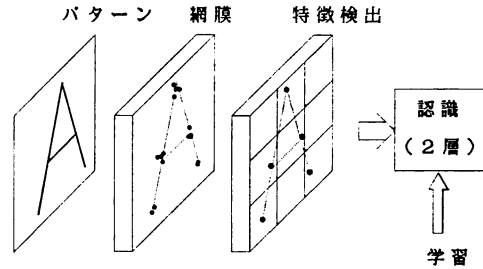


図2 文字認識システム

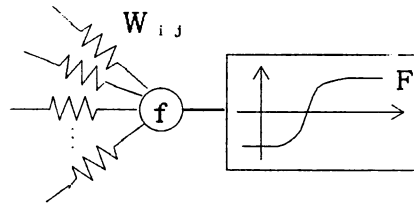


図3 ユニットのモデル

目の層にある i 番目のユニットは $k-1$ 番目の層のユニットのうち、ある領域 S_i^k に存在する一群のユニットからだけ入力を受けている。 S_i^k は $k-1$ 層にあるユニットのうち、 k 層の i 番目のユニットに結合しているユニットの集合を表わし、これを受容野と呼ぶことにする。一つの i に対して、

$$W_{ij}^{k-1} \neq 0 \quad (j \in S_i^k)$$

となる。層の動きにより、ユニットの受容野の大きさは異なる。層と層の間のユニットの結合には、初期状態においてすでに結合が完成している固定結合と外部から与えられる刺激の状態に応じて結合の強さが変化していく可変結合の両者がある。可変結合をもつ部分の結合係数が学習によって変わり、システムの学習機能を実現する。

各ユニットは入力 O_j^{k-1} に対する活動レベル u_i^k をもっている。ユニットは活動レベルにより、出力関数 F^k を通して出力 O_i^k を出す。この出力がまた、次の層の多数のユニットの入力として使われる。各層はそれぞれの機能に応じた出力関数をもつ。たとえば、認識層の場合はシグモイド関数が出力関数として用いられた。

3.1 網膜層

文字パターンの場合には端点や交差点などが特徴点としてみられる。このような特徴点を検出するために、on 中心型と off 中心型神経節細胞の工学的モデルをユニットとして使った。このような細胞のモデルはエッジ検出にも使われている¹²⁾。図4は棒状およびT字状の刺激パターンとユニットの反応の位置関係を示している。on 中心型の場合では、棒状パターンの端部で強い信号が出

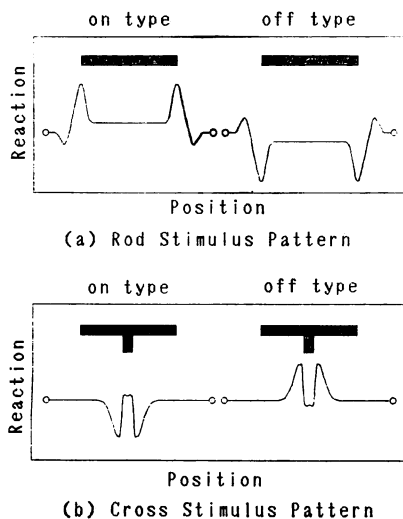


図4 刺激パターンの位置に対するユニットの反応

るが、off 中心型の場合では、T字状のパターンの交差点の付近で強い信号が出る^{*1}。網膜層はこのような2種類の細胞を2次元配列した処理面で、入力されたパターンの端点や交差点を検出する処理を行なう。

ユニットの出力は次の式になる。

$$O_i^1 = F^1 \left(\sum_j^{N^1} W_{i^1, j^0} X_j - \theta^1 \right) \quad (3)$$

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

ここで、0層の細胞 X_j は入力パターン、 N^1 は受容野のサイズである¹³⁻¹⁵⁾。

出力がゼロより大きい場合はこのユニットが興奮し発火しているといい、発火している細胞を発火点と呼ぶ。パターン信号は網膜を通して発火信号になり、次の層へ伝わる。

3.2 特徴検出

網膜層は互いに重ならない領域に分けられ、各領域の出力は特徴検出層の一つのユニットに結合される。特徴検出層ではこの領域(特徴検出層の受容野)内で最大の出力をもつ点を特徴点として抽出すると同時にその種類を判断、分類する。

特徴を判断する方法は多くあるが、本モデルでは特徴点から出る入力パターンの線分の数で特徴を分類し、判断する。特徴検出層のユニットの出力は特徴点として選ばれた点から出る線分の数に比例し、その大きさが特徴の種類を表わす。パターン情報は特徴情報になり、認識

*1 on 中心型では“孤立した点”，off 中心型では“孤立した点”でも大きな出力が出るが、これらは後処理過程でノイズとして消去される。

層へ伝わる。

3.3 認識

認識部分は C^1 と C^2 の二層からなり、それぞれユニットの数が異なる。 C^1 層は隠れ層として動作し、任意のパターンを認識できる能力をもたせるために加えたものである。

出力関数としてはシグモイド関数

$$F^k(u) = 2 / \{1 + \exp(-u)\} - 1$$

が用いられるのが一般的であるが、シミュレーションを高速化し、かつハード化を容易にするために、 C^1 層の出力関数 F^3 として次のような3値関数を用いた。

$$F^3(u) = \begin{cases} 1 & u > a \\ 0 & -a < u \leq a \\ -1 & u \leq -a \end{cases} \quad (4)$$

ここで、 u は式(2)で示したユニットの活動レベルである。 a は学習した後の結合係数で決める。 C^2 層(出力層)の各ユニットはそれぞれ一つのパターン(たとえば文字)を表わす。そのなかで出力値が最大になるユニットがシステムの出力となる。

4. 学習方式

さまざまなデータ処理を実行し、柔軟性をよくするには、ネットワークの結合係数を適応的に変化させる必要があるが、多層構造のネットワークの全ての結合係数を変化させることはたいへんなことである。とくにハード化するには固定結合と可変結合ではその結合が占める空間も必要な技術もまったく異なる。たとえば、固定連想メモリの場合、22個のニューロンを有するLSIは100 μm 角のチップ内に収められるのに対して、プログラマブル連想のメモリの場合には、一つのシナプス結合に100 μm 角の面積を要する⁹⁾。

構造を簡単にするため、本モデルでは C^1 層および C^2 層のユニットの結合係数だけをバックプロパゲーション(誤差逆伝播)学習法により変化させるようにした¹⁶⁾。

バックプロパゲーション法とは、教師ありの学習で、ネットワークに誤差フィードバックを導入し、このフィードバックに基づいて、結合係数を調整するのである。結合係数の修正量は

$$\Delta W_{i^k, j^{k-1}}(t+1) = \eta d_i^k O_j^{k-1} + \alpha \Delta W_{i^k, j^{k-1}}(t)$$

である。ここで、 O_j^k k 層の j 番目のユニットの出力、 t は学習段階を表わす変数、 η は学習速度を決める学習定数、 α は学習速度が大きい場合にも安定に動作するように導入された安定化定数である。

C^2 層の誤差信号 d_i^m は、教師信号 T_i と C^2 層の出

力 O_i^m を用いて

$$d_i^m = (T_i - O_i^m) F^{m'}(u_i^m)$$

C^2 以外では

$$d_i^k = \left(\sum_j d_j^{k+1} W_{ij}^{k+1} \right) F^{k'}(u_i^k)$$

で定義される。ここで、 $F^{k'}$ は出力関数 F^k の微分である。

第3層 (C^3 層) の出力関数 F^3 が微分できないので、信号処理流れと誤差修正流れの数学処理を分け、誤差信号の計算には、次の関数

$$F(u) = 2 / \{1 + \exp(-u)\} - 1 \quad (4')$$

を用いた。

学習の結果は出力層における二乗誤差評価関数で評価する。 E_p を p 個の入力パターンに関する二乗誤差

$$E_p = \sum_i \sum_j \{T_j(i) - O_j^m(i)\}^2$$

とする。ここでは、 $O_j^m(i)$ はパターン i に対する出力層の j 番目ユニットの出力であり、 $T_j(i)$ はこれに対する教師信号である。 E_p が極小になるまで、結合係数を変化させ、学習続ける。

学習の収束を速くする方法はいろいろ検討されている^{17,18)}。 η が大きいほど収束が速いが、大きすぎると振動してしまう。 α は振動を減衰させるが、大きすぎると遅くなる。いろいろな値で実験した結果、それらの定数の最適値はパターンの数にも、誤差の大きさにも関係がある。本研究では誤差が小さくなるとともに η と α を徐々に大きくすることにより、振動せず速く収束させることができた。全てのパターンに対してネットワークのダイナミクスを物理的に理解して定数を決めることができないので、各パラメータは試行で設定している。これは今後の課題となる。

5. シミュレーション

これまで述べてきたシステムによるアルファベット大文字認識のシミュレーションをパーソナルコンピュータで行なった。実際のパラメータは表1に示されている。まず、前処理部分(網膜層、特徴検出層まで)のパターンの特徴を分離する能力を確かめるために、文字パターンについてそれぞれの特徴を取り出し、特徴ベクトルにして、性質を調べた。実際、6人の書いた文字について

表1 シミュレーションのパラメータ

	網膜層	特徴検出層	C^1 層	C^2 層
ユニット数	49×49	3×3	15	13
受容野サイズ	21	16×16	3×3	15

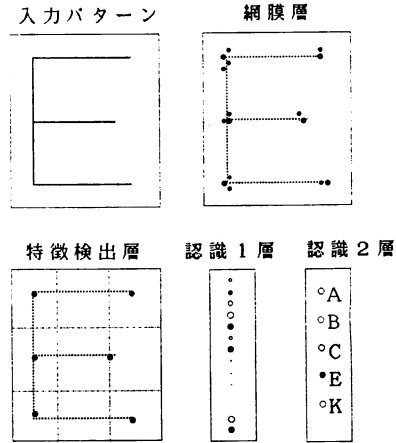


図5 各層の活動パターン

実験し、26種類のパターンがそれぞれ重なりがなく、独特の特徴で表わされていること、つまり、認識アルゴリズムが良ければ、これらの文字をすべて識別して認識できることを確認した。

次に、実際の手書き文字とプリンタ文字について認識の実験を行なった。手書き文字については、A, B, C, E, Kの五つの文字を選んで学習した後、6組の文字の認識率は90%であった。プリンタ文字の場合は、C, D, E, F, G, I, J, P, Q, R, T, U, Vの計13個の半角文字2組を学習させた後、半角、全角、倍角、縦倍角、4倍角各種の文字について50~78%の認識率が得られた。これらのなかでは、倍角文字の認識率が最も悪かった。倍角パターンについて学習させれば、認識確率は70%以上だと予想される。

図5は文字Eを入力したときのネットワーク各層の活動パターンを示している。黒い点は正の出力、白い点は負の出力を表わし、点のサイズが出力の大きさを表わしている。 C^2 層ではEを表わすユニットが他のユニットより大きい出力を出していることがわかる。他のアルファベット大文字やカタカナなども学習すれば認識できると思われる。

6. ハードウェア

本来並列処理をしている人間や生物の脳にあるニューラル・ネットワークをまねて作ったシステムを、コンピュータで逐次的に処理すると非常に遅い。また、ネットワークの規模が大きくなれば、これは不可能になる。本研究の最終的な目的は、システムをデバイス化して並列的に高速に、画像の入力・処理・認識を実現することである。この目的実現のために、まずデバイスを部分的に

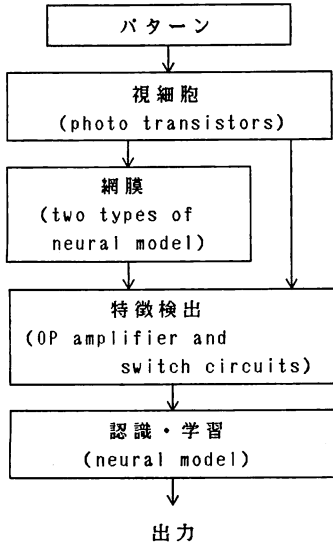


図6 デバイスで実現するシステム

試作してハードシミュレーションを試みた。

ハードシステムの構成を図6に示す。これまでに、網膜層と特徴検出層を試作した。実験では、網膜細胞の数を4×4、細胞の受容野を3×3とした。また、細胞の種類は交差点近傍で発火するoff中心型とした。

網膜層は、神経節細胞工学的モデルそのものである。パターンへの入力は光で行ない、入力細胞としてはフォトトランジスタを使い、光信号を電気信号に変える。入力は、それぞれ重み付けされ、加算され、閾値を越えたときに出力を出す。

特徴検出層では、特徴点とその種類を検出する。特徴点は網膜層からの信号の中から最大値を見つけることで得られる。最大値の検出にはコンパレータなどを使った。また、特徴点の種類を判断するため、細胞のまわりをある一定の大きさの四角形で区切っておく。ここで、特徴点が発火すれば、発火した特徴点に限って、その四角形のまわりの4辺それぞれの細胞の出力(0か1)のORをとり、それを加算する。その出力の大きさによって種類がわかる。たとえば、その出力が1ならば“-”，2ならば“7”，3ならば“1”，4ならば“+”という形である。その情報は、次の認識層に伝えられる。

実際、二本交差(鍵型“7”，十字型“+”など)パターンを入力し、システムの出力を調べた。図7は鍵型のパターンが入力されたときの特徴点検出までの各層の活動パターンである。この図から交差点という特徴点が発出できることが確認でき、ハードシステムが期待どおりに動作していることがわかった。

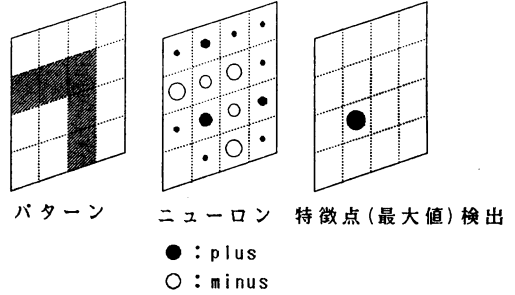


図7 デバイスで実験した結果

7. 結 び

ニューラル・ネットワークを使って、アルファベット大文字をほぼ正しく認識できることを示した。認識のためのアルゴリズムには、特徴検出モデルを採用した。パターンの部分部分の特徴だけを捉えて判断するので、パターンの大きさは認識には影響しない。また回転しなければ位置がずれても構わない。認識部分のC¹層とC²層のユニットの結合係数は学習で決めるので、原理的にはどんなパターンでも認識できる。

デバイス化することを目的として、ハードウェアでシステムを部分的に試作し、これが期待どおりに動作することを確認した。モデル自身が簡単で並列処理アルゴリズムであるので、人工網膜や特徴検出センサーの入力部分など3次元デバイスへの応用も期待できる。

現在のシステムでは、網膜層に2種類の神経細胞のモデルしか用いていないので、交差点と端点以外の特徴を検出することができない。そのため、曲線部分が多いかな文字の認識率は低く、アルファベットでも、“O”と“D”の区別などは難しい。現在、もっと柔軟性のあるアルゴリズムを検討中である。

文 献

- 1) J. J. Hopfield and D. W. Tank: "Computing with neural circuits: A model," *Science*, **233** (1986) 625-633.
- 2) G. Krishnan and D. Walters: "Psychologically plausible features for shape recognition in a neural network," *IEEE Int. Conf. on Neural Networks*, **2** (1988) 127-134.
- 3) 福島邦彦, 三宅 誠, 伊藤崇之, 河野隆志: "ネオコグニトロン原理を用いた手書き数字認識", *情報処理学会論文誌*, **28** (1987) 627-635.
- 4) C. A. Mead M. A. Mahowald: "A silicon model of early visual processing," *Neural Networks*, **1** (1988) 91-97.
- 5) 清水和彦, 合原一幸, 豊田雅嗣, 小谷 誠: "カオスニューラルネットワーク", 第3回生体・生理工学シンポジウム論文集 (1988) pp. 119-122.

- 6) H. P. Graf and P. deVegvar: "A CMOS implementation of a neural network model," Advanced Research in VLSI, Proceedings of the 1987 Stanford Conference.
- 7) T. F. Pawlicki, D. S. Lee, J. J. Hull and S. N. Srihari: "Neural network models and their application to handwritten digit recognition," IEEE Int. Conf. on Neural Networks, 2 (1988) 63-70.
- 8) 米津宏雄: "チップとデバイス技術の現状と将来—VLSIと仮想ハード—", Comput. Today, No. 27 (1988).
- 9) 塚原仲晃: 脳の情報処理 (朝倉書店, 東京, 1984).
- 10) K. T. Spoehr and S. W. Lehmkuhle (宇阪直行訳): 視覚の情報処理 (サイエンス社, 東京, 1986).
- 11) 塚原仲晃: 脳の情報処理 (朝倉書店, 東京, 1984).
- 12) 藤井真人, 福島邦彦, 伊藤崇之: "視覚神経に学んだエッジ検出", 第3回生体・生理工学シンポジウム論文集 (1988).
- 13) 梅野正義, 朱 小燕, 中村雅司: "生体の視覚系モデルによる文字パターン認識", 電気学会論文誌C (1988) 453-456.
- 14) 梅野正義, 清水昇市, 野間敏弘, 太田正光: "人工網膜のアプローチ", 名古屋工業大学学报, 35 (1984) 191-196.
- 15) 梅野正義, 枇杷木考恒: "文字パターン認識における自己組織化を利用した特徴量の検出", 電子通信学会論文誌, J69D, 10 (1986) 1443-1449.
- 16) D. E. Rumelhart and J. L. McClelland: *Parallel Distributed Processing* (MIT Press, 1986).
- 17) N. Watanabe, S. Nagata and K. Asakawa: "Mobile robot control by neural networks and their associated learning algorithm," IECON '88 (1988) p. 1054.
- 18) 賈 楨, 戸田尚宏, 白井文朗: "Back-Propagation アルゴリズムの加速化", 昭和63年度電気関係学会東海支部連合大会 (1988) p. 512.